UNIVERSITE
BRETAGNE
LOIRE / MATHSTIC

Université
Bretagne Sud
ubs :

THÈSE DE DOCTORAT DE

L'UNIVERSITE BRETAGNE SUD
COMUE UNIVERSITÉ BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : Mathématiques et leurs Interactions
Par

Jamila Mifdal

Applications du transport optimal et des méthodes non locales à la fusion d'images hyperspectrales et Multispectrales

Thèse présentée et soutenue à Vannes, le 25/11/2019
Unité de recherche : UMR CNRS 6205
Thèse N° : 537

Rapporteurs avant soutenance :

Coloma Ballester   Maître de conférences Universitat Pompeu Fabra, Barcelone
Julie Delon   Professeure   MAP5, Université Paris Descartes, Paris

Composition du Jury :
Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition ne comprend que les membres présents

Coloma Ballester   Maître de conférences Universitat Pompeu Fabra, Barcelone
Rapportrice
Antoni Buades   Professeur   Universitat des îles Baléares, Palma de Majorque
Membre du jury
Bartomeu Coll   Professeur   Universitat des îles Baléares, Palma de Majorque
Co-directeur
Nicolas Courty   Professeur   Université Bretagne Sud, Vannes
Co-encadrant
Julie Delon   Professeure   MAP5, Université Paris Descartes, Paris
Rapportrice
Jacques Froment   Professeur   Université Bretagne Sud, Vannes
Co-directeur

À Maman,

Pour la bonté de son coeur, pour sa patience et pour sa bonne humeur permanente.

À Papa,

Pour sa force, son ouverture d'esprit et pour son humour contagieux.

À toute la fratrie,

Pour nos rires en toutes circonstances.

Résumé

Le monde dans lequel nous vivons est observé par de nombreux satellites. En effet, grâce aux missions satellitaires, certaines zones de la Terre, à savoir les champs d'agriculture, le désert ainsi que les zones urbaines peuvent être surveillées efficacement. La surveillance de ces zones est faite grâce aux caméras embarquées à bord des satellites destinés aux missions d'observation de la Terre. En revanche, à cause des contraintes techniques et financières, le développement des capteurs d'imagerie haute résolution est assez limité. Par conséquent, des compromis sont constamment pris par les fabricants afin de construire des capteurs adaptés à la mission à laquelle ils sont destinés. Ainsi, les méthodes consistant à fusionner des données issues de plusieurs capteurs, permettent de contourner les limitations et produisent des images haute résolution.

En traitement d'images, la plupart des problèmes sont inverses et mal-posés. Inverses car, à partir du modèle physique d'acquisition de l'image et des observations disponibles, la scène observée doit être reconstruite. Les problèmes inverses sont considérés comme mal-posés lorsque la solution n'est pas unique et lorsque de faibles perturbations au niveau des observations peuvent conduire à des résultats très différents. L'omission ou la simplification des informations concernant la scène observée peut également conduire à une nouvelle image reconstruite. L'application de la régularisation à un problème mal-posé permet de le rendre bien-posé. La régularisation se traduit par l'ajout d'un terme de

régularisation dans le problème à résoudre. Le terme de régularisation décrit les propriétés que l'image à reconstruire doit satisfaire.

La problématique traitée dans la présente thèse est celle de la fusion d'images hyperspectrales et multispectrales. Une image hyperspectrale (HS) possède une haute résolution spectrale et une faible résolution spatiale, alors qu'une image multispectrale (MS) a une haute résolution spatiale et une faible résolution spectrale. Le but est donc de fusionner l'information pertinente contenue dans chacune des images afin de produire une image haute résolution avec la résolution spectrale de l'image hypespectrale et la résolution spatiale de l'image multispectrale.

Dans le présent document, une variété de méthodes pour la fusion d'images hyperspectrales et multispectrales est présentée. Le document de thèse est constitué de deux chapitres, chaque chapitre est divisé en deux sections qui décrivent chacune un nouveau modèle de fusion d'images hyperspectrale et multispectrale.

Dans le chapitre 2, Section 2.4, le problème de la fusion est résolu avec des outils de la théorie du transport optimal. Parmi ces outils figure la distance de Wasserstein régularisée définie comme suit

$$W_\gamma(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} <\pi, D> -\gamma E(\pi),$$

où μ et ν sont deux mesures de probabilités, D est une matrice de coût, π est le plan de transport, E est l'énergie entropique et γ est le coefficient de régularisation entropique.

L'énergie entropique est définie comme suit

$$E(\pi) = -\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{i,j}(\log \pi_{i,j} - 1) - \iota_{\mathbb{R}+}(\pi_{i,j}),$$

où ι est la fonction indicatrice de $\mathbb{R}_+$ telle que

$$\forall x, \quad \iota_{\mathbb{R}_+}(x) = \begin{cases} 0 & \text{si} \quad x \in \mathbb{R}_+, \\ +\infty & \text{sinon.} \end{cases}$$

Le problème de fusion est ainsi modélisé par la minimisation de la somme de deux distances de Wasserstein régularisées comme suit

$$\min_{\mathbf{u} \in \Sigma_u} G(\mathbf{u}) = \lambda W_{\gamma_M}(\mathbf{u}, \widetilde{\mathbf{S}}(\mathbf{f})) + (1 - \lambda) W_{\gamma_H}(\mathbf{u}, \widetilde{\mathbf{T}}(\mathbf{g})), \tag{1}$$

où $\mathbf{u}$ est l'image fusionnée à reconstruire, W_{γ_M} et W_{γ_H} sont les distances de Wasserstein régularisées relatives à la composante multispectrale et hyperspectrale respectivement, γ_M et γ_H sont des coefficients de régularisation entropique relatifs à W_{γ_M} et à W_{γ_H} respectivement, $\widetilde{\mathbf{S}}$ et $\widetilde{\mathbf{T}}$ sont des opérateurs d'inversion, λ est un paramètre déterminant le poids de chaque distance de Wasserstein et Σ_u est le simplex où $\mathbf{u}$ est défini.

La résolution du problème de minimisation (1) a été faite avec l'algorithme de Sinkhorn. Ce dernier permet l'utilisation des opérations matrices-vecteurs ce qui permet d'accélérer le calcul. La figure 1 montre les résultats visuels de la méthode de fusion proposée. Nous pouvons remarquer que visuellement, l'image fusionnée est similaire à l'image de référence et que la plupart des détails géométriques ont été récupérés à l'issu du process de fusion. Dans la Section 2.8.2, une étude détaillée des résultats expérimentaux est fournie. La

méthode proposée est également comparée, visuellement et quantitativement avec d'autres méthodes de l'état de l'art. La comparaison a montré que les résultats de la méthode suggérée se comparent favorablement avec d'autres méthodes de l'état de l'art.

Figure 1: Évaluation de la fusion HS-MS avec le transport optimal sur l'image Pavia $256 \times 256 \times 93$. De haut à gauche vers la droite. L'image référence, l'image HS, l'image MS et le résultat de la méthode proposée.

Dans la Section 2.9 du même chapitre, une version améliorée du précédent modèle de fusion est proposée. Dans le nouveau modèle de fusion, la nouveauté concerne la matrice de coût qui, dans le précédent modèle était définie comme suit

$$\xi = e^{-\frac{D}{\gamma}}$$

Dans le nouveau modèle de fusion, un terme nonlocal means décrivant la relation entre les patchs de l'image est ajouté. Le but du terme nonlocal est d'inclure les valeurs des pixels dans le transport et de prendre en compte le bruit

contenu dans les images de départ. Pour deux pixels i et j dont les coordonnées sont (x_i, y_i, z_i) et (x_j, y_j, z_j) respectivement, le nouveau coût du transport est le suivant

$$\xi(i,j) = \exp\left(-\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) - \frac{\alpha}{\gamma}(z_i - z_j)^2 - \frac{\|P(i) - P(j)\|_2^2}{h_{sim}^2}\right)$$

où α est un coefficient permettant d'équilibrer la distance spectrale par rapport à la distance spatiale, P désigne le patch extrait de l'image en question et h_{sim} est un paramètre de filtrage. La difficulté de cette méthode réside dans le nombre élevé des paramètres relatifs au modèle de fusion qui est égal à huit. Une étude de sensibilité des paramètres a montré que le nouveau coût n'a pas amélioré les résultats de la fusion en comparaison avec le premier modèle. En revanche, l'inclusion des poids NLM pourrait être bénéfique pour diminuer les effets du bruit géométrique, tel que les erreurs d'enregistrement qui peuvent affecter les images hyperspectrales et multispectrales lors de la phase d'acquisition. Cette nouvelle piste sera explorée dans la continuité de l'actuel travail de recherche.

Dans le chapitre 3, deux modèles de fusion basés sur le gradient nonlocal ont été proposés. Le premier modèle de fusion, présenté dans la Section 3.2, est exprimé comme la minimisation d'une fonctionnelle constituée de quatre termes d'énergie.

$$\min_{u \in \mathbb{R}^{H \times N}} \sum_{h=1}^{H} \|\nabla_{\omega_h} u_h\|_1 + \frac{\mu}{2} \sum_{h=1}^{H} \|DBu_h - g_h\|_2^2 + \tag{2}$$
$$\frac{\gamma}{2} \sum_{m=1}^{M} \|(Su)_m - f_m\|_2^2 + \frac{\lambda}{2} \sum_{h=1}^{H} \|\tilde{P}_h u_h - P_h \tilde{g}_h\|_2^2,$$

Le terme $\|\nabla_{\omega_h} u_h\|_1$ est un régularisateur défini comme suit

$$\|\nabla_{\omega_h} u_h\|_1 = \sum_i \left|\nabla_{\omega_h} u_{h,i}\right|$$

où $|\cdot|$ dans ce cas désigne la norme L^2 et $\nabla_{\omega_h} u_h \in \mathbb{R}^{N \times N}$ est le gradient nonlocal calculé pour chaque pixel tel que

$$\nabla_{\omega_h} u_{h,i} = \left((\nabla_{\omega_h} u_{h,i})_1, \ldots, (\nabla_{\omega_h} u_{h,i})_N \right),$$

avec

$$(\nabla_{\omega_h} u_{h,i})_j = \sqrt{\omega_{h,i,j}} \left(u_{h,j} - u_{h,i} \right),$$

où $\omega_{h,i,j}$ est le poids nonlocal défini comme suit

$$\omega_{h,i,j} = \begin{cases} \dfrac{1}{\Gamma_i} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{m=1}^{M} s_{m,h} \sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|f_{m,i+t} - f_{m,j+t}\|_2^2}{h_{sim}^2 (2\nu_c + 1)^2 \sum_{m=1}^{M} s_{m,h}} \right) & \text{si} \\ & \|i-j\|_\infty \leqslant \nu_r \\ & (i \neq j) \\ 0 & \text{sinon} \end{cases}$$

$$\Gamma_i = \sum_{\{j : \|i-j\|_\infty \leqslant \nu_r\}} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{m=1}^{M} s_{m,h} \sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|f_{m,i+t} - f_{m,j+t}\|_2^2}{h_{sim}^2 (2\nu_c + 1)^2 \sum_{m=1}^{M} s_{m,h}} \right)$$

$$\omega_{h,i,i} = \max \left\{ \omega_{h,i,j} : \|i-j\|_\infty \leqslant \nu_r \text{ et } j \neq i \right\}$$

Contrairement au gradient classique qui permet une interaction locale entre les pixels, le gradient nonlocal peut être calculé entre deux pixels quelconques dans l'image quelque soit leur position.

Les deux termes d'attache aux données $\|DBu_h - g_h\|_2^2$ et $\|(Su)_m - f_m\|_2^2$ sont relatifs à la génération de l'image multispectrale et hyperspectrale. L'opérateur B est un filtre passe-bas, D est un opérateur de sous-échantillonnage et S représente la réponse spectrale du capteur multispectrale. Le rôle des termes d'attache aux données est de pénaliser la déviation de la solution **u** de l'image hyperspectrale

g_h et multispectrale f_m. La pénalisation est faite à travers la norme L^2.

Le dernier terme $\|\tilde{P}_h u_h - P_h \tilde{g}_h\|_2^2$ est une contrainte radiométrique pénalisée avec la norme L^2. Le but de cette contrainte est de forcer l'image fusionnée **u** à partager les mêmes hautes fréquences spatiales que l'image multispectrale **f**, l'égalité suivante est donc imposée pour chaque bande hyperspectrale h

$$\frac{u_h}{P_h} = \frac{\tilde{g}_h}{\tilde{P}_h}, \ \forall h \in \{1, \ldots, H\}, \tag{4}$$

L'égalité (4) est équivalente à $u_h - \tilde{g}_h = \frac{\tilde{g}_h}{\tilde{P}_h}\left(P_h - \tilde{P}_h\right)$. Cette dernière signifie que l'on impose que les hautes fréquences de chaque bande hyperspectrale $u_h - \tilde{g}_h$ coïncident avec les hautes fréquences de l'image multispectrale $P_h - \widetilde{P}_h$ de telle façon à ce que les détails géométriques soit injectés dans l'image fusionnée. Le coefficient de modulation $\frac{\tilde{g}_h}{\tilde{P}_h}$ prend en compte l'énergie hyperspectrale et multispectrale de chaque bande h. Les différentes composantes de la contrainte radiométrique sont définies comme suit

- $\mathbf{P} \in \mathbb{R}^{H \times N}$ est une combinaison linéaire de l'image multispectrale telle que,

$$P_h = \sum_{m=1}^{M} \alpha_{m,h} f_m, \ \forall h \in \{1, \ldots, H\}. \tag{5}$$

- $\tilde{\mathbf{P}} \in \mathbb{R}^{H \times N}$ est calculé comme suit

$$\tilde{P}_h = \sum_{m=1}^{M} \alpha_{m,h} \tilde{f}_m, \ \forall h \in \{1, \ldots, H\}. \tag{6}$$

- $\tilde{\mathbf{g}} \in \mathbb{R}^{H \times N}$ est l'image hyperspectrale ramenée à la grille fine par interpolation bicubique.

avec $\alpha_{m,h}$ défini comme suit

$$\alpha_{m,h} = \frac{s_{m,h}}{\sum_m s_{m,h}}$$

où $s_{m,h}$ sont des coefficients extraits de la réponse spectrale du capteur multi-spectral.

L'image $\tilde{\mathbf{P}}$ est calculée comme dans (6) mais cette fois-ci avec l'image $\tilde{\mathbf{f}}$. Cette dernière est le résultat de la dégradation spatiale de l'image multispectrale comme dans le modèle de génération (1.7) suivi par une interpolation bicu-bique afin de revenir sur la grille de l'image multispectrale. Cette succession d'opérations dégrade amplement les hautes fréquences spatiales tout en préser-vant les basses fréquences. La contrainte radiométrique est détaillée davantage dans le reste du document. La figure 2 montre les résultats visuels de la méthode proposée. Nous remarquons que l'information spectrale ainsi que l'information spatiale ont été reconstruites et que l'image fusionnée ressemble visuellement à l'image référence. Des mesures objectives avec des indices de qualité ont montré que la méthode proposée se compare favorablement aux méthodes de l'état de l'art.

Dans la Section 3.7, des modifications ont été apportées au modèle initial (2) afin d'aboutir au modèle suivant

$$\min_{u \in \mathbb{R}^{H \times N}} \sum_{h=1}^{H} \|\nabla_{\omega_h'} u_h\|_1 + \frac{\mu}{2} \sum_{h=1}^{H} \|DBu_h - g_h\|_2^2 +$$

$$\frac{\gamma}{2} \sum_{m=1}^{M} \|(Su)_m - f_m\|_2^2 + \frac{\lambda}{2} \sum_{h=1}^{H} \|\tilde{P}_h u_h - P_h \tilde{g}_h\|_1,$$

avec les nouveaux poids nonlocaux définis comme suit

$$\omega_{h,i,j}' = \begin{cases} \dfrac{1}{\Gamma_i'} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|\tilde{g}_{h,i+t} - \tilde{g}_{h,j+t}\|_2^2}{h_{sim}^2 (2\nu_c + 1)^2}\right) & \text{si } \|i-j\|_\infty \leqslant \nu_r \ (i \neq j) \\[4ex] 0 & \text{sinon} \end{cases}$$

Figure 2: Évaluation visuelle des performances de la fusion HS-MS avec le gradient nonlocal sur l'image Chikusei $304 \times 304 \times 93$. De haut à gauche vers la droite. L'image référence, l'image HS, l'image MS et le résultat de la méthode proposée.

et

$$\Gamma_i' = \sum_{\{j:\|i-j\|_\infty \leqslant \nu_r\}} \exp\left(-\frac{\|i-j\|_2^2}{h_{spt}^2} - \frac{\sum_{\{t\in\mathbb{Z}^2:\|t\|_\infty \leqslant \nu_c\}} \|\tilde{g}_{h,i+t} - \tilde{g}_{h,j+t}\|_2^2}{h_{sim}^2(2\nu_c+1)^2}\right)$$

où

$$\omega_{h,i,i}' = \max\left\{\omega_{h,i,j} : \|i-j\|_\infty \leqslant \nu_r \text{ et } j \neq i\right\}$$

Les différents termes impliqués dans la contrainte radiométrique à savoir $\mathbf{P}$, $\tilde{\mathbf{P}}$ et $\tilde{\mathbf{g}}$ sont les résultats de transformations de l'image multispectrale et hyperspectrale comme cela a été montré avant. Par conséquent, il est possible qu'un nouveau bruit ait été créé à travers ces transformations. La norme L^2 est efficace dans le cas d'un bruit gaussien et moins efficace pour d'autres types de bruit. Ainsi, la norme L^2 n'est peut être pas suffisamment adaptée au type de bruit fabriqué lors des transformations mentionnées ci-dessus. Par conséquent, dans le nouveau modèle, la contrainte radiométrique a été pénalisée avec la norme L^1. Les expérimentations dans cette partie ont porté sur des images

issues d'une caméra hyperspectrale représentant des scènes de la vie de tous les jours. La figure 3 montre le résultat visuel de la modification proposée. Nous remarquons que l'image fusionnée ressemble visuellement à l'image référence et que les détails spatiaux ainsi que les détails spectraux ont été reconstruits. L'analyse des résultats quantitatifs montre que la modification proposée du modèle se compare favorablement aux méthodes de l'état de l'art et dépasse même les résultats du modèle initial dans quelques mesures de qualité.
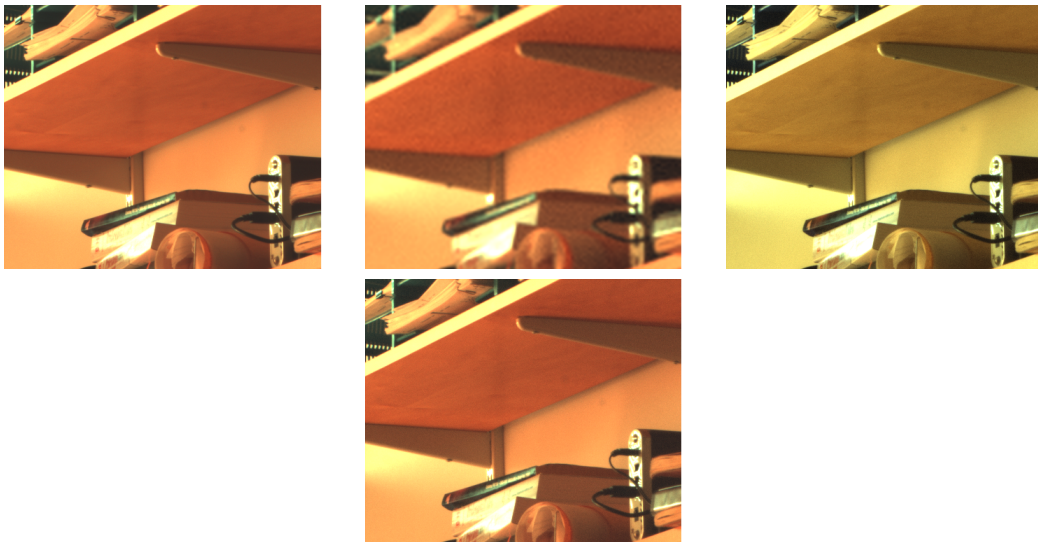
Figure 3: Évaluation visuelle des performances de la fusion HS-MS dans le cas de la pénalisation de la contrainte radiométrique avec la norme L^1 sur l'image "Bookshelves" de résolution spatiale $512 \times 512 \times 31$. De haut à gauche vers la droite. L'image référence, l'image HS, l'image MS et le résultat de la méthode proposée.

Abstract

The world we live in is constantly under observation. Many areas such as offshore zones, deserts, agricultural land and cities are monitored. This monitoring is done throughout remote sensing satellites or cameras mounted on aircrafts. However, because of many technological and financial constraints, the development of imaging sensors with high accuracy is limited. Thus, compromises are constantly made by the manufacturers to adapt the sensors to the application they are destined for. Therefore, solutions such as multi-sensor data fusion overcome the different limitations and produce images with high quality. The idea behind image fusion is the combination of spatial and spectral characteristics from different sensors into one image with high spatial and spectral quality.

Most of image processing tasks are ill-posed inverse problems. Inverse because from the physical imaging process and available observations one seeks to reconstruct the original observed scene or object. Inverse problems are called ill-posed because the solution depends on the observations and small change of these latter may lead to a different outcome. Discarding or simplifying information about the scene may also lead to a different solution of the image to be reconstructed. In order to account for the ill-posedness of inverse problems, a regularization term is imposed to get the desired properties of the original image in the final solution.

In this dissertation various methods for dealing with hyperspectral and multispectral image fusion are presented. The first part of the thesis uses tools from the optimal transport theory namely the regularized Wasserstein distances. The fusion problem is thus modeled as the minimization of the sum of two regularized Wasserstein distances. Each one of these distances recovers either the spectral or the spatial characteristics contained in the input images. The minimization problem is strongly convex and it is solved with Sinkhorn-Knopp's algorithm. The visual and quantitative evaluation of the fused model on satellite images compares favorably to the state-of-the-art fusion methods. A modification of the previous fusion model is also suggested. It consists in adding nonlocal means (NLM) weights in the cost matrix. Including the NLM weights allows the mixing of the distances between the pixels and their values. The goal of the NLM weights is to improve the transport between the pixels and take into account the noise of the input data. The experimental results showed that the NLM weights did not improve the transport and did not reduce the noise. However, the NLM weights might mitigate the misregistration that affects data during the acquisition process. This line of research will be tested in a future work. In the second part of this thesis, the hyperspectral and the multispectral fusion problem is presented differently. The latter is modeled as the minimization of four energy terms. The first term is a nonlocal regularizer that accounts for the ill-posedness of the problem. Throughtout the term of regularization, the fused image is forced to share the fine geometric details of the multispectral image. The second and the third terms are related to the data generation model, their minimization forces the fused image to stay as close as possible to the hyperspectral and the multispectral image. The last term of the fusion model is a radiometric constraint. Its goal is to force the fused image to share the geometric details and the high spatial frequencies

of the multispectral image, for the latter contains the relevant spatial details. The regularization term was penalized with the $L^{1,2}$ norm and the rest of the terms of the model were penalized with the L^2 norm. In a second step, the fusion model was upgraded by penalizing the radiometric constraint with the L^1 norm. Both fusion problems were solved with the gradient-descent algorithm of Chambolle and Pock. Experimental results were conducted on multiple datasets and the fusion was assessed visually and quantitatively. The performance of both models compares favorably with the state-of-the-art methods.

Acknowledgements

Je voudrais commencer par remercier profondément Docteur Coloma Ballester et Professeure Julie Delon pour avoir rapporté mon manuscrit de thèse. Je remercie également Professeur Antoni Buades pour avoir accepté de faire partie de mon jury. Je suis très fière de soutenir devant vous.

Je voudrais remercier sincèrement mes deux co-directeurs de thèse Professeurs Jacques Froment et Bartomeu Coll ainsi que mon co-encadrant Professeur Nicolas Courty, d'avoir rendu cette thèse possible. Je les remercie de m'avoir donné l'opportunité de faire partie du laboratoire LMBA et de faire partie de l'équipe pétillante OBELIX à l'IRISA. Les échanges et les intéractions avec les collègues au sein de ces deux laboratoires ont été enrichissants aussi bien sur le plan professionnel que sur le plan personnel. Je les remercie également de m'avoir permis d'être en cotutelle avec l'Université des Îles Baléares, où une grande partie du contenu scientifique de ma thèse a été développée, et puis, les séjours à Palma ont toujours été très agréables.

Je remercie Jacques, de m'avoir proposé un sujet de thèse qui m'a toujours autant passionné et dont les applications diverses et variées ne cessent de m'attirer. Je le remercie pour ses disponibilités et pour les discussions intéressantes qui m'ont permis de développer un

esprit critique et d'aborder les questions scientifiques par plusieurs angles différents et avec une nouvelle façon de penser.

Je tiens à remercier Nicolas d'avoir été mon co-encadrant et avec qui c'était un plaisir de travailler. Je le remercie de m'avoir fait découvrir la théorie du transport optimal et de m'avoir guidé durant mes premiers pas dans l'application de cette théorie à la fusion HS-MS. Nos discussions m'ont toujours permis d'explorer de nouvelles pistes.

Je voudrais exprimer ma sincère gratitude et ma profonde reconnaissance à mon deuxième co-directeur de thèse Tomeu, avec qui, c'était juste un énorme plaisir de travailler. Son dynamisme, sa bienveillance, sa curiosité, son humanité et sa grande modestie m'ont donné encore plus envie de continuer dans la recherche, ses qualités resteront une source d'inspiration et un modèle à suivre dans ma carrière scientfique. Je le rermercie d'avoir été toujours à l'écoute et de s'être toujours préoccupé de l'état d'avancement de mes travaux. Je le remercie pour la qualité de son encadrement et pour sa générosité dans le partage du savoir, ses encouragements m'ont toujours donné la force de persévérer et de réussir. Moltes gracies per tots.

I would like to thank Joan, with whom I collaborated and worked during my PhD and during my stays in Palma. His scientific knowledge and intuitions along with his modesty, made working with him an enriching experience and a great work opportunity in my PhD thesis. Thanks to my collaboration with him, I was able to discover the world of variational models and the applications of non-local gradients to image processing. I'm sincerely grateful for everything I learnt from him. Moltes gracies.

Durant mes années de thèse, j'ai eu la chance de donner des cours à l'Université de Bretagne Sud. Pour cela, je voudrais remercier tous mes étudiants et aussi les collègues : Emmanuel, Bertrand, Béatrice et Quansheng avec qui c'était un plaisir d'enseigner. Je tiens également à remercier le pôle doctoral de leur écoute et de leur aide.

Mes années de thèse m'ont permis de rencontrer beaucoup de belles personnes avec qui j'ai vécu des moments agréables et avec qui j'ai noué des relations d'amitiés qui, j'espère, dureront pour le restant de mes jours. Je tiens à remercier Ronan, pour les fous rires et les délires interminables, et aussi pour les parties de pétanque dans les couloirs de l'université après 20h, Raounak, pour son sourire, sa joie de vivre ainsi que pour nos sujets de discussion choisis avec grand soin, Hélène, pour les discussions profondes et les agréables pauses "capoutche" qui ont toujours duré plus de dix minutes, à Lucie, la voix de la raison, pour son grand coeur et son côté loufoque insoupçonné, Claire, pour sa clairvoyance et sa sagesse, Nan, pour être toujours fidèle à elle-même, Elyes, pour sa grande gentillesse et son humour atypique, Kévin, qui a toujours su faire preuve de patience lorsque je ruinais ses Rubik's-Cubes, Jean-Christophe, pour sa modestie débordante, Anne-Charlotte, pour son énergie et pour m'avoir toujours accompagné dans mes délires, Rihem, la douce, pour la bonté de son coeur, Thuy, pour sa bonne humeur permanente, Xiao pour sa sympatie et Erwan pour sa discrétion. Je remercie également ma copine de longue date Mathilde, malgré la distance, elle m'a toujours suivie de près. Je tiens à remercier Véronique pour sa bienveillance, malgré le fait qu'elle soit partie, elle a toujours pris de mes nouvelles. Je remercie également Sandrine,

avec qui cela a toujours été agréable de discuter.

Enfin, un grand merci à mes deux frères, que j'aime profondément, d'avoir toujours veillé sur moi. Merci à Momo, le prévenant, pour son affection et sa gentillesse, à mon petit frère le Russe, pour son amour et pour avoir toujours été mon fidèle complice, à Aya, notre petit rayon de soleil, qui, même en ne comprenant rien à ce qu'il se passe autour d'elle, elle m'a toujours comblée de son amour et souhaitée bonne chance, et enfin, j'aimerais remercier mes merveilleux parents pour leur amour et leur soutien inconditionels. Je les remercie pour toutes les libertés qu'ils m'ont données et pour avoir toujours été présents à mes côtés. Je les remercie d'avoir été compréhensifs et patients surtout dans les moments où j'étais très (trop ?) insupportable. Les rendre et les voir fiers a toujours été ma plus grande source d'énergie pour réussir.

Bibliogrpahy of the Thesis

The following publicatins were produced on the research presented in this thesis:

- **International conference articles**

    - J. Mifdal, B. Coll, N. Courty, J. Froment, and B. Vedel, "Hyperspectral and multispectral Wasserstein barycenter for image fusion," in *Proc. IEEE Geosci. Remote Sens. Symp. (IGARSS)*, (Texas, TX, USA), 2017

    - J. Mifdal, B. Coll, and J. Duran, "A variational formulation for hyperspectral and multispectral image fusion," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3328–3332, IEEE, 2018

- **International journal papers**
  J. Mifdal, B. Coll, J. Froment, J. Duran, "A variational approach for hyerspectral data fusion and supperesolution", The journal paper is currently being written

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The use of digital images to understand the world around us has been rapidly growing during the last years. In fact, the availability of developed imaging devices gave birth to high quality images that allowed many applications in many fields. The use of machines destined to process and interpret these images is what is called computer vision. For humans, visualizing and analyzing data such as counting the number of individuals in a crowd, distinguishing their skin colors etc is automatic and we do not even think about it. However, such tasks for a computer are extremely difficult and challenging. Thanks to the big development of image processing methods that are integrated in computer vision machines, many devices like drones, medical machines etc are able to receive, analyze and process visual data on their own in order to make decisions.

Reconstructing images through the eye or through a computer vision machine can be seen as an *ill-posed inverse problem*. In fact, it is widely known in the image processing society that the human eye perceives visual information from a limited number of points. How come we can see everything around us then? It is because, for a human being with good health (mental and physical) conditions, the brain completes the perceived image by interpolating and extrapolating

the data received from different points. If the scene is familiar to us then the true image of the latter can be adequately reconstructed. Thus, the problem of reconstructing the image of a scene is ill-posed because of the non-uniqueness and the instability of the solution. The human brain is usually quick in solving ill-posed problems by using its previous experience (*a priori* information).

In the majority of image processing tasks, ill-posed inverse problems are very present. The ill-posedness is due to the instability of the solution which is itself due to the errors that might affect the observations. These problems are inverse because one seek to deduce the original observed scene from a set of physical data. Ill-posed inverse problems are solved by choosing priors that impose and describe the desired characteristics in the final solution. These priors *regularize* the problem and make it *well-posed* because, thanks to the regularization, in most of the cases, a unique solution can be found.

1.1 Image formation models

1.1.1 The human visual system

The human visual system is very complex and it has been the subject of many studies that led to much progress in developing image processing techniques and imaging devices. The human eye perceives colors because it is composed of photoceptors named cones which are sensitive to three different bands of the electromagnetic spectrum as shown in Figure 1.1. The B-Cones are sensitive to the spectrum in the range of wavelengths between 400 nm and 550 nm peaking at 440 nm which covers violet, blue and cyan light. The G-Cones are sensitive to the spectrum in the range of wavelengths between 430 nm and 670 nm peaking at 550 nm which covers cyan, green, yellow and orange light. As to the R-Cones they are sensitive to the spectrum in the range of wavelengths between 500 nm

and 760 nm peaking at 575 nm. This spectrum covers green, yellow, orange and red light.
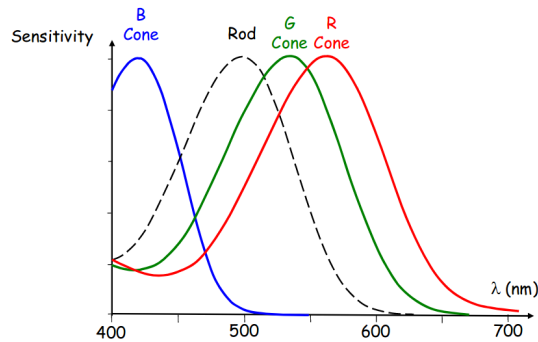
Figure 1.1: The spectral responses of rods and cones

The spectral response of the human eye can be found throughout an averaging of the three cone sensitivities as shown in Figure 1.2. As one can notice, the eye's visual response peaks in the region corresponding to the green light, this shows that the eye responds strongly to the green light and that we are more sensitive to the green light than to any other color in the visible spectrum. If we are presented with three light sources: red, green and blue with equal optical power, the green would be perceived as the brightest.
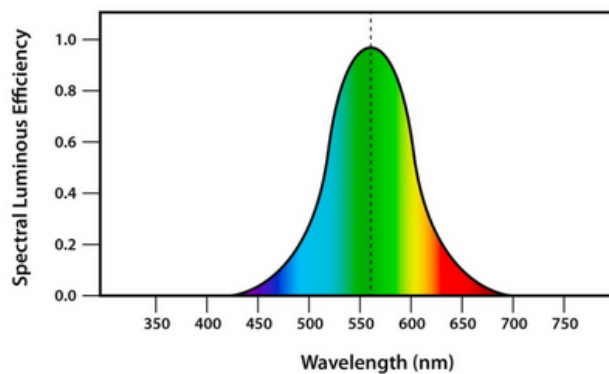
Figure 1.2: The spectral sensitivity of the human eye

The human visual system has a fascinating and a complex structure. Under-

standing the latter led to the development of algorithms and imaging devices that replicate roughly the functioning of the human eye and help visualize and interpret the world around us. For example, the physiological understanding of the eye's cells inspired the theory of edge detection of Marr [3]. It also led to spatial filtering, introduced in the early work of Hubel and Wiesel [4], which is a fundamental operation that is heavily used in image processing.

1.1.2   Digital imagery

A digital image is a binary representation of visual information such as scenes and objects and it is electronically saved on any storage device. Each digital image is recorded as many numbers, it is divided into a matrix or array of small picture elements called pixels. A pixel is the smallest component in an image that can be displayed and represented on a digital display device. The value of a pixel is related to the brightness or color of the represented scene.

Like in the visual human eye system, an aperture at the front of the camera opens to let light through. Then, the electronic equipment of the camera captures the incoming light rays and turns them into electrical signals. The light detector is one of two types, either a charge-coupled device (CCD) or a complementary metal-oxide semiconductor (CMOS) image sensor. Both CCD and CMOS image sensors convert light into electric charge through photodiode and process it into electronic signals. Both technologies use an array of millions of tiny photo sensors (photodiodes). Each one of these sensors creates an electrical current when exposed to light. In a CCD device, the charge is transported across the chip and read at one corner of the array. Then, each pixel's value is turned into a digital value with an analog-to-digital converter. In CMOS devices, there are several transistors at each pixel that amplify and move the charge using more traditional wires and each pixel can be read individually. CCD sensors create

high-quality, low noise images whereas CMOS sensors are more susceptible to noise. Furthermore, the light sensitivity of a CMOS sensor is low. In fact, each pixel on a CMOS sensor has several transistors next to it, many electrons hit the transistors instead of the photodiode, thus the amount of electrons hitting the photodiode is low. CMOS sensors consume little power whereas CCDs use a process that consumes lots of power. CCDs consume as much as 100 times more power than an equivalent CMOS sensor.

In each digital imaging system, a number of bits is made available to represent each pixel in the image, it is what is called the pixel bit depth. Thus, each pixel is represented as a series of bits, in most imaging devices the pixel bit depth is equal to eight bits which means that each pixel can have 256 different brightness levels.

Each image sensor has small individual detectors that respond to light that hits them.The light energy generates an electric current at each detector. The current is then measured electronically and converted to a digital value that represents the amount of light detected. However, the sensor responds to all wavelengths in the visible spectrum and produces only black and white (grayscale) images. The sensor itself does not discriminate colors, thus, detectors are needed to separate red, green and blue wavenlengths. Adding three sensors is one solution to discriminate colors but it makes the camera bulky and very expensive. Instead, most cameras use a filter array as shown in Figure 1.3. The filter array is a single sensor that has pixels that respond to either red, green or blue color. Figure 1.3 shows the most common type of color filter arrays which is called "Bayer array". The latter is placed on top of the image sensor to make each element of the sensor detect either red, green or blue color. An algorithm incorporated in the camera has to approximate the other two primary colors in order to have full color at every pixel. As one can notice in Figure 1.3, there are

more green elements than there are red or blue. This is due to the fact that the human eye is more sensitive to the green color than to the two other ones as illustrated in Figure 1.2. The redundancy with green pixels produces an image which appears less noisy.
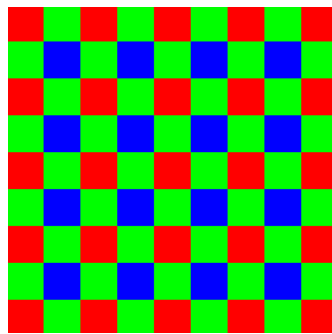
Figure 1.3: Bayer filter

Since at each pixel the information of the three primary colors is needed (red, green and blue), image processing is carried out in order to convert the detected image (with the color filter) to a single full-color image. This process is known as demosaicking. The latter consists in translating the Bayer array of primary colors into a final image which contains full color information at each pixel.

1.1.3   Satellite imagery

Satellite sensors can provide three types of images with different spectral and spatial resolutions: the panchromatic, the multispectral and the hyperspectral image. The panchromatic (Pan) image is a grayscale image that has a very high spatial resolution but only one spectral band, most of the time, in the visible spectrum. The multispectral (MS) image has a high spatial resolution, lower than the panchromatic's, and few spectral bands. Finally, the hyperspectral (HS) image has a low spatial resolution, compared to the multispectral image, but it has many contiguous spectral bands. Figure 1.4 shows an example of a

hyperspectral image with a spatial resolution of 30 m and 224 spectral bands. The figure also shows that each object of the image has its own spectral signature which is useful for instance to distinguish between the objects of a captured scene.
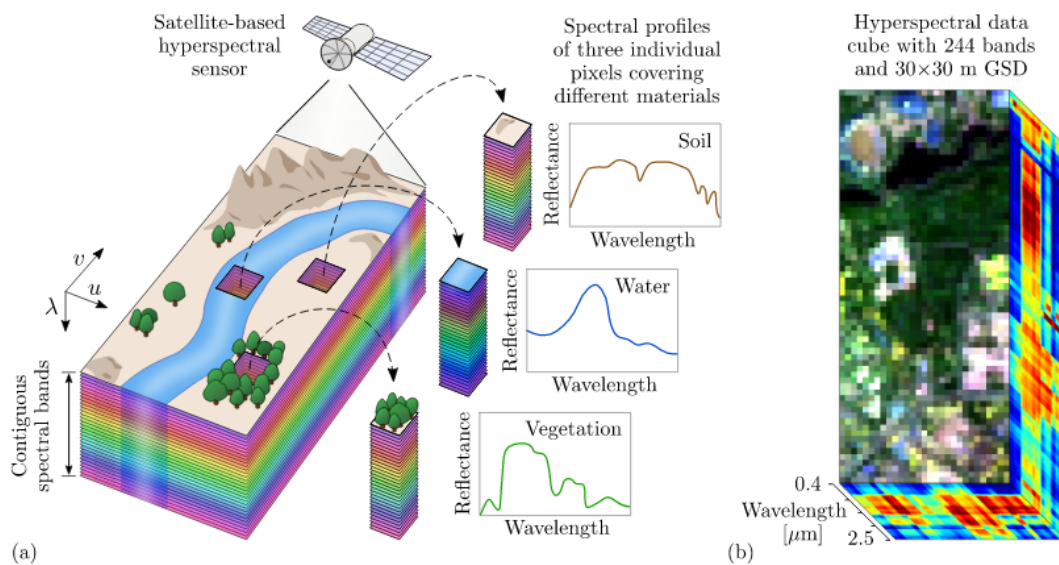
Figure 1.4: Representation of hyperspectral data spatially degraded but with many contiguous bands [5].

For several decades, satellite imagery has been used in various applications and it made significant contributions to our understanding of earth processes and human-environment interactions. The high spectral resolution in the hyperspectral images allows to see the unseen, the availability of the multiple spectral bands reveal valuable information about the acquired object or scene. Hyperspectral images are used, for instance, in quality food inspection, study of composition of materials etc. The panchromatic or the multispectral image are used for their high spatial resolution. The latter describes how much detail is visible to the human eye, the higher the spatial resolution the finer the geometric details. Having fine geometric details in an image is very important because it

helps in different tasks such as vegetation mapping, objects recognition etc.

Most observation satellites like Ikonos, Landsat, Quickbird and Pléiades decouple the acquisition of panchromatic, multispectral and hyperspectral images. The decoupling of the acquisition is due to technical constraints of satellite sensors such as the on-board storage capacity and the transmission bandwidth. The hyperspectral cameras sample the electromagnetic spectrum in very contiguous and narrow spectral bands. Most of imaging sensors contain pushbroom scanners as highlighted in Figure 1.5. This kind of scanners contain linear array sensors, which, in most cases correspond to CCD devices. With this kind of scanning, the recording of the image is carried out line by line as the spacecraft that contains the imaging sensor flies forward. For each line perpendicular to the flight direction, the pixels are measured simultaneously. Pushbroom scanners have an important drawback, the detectors can have varying sensitivity which can cause some noise in the final image. Thus, the narrower the bands the higher the noise because of the difference of sensitivity of different sensors. One way of reducing the impact of noise is the resampling of the spatial grid of the image. In the case of hyperspectral images, the spatial solution is compromised and it has to undergo resampling in order to reduce noise and keep the spectral resolution. Unlike the hyperspectral case, multispectral cameras contain a few optical sensors that are sensitive to very spaced wavelengths and that apply less resampling than in the hyperspectral sensors, which leads to high spatial resolution. In the case of the panchromatic image, only one band is registered with a very high spatial resolution.

For the visualization of multiband images that have more than three bands, three spectral bands are randomly selected and the image is plotted as a RGB one. Thus, the colors displayed on the images are false colors and they change following the selected three bands.
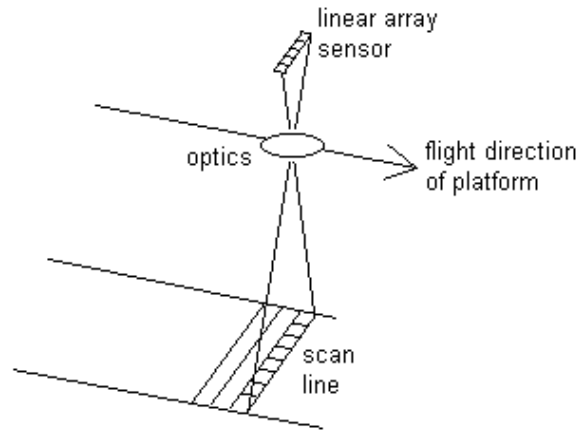
Figure 1.5: A pushbroom scanner

1.2 Inverse problems

Inverse problems involve the estimation of "real data" from given observations, these latter are usually noisy and suffer from incomplete information due to physical limitations of imaging devices. Electronic sensors acquire properties of "reality" that are not the same as the natural sensors: our eyes. This statement has even a philosophical side, Inmanuel Kant said: "Things that we see are not by themselves what we see" [6]. The process of acquiring an image can be modeled by the integral of a function describing the physical properties of the scene to be acquired. For example, the light detector CCD in a camera integrates light on a finite interval of the electromagnetic spectrum and gives as a result an electrical signal, which is by itself converted to a digital signal, the so-called gray level. To sum up, what we see are not the objects by themselves but their integrals. Thus, extracting the information about the objects involves inversing the image generation model, which can be difficult to solve.

In order to illustrate the idea behind inverse problems, let us take the example of the restoration of a "clean" image given noisy observations. The forward

model consists in generating a noisy image f from a sharp image u through the convolution model $f = \kappa * u + \eta$, where κ is a convolutional kernel representing the imaging device and η is some noise affecting the generated data. The example is illustrated in Figure 1.6 and it shows that the result of recovering u is not pleasant.

| Clean | Noisy | Denoised |
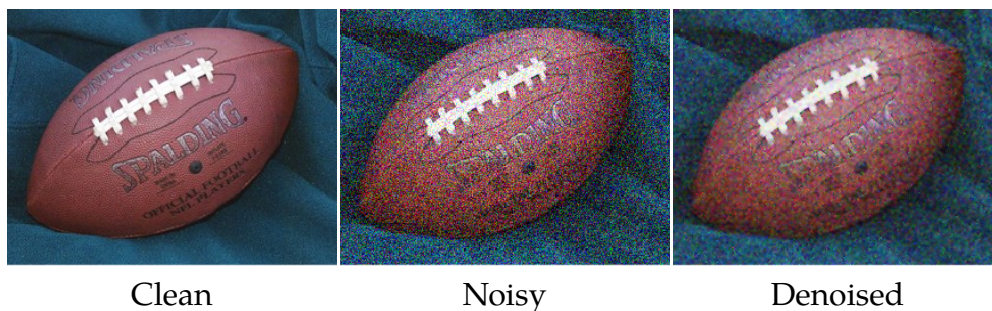
Figure 1.6: From the forward generation model $f = \kappa * u + \eta$ a noisy image f was generated from a clean image u with some noise. The inverse process consists in recovering u given f. The denoising was carried out with the Wiener filter

Inverting the forward model in order to recover a denoised image is not straight forward because one has to invert the operator κ which is not always invertible. For instance, in the case where κ is the Gaussian kernel, the inversion problem is equivalent to inverting the heat equation and it is well known that this is an ill-posed problem.

1.2.1 Ill-posed inverse problems

In the field of image processing, hypotheses are made on the physical process of generation of observations. Thus, problems are usually modeled by a system of equations to be solved, it is the so-called *image model*. Then, the goal is to find "the real initial" data given the observations. However, most of the time, the image model is not invertable which causes the ill-posedness of the image problem in the sense that a unique solution cannot be found. In fact, according

to Hadamard [7], a mathematical problem is well-posed when

i) the solution exists,

ii) the solution is unique,

iii) the solution depends continuously on the initial data, which means that the solution must be robust to the noise.

A problem that violates any of the well-posedness properties is an ill-posed problem. The first and the second condition deal with the feasability of the inverse problem, whereas the third condition is about the stability of the solution. When the solution does not depend "smoothly" on the available data, small perturbations on the latter may lead to huge variations in the solution that becomes therefore highly instable.

Most image processing problems are linear, thus, from now on, the study of inverse problem will be restricted to linear ones. The majority of image models can be represented as follows. Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces and $\Phi : \mathcal{H} \to \mathcal{K}$ a linear forward operator. The operator Φ maps data u contained in $\mathcal{H}$ to another kind of data f in $\mathcal{K}$. We consider the following equation

$$\Phi u = f \tag{1.1}$$

The model (1.1) represents a forward problem and the goal is to compute an approximation of the true image u such that

$$u = \Phi^{-1} f \tag{1.2}$$

Problem (1.2) is usually ill-posed because the operator Φ is not always invertible. In order to make (1.2) well-posed, Φ has to be bounded and surjective on $\mathcal{K}$, which ensures the existence of a solution, and injective on $\mathcal{H}$ to ensure

the uniqueness of the solution. From the closed graph theorem it follows that Φ^{-1} is continuous which ensures that the third condition of well posedness is satisfied.

Data recovered from imaging systems are usually corrupted with noise and can thus be represented as follows

$$f_\eta = \Phi u + \eta \tag{1.3}$$

where f_η is the noisy observed data and η is a realization of an additive noise introduced by transmission, electronical imaging devices etc. In the rest of the document f_η will be denoted f in order to point out noisy data.

1.2.2   Regularization approach to ensure well-posedness

In image processing, most problems are inverse ones, because the aim of these problems is to estimate data from inadequate or noisy observations. In the majority of cases, inverse problems are ill-posed and they can be reduced to well-posed problems by transforming the functional to be optimized. One way of making an ill-posed problem well-posed, is to provide additional information to the image model, such as hypotheses that take into account the properties of the image to be found or constraints that impose that the final image takes a particular structure. Adding extra information to the image model restricts the dimension of the space of admissible solutions. These additional properties are generally expressed as new terms in the case of an optimization problem.

Let us go back to the image model (1.3). In this case, recovering the original image u involves inversing the operator Φ. However, Φ^{-1} is either not continuous or does not always exist. One way to hijack the nonexistence of Φ^{-1} is to approach the unknown image u by solving the following least-square

problem

$$\underset{u\in\mathcal{H}}{\text{argmin}}\|f_\eta - \Phi u\|_{\mathcal{K}}^2.$$

The solution $u^\dagger$ is also called Moore-Penrose generalized solution such that $u^\dagger = \Phi^\dagger f$ where $\Phi^\dagger : \mathcal{K} \to \mathcal{H}$ is continuous if and only if the range of Φ is closed [8]. Nevertheless, in most of the applications in image processing, the range of Φ is not closed which means that the solution of the inverse problems is not stable. Thus, the conditions of Hadamard for well-posedness do not hold anymore.

One important property of ill-posed inverse problems is the instability of the solution. In order to find an approximate and stable one, one way consists in modifying the problem to be solved by regularizing it. The regularization consists in imposing an *a priori* information to the solution which restricts the space of admissible solutions. Let us consider the following regularization scheme

$$\min_{u\in\mathcal{H}}\|\Phi u - f\|_{\mathcal{K}}^2 + \lambda\Theta(u), \tag{1.4}$$

where $\Theta(u)$ is a penalization term that imposes the desired characteristics in the final solution and $\lambda = \lambda(\eta, f)$ is a regularization parameter that acts as a trade-off between the two terms of the regularized problem (1.4). One classic regularization is introduced by Tikhonov where $\Theta(u) = \|u\|^2$ [9, 10] is considered. The Tikhonov regularization problem is strictly convex and the solution u_λ is given by

$$u_\lambda = (\Phi^*\Phi + \lambda I)^{-1}\Phi^* f, \tag{1.5}$$

where Φ^* is the adjoint operator of Φ.

The choice of the regularization term Θ is very important because it deter-

mines the characteristics that the final image should have. Moreover, Θ has to be chosen carefully in order to take into account borders, texture of the image etc. The influence of Θ in the final image is controlled by the regularization parameter λ.

1.3 Remote sensing and image fusion

1.3.1 Remote sensing

In remote sensing, energy emanating from the earth's surface is measured using a sensor mounted on an aircraft or spacecraft platform. In other words, remote sensing consists in the collection of information about an object or a phenomenon without the need of making any contact with it. With remote sensing we can have access to images and gather information about dangerous, inaccessible or large areas which most of the time cannot be done by humankind. Thus, with the collection of these data, better analysis and explanation of many natural phenomena can be provided. Remote sensing can also be carried out in the sea where sensors on board of sonar systems are able to provide cartography of the ocean, which can help study ecosystems and different populations of different species that grow in the depths of the sea without having to disturb them.

In remote sensing, there are two types of sensors: passive and active sensors. Passive sensors capture the reflection emitted by the object due to the exposition to sunlight whereas active sensors emit a signal to the object and capture the reflected signal. Radar and Lidar are examples of active sensors, the time delay between the emission and the reception of the reflected signal helps establish the location, speed and direction of an object. Moreover, the strength of the returned signal describes the physical properties of the remotely sensed objects.

Passive sensors are especially used for land covering and monitoring as well as object identification, whereas active sensors are used for vegetation structure and ground surface elevation study.

Hyperspectral remote sensing, also known as imaging spectroscopy, and multispectral remote sensing are among the most advanced technologies in imaging science and they can be of big help and use in our everyday life and in many other fields. For example, natural phenomena such as deforestation, floods and hurricanes can be monitored, forest fires can be mapped from above and adequate interventions to control them can be organized. With remotely sensed images we can also predict the weather, assess the evolution of erupting volcanoes or follow closely and map the growth of populations, vegetations, urban cities, forests etc.

Unlike RGB cameras that provide images with only three bands, hyperspectral cameras sample the electromagnetic spectrum into narrow and contiguous bands (roughly 200 spectral bands) with an interval of 10 to 20 nm between each two bands. The range of the bands goes from the ultra-violet to radio waves through the visible range and the infrared, whereas RGB cameras sample the visible spectrum only. Each object has its own spectral signature which depends on its physical properties and on how it emits and reflects radiations, thus, the high number of spectral bands allows the distinction between objects throughout the reflectance spectrum of each one of them. This particularity about hyperspectral images has proven very useful in many fields such as in tracking [11], food inspection [12] and also in the study of biodiversity where scientists use hyperspectral images to distinguish between species based on their spectral reflectance [13]. Due to the technical constraints of satellite sensors and onboared storage capacity restrictions, hyperspectral images are quite degraded spatially which sometimes limits their applicability. Regarding the multispectral

sensors, they provide images with much less spectral bands than the hyperspectral sensors (between 3 and 10 bands) from the visible range and the infrared. However, unlike the hyperspectral case, multispectral sensors provide images with a high spatial quality. A high spatial resolution provides fine geometric details that help discern individual objects or features in a captured image. In general, the analysis of the spatial structure is very important in the remote sensing field. For instance, in the case of an urban area that contains roads, office building and parks, having an image with fine geometric details helps recognize and classify different areas with a good accuracy from the captured image.

1.3.2 Satellite sensors constraints

Despite the important progress in the remote sensing technology, satellite sensors remain subjects to many technical constraints. Thus, trade-offs are constantly made by the manufacturers in order to produce relevant data. Therefore, satellite provide data that lack relevant details either spectrally or spatially, depending on the sensor. Two of the determining factors in imaging devices are the SNR (Signal to Noise Ratio) and the GSD (Ground Sampling Distance). The good spatial and spectral quality in satellite images come at the price of a low SNR which means that the images contain noise. Therefore, in order to provide less noisy images with relatively good spectral and spatial resolutions, some compromises should be made. The SNR in decibels is computed as follows

$$\text{SNR}_{\text{dB}} = 10 \log_{10}(\frac{\text{P}_{signal}}{\text{P}_{noise}}),$$

where P_{signal} is the power of the signal and P_{noise} is the power of the noise that affects the signal.

The GSD factor represents the distance between the centers of two consecutive pixels measured on the ground, the bigger the GSD the more spatially degraded the image. For example, a GSD of 20 m means that each pixel of the image represents an area of 20×20 m^2 on the ground, which means that the details and information of the 20×20 m^2 ground scene are "squeezed" into one pixel of the image. Unfortunately, we cannot always choose to have a small GSD and thus have a better spatial resolution without making sacrifices elsewhere such as having a low and bad SNR.

The GSD is linked to the SNR through the NIIRS (National Imagery Interpretability Rating Scale) which is given by the following general image quality equation (GIQE) [14]

$$\text{NIIRS} = c_0 + c_1 \log_{10}(\text{GSD}) + c_2 \log_{10}(\text{RER}) + c_3 \frac{G}{\text{SNR}} + c_4 H, \qquad (1.6)$$

where c_0, c_1, c_2 and c_4 are coefficients relative to the imaging device, RER is the system post-processing relative edge response, G is the system post-processing noise gain and H is the system post processing edge overshoot factor, see [15]. We can notice that GSD and SNR are inversely proportional and therefore trade-offs should be made according to the imaging device and the task the latter is destined to accomplish. Thus, imaging spectrometers provide data with an important GSD and with a SNR that is relatively small. However, the hyperspectral data contains many contiguous spectral bands captured at different wavelengths which is interesting for various applications, namely the study of the composition of the soil, the existence and concentration of gazes in the atmosphere etc. Multispectral sensors provide images with very few spectral channels but with a small GSD which allows data to have a high spatial

quality that can be harnessed for different purposes such as target detection, classification, facial recognition etc. One should bear in mind that the capacity storage and the bandwidth transmission are additional limitations which are taken into consideration. As a consequence, they might affect the choice of the GSD and the SNR factors and thus the quality of the produced images.

1.4 Goals of the thesis and contributions

Remotely embedded imaging systems acquire multi-band images which, in general, can be considered as three-dimensional data cubes. Three variables (x, y, λ) can be used to represent a multi-band image where x and y are the spatial dimensions of the scene and λ is the spectral dimension as illustrated in Figure 1.7. Hyperspectral (HS) images [16] and multispectral (MS) ones [17] are typical examples of multi-bands images.
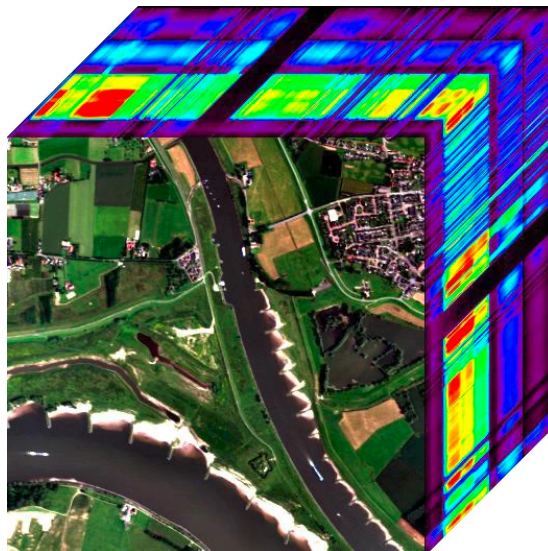
Figure 1.7: Hyperspectral data cube

Due to the technical constraints of imaging satellite sensors, multi-band images suffer from limited spatial or spectral resolution. Thus, the fusion of a

high-spatial and a low-spectral resolution (MS) image, with a low-spatial and a high-spectral resolution (HS) image, has emerged in order hijack the limited resolutions in satellite sensors. Satellite image fusion has been explored for many years and it is still a challenging and active field of research [18, 19, 20]. Figure 1.8 shows an example of a high-spectral but low-spatial resolution image, a high-spatial but low-spectral resolution and a high-spatial and high-spectral resolution images. Pansharpening was the first fusion technique and it consists of fusing a high-spatial resolution panchromatic (Pan) image and a low-spatial resolution (MS) image (MS pansharpening) [21, 22, 23, 24]. The problem of fusing hyperspectral and panchromatic images (HS pansharpening) has also been explored [20, 25, 26].

The fusion of multispectral and hyperspectral images (HS-MS fusion) has been studied in the literature, however, it is still a challenging problem due to the high dimensionality of images. In fact, in HS-MS fusion, the hyperspectral data has many spectral bands and the multispectral data contains a rich spatial information which in total, makes a higher quantity of data than in MS or HS pansharpening. Thus, the objective of this thesis is to address the HS-MS fusion. For this matter, the suggested methods take into account the high dimensionality of data and combine the rich spectral information contained in the hyperspectral image and the spatial information contained in the multispectral image in order to infer a high resolution hyperspectral image.

1.4.1 Image fusion and problem formulation

In order to address the HS-MS fusion problem, it is important to specify the different notations needed for its formulation. The target image, also called the unknown, fused or high resolution hyperspectral image $\mathbf{u} = (u_1, \ldots, u_H)^\mathsf{T} \in \mathbb{R}^{H \times N}$, is rearranged in a $H \times N$ matrix where H is the number of spectral bands and N is

Figure 1.8: (Left) Hyperspectral image (size: 99×46×224). (Middle) Panchromatic image (size: 396×184). (Right) Target image (size: 396×184×224)

the number of overall pixels of **u**. This matrix representation of the target image makes the distinction between spatial and spectral dimensions clear. The way **u** is rearranged in the matrix means that each row $u_i \in \mathbb{R}^N$ is a one-dimensional vector that contains the pixels of the ith spectral channel with $i \in \{1, \ldots, H\}$. The observed images are generally considered as linear degradations of the target high-spatial and high-spectral image. Thus, the HS-MS fusion can be seen as restoring a three dimensional data-cube from two degraded data-cubes. The linear formulation of the observed images allows the fusion problem to be easily formulated.

The multispectral image referred to as **f** such that $\mathbf{f} = (f_1, \ldots, f_M)^\intercal \in \mathbb{R}^{M \times N}$ with M bands and N pixels is supposed to be a spectrally degraded and noisy version of the target image **u**. On the contrary, the hyperspectral image denoted as **g** such that $\mathbf{g} = (g_1, \ldots, g_H)^\intercal \in \mathbb{R}^{H \times N_l}$, is a blurred, down-sampled and noisy version of **u** with H bands, where $H >> M$ and $N_l = \frac{N}{l^2}$ pixels where l is the

spatial sub-sampling factor. The relationship between $\mathbf{u}$, $\mathbf{f}$ and $\mathbf{g}$ is given by the following common observation model [27]

$$
\begin{aligned}
g_h &= u_h BD + \varepsilon_h, \ \forall h \in \{1, \ldots, H\}, \\
f_m &= (Su)_m + \varepsilon_m, \ \forall m \in \{1, \ldots, M\},
\end{aligned}
\tag{1.7}
$$

where u_h represent the hth band of the unknown full resolution hyperspectral image $\mathbf{u}$. The operator $B \in \mathbb{R}^{N \times N}$ is a low-pass filter that acts on each band and it models the point spread function of the hyperspectral sensor. The operator $D \in \mathbb{R}^{N \times N_l}$ is a down-sampling operator with the down-sampling factor denoted by l. For the spectral degradation step, the operator $S \in \mathbb{R}^{M \times H}$ is used as the spectral response of the multispectral sensor. The two terms, $\varepsilon_h \in \mathbb{R}^{N_l}$ and $\varepsilon_m \in \mathbb{R}^N$, are the realization of i.i.d. zero-mean band-dependent Gaussian noise. As to h and m, they denote the number of the spectral band in $\mathbf{g}$ and $\mathbf{f}$, respectively.

1.4.2 Contributions of the thesis

In this thesis, several algorithms for hyperspectral and multispectral image fusion were suggested. The contributions of this work are described in two main chapters, each chapter is divided into two components that describe each a new fusion model. Some notations might be used in many places. In oder to avoid confusion, in each chapter the notations are introduced and specified before use.

In Chapter 2, we study the image fusion problem with tools from the optimal transport theory. One of these important tools is the regularised Wasserstein distance defined as follows

$$W_\gamma(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} < \pi, D > -\gamma E(\pi),$$

where μ and ν are two probability measures, D is a cost matrix, π is the transport plan, E is the entropy and γ is the regularization coefficient linked to the entropy.

The entropic energy used to regularize the Wasserstein distance is defined as follows

$$E(\pi) = -\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{i,j}(\log \pi_{i,j} - 1) - \iota_{\mathbb{R}^+}(\pi_{i,j}),$$

where ι is the indicator of $\mathbb{R}_+$ such that

$$\forall x, \quad \iota_{\mathbb{R}_+}(x) = \begin{cases} 0 & \text{if} \ \ x \in \mathbb{R}_+, \\ +\infty & \text{otherwise.} \end{cases}$$

The fusion problem is thus modeled by the minimization of the sum of two regularized Wasserstein distances as follows

$$\min_{\mathbf{u} \in \Sigma_u} \left\{ G(\mathbf{u}) = \lambda W_{\gamma_M}(\mathbf{u}, \widetilde{\mathbf{S}}(\mathbf{f})) + (1 - \lambda) W_{\gamma_H}(\mathbf{u}, \widetilde{\mathbf{T}}(\mathbf{g})) \right\}, \tag{1.8}$$

Where $\mathbf{u}$ is the fused image to be recovered, W_{γ_M} and W_{γ_H} are regularized Wasserstein distances, with regularization coefficients γ_M and γ_H, related to the multispectral and hyperspectral components, respectively. Furthermore, the operators $\widetilde{\mathbf{S}}$ and $\widetilde{\mathbf{T}}$ are inversion ones, λ is a trade-off parameter and Σ_u is the simplex where $\mathbf{u}$ is defined.

The resolution of the minimization problem (1.8) is carried out with

Sinkhorn's algorithm. The latter allowed the use of matrix-vector operations which sped up the computations. Figure 1.9 shows the visual results of the proposed fusion method. We notice that, visually, the fused image looks similar to the reference image and that the main geometric and spectral details were recovered throughout the fusion process. In Section 2.8.2, a detailed experimental study of the fusion method is provided. It is also compared with other state-of-the-art methods visually and also with objective quality indices. The comparison showed that the results of our proposed algorithm compare favorably to the state-of-the-art methods.

Figure 1.9: Performances of HS-MS fusion with optimal transport on Pavia dataset $256 \times 256 \times 93$. From top left to right. Reference image, HS image, MS image and the result of the proposed method.

In Section 2.9, an upgrade of the fusion model was suggested. For the new fusion model the novelty was in the cost matrix which in the previous fusion model was defined as follows

$$\xi = e^{-\frac{D}{\gamma}}$$

where D is the initial cost matrix that contains the euclidean distances between the pixels and γ is the regularization coefficient.

In a modification of the proposed fusion model, a nonlocal means (NLM) term that describes the relationship between patches of the image is added. The goal of adding this term is to include the values of the pixels in the transport, to account for the noise in the images and the misregistration that might affect data during acquisition. For two pixels i and j with the three dimensional coordinates (x_i, y_i, z_j) and (x_j, y_j, z_j) respectively, the modified cost matrix is defined as follows

$$\xi(i,j) = \exp\left(-\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) - \frac{\alpha}{\gamma}(z_i - z_j)^2 - \frac{\|P(i) - P(j)\|_2^2}{h_{sim}^2}\right),$$

where α is a coefficient that balances the spatial distance with respect to the spectral one, P denotes the patch extracted from the image and h_{sim} is a filtering parameter. The difficulty of this method is the high number of parameters related to the fusion model which in total is equal to eight parameters. A study of sensitivity of these parameters showed that the method did not improve the results of the previous fusion model. Nevertheless, the NLM weights might contribute to downplaying the effect of misalignment that affects data during the acquisition process. This new hypothesis will be explored in future work.

In Chapter 3, a fusion model based on the nonlocal gradient with two variants is suggested. The first fusion model, presented in Section 3.2, is expressed as

the minimization problem as follows

$$\min_{u\in\mathbb{R}^{H\times N}} \sum_{h=1}^{H} \|\nabla_{\omega_h} u_h\|_1 + \frac{\mu}{2}\sum_{h=1}^{H}\|DBu_h - g_h\|_2^2 +$$

$$\frac{\gamma}{2}\sum_{m=1}^{M}\|(Su)_m - f_m\|_2^2 + \frac{\lambda}{2}\sum_{h=1}^{H}\|\tilde{P}_h u_h - P_h\tilde{g}_h\|_2^2. \tag{1.9}$$

The term $\|\nabla_{\omega_h} u_h\|_1$ is a regularizer defined as follows

$$\|\nabla_{\omega_h} u_h\|_1 = \sum_{i}\left|\nabla_{\omega_h} u_{h,i}\right|,$$

where $|\cdot|$ denotes the L^2-norm in this case and $\nabla_{\omega_h} u_h \in \mathbb{R}^{N\times N}$ is the nonlocal gradient computed for each pixel such as $\nabla_{\omega_h} u_{h,i} = \left((\nabla_{\omega_h} u_{h,i})_1, \ldots, (\nabla_{\omega_h} u_{h,i})_N\right)$ and it is defined as follows

$$(\nabla_{\omega_h} u_{h,i})_j = \sqrt{\omega_{h,i,j}}\left(u_{h,j} - u_{h,i}\right),$$

where $\omega_{h,i,j}$ is a similarity measure defined as follows

$$\omega_{h,i,j} = \begin{cases} \dfrac{1}{\Gamma_i}\exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{m=1}^{M} s_{m,h}\sum_{\{t\in\mathbb{Z}^2:\|t\|_\infty\leqslant\nu_c\}}\|f_{m,i+t} - f_{m,j+t}\|_2^2}{h_{sim}^2(2\nu_c + 1)^2\sum_{m=1}^{M} s_{m,h}}\right) & \text{if} \\ & \|i-j\|_\infty\leqslant\nu_r, \\ & (i\neq j) \\ 0 & \text{else,} \end{cases}$$

and

$$\Gamma_i = \sum_{\{j:\|i-j\|_\infty\leqslant\nu_r\}}\exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{m=1}^{M} s_{m,h}\sum_{\{t\in\mathbb{Z}^2:\|t\|_\infty\leqslant\nu_c\}}\|f_{m,i+t} - f_{m,j+t}\|_2^2}{h_{sim}^2(2\nu_c + 1)^2\sum_{m=1}^{M} s_{m,h}}\right)$$

$$\omega_{h,i,i} = \max\left\{\omega_{h,i,j} : \|i-j\|_\infty\leqslant\nu_r \text{ and } j\neq i\right\}$$

.

Unlike the classical gradient that interacts locally with the pixels, the nonlocal one allows the computation of the gradient between any two pixels in the image.

The terms $\|DBu_h - g_h\|_2^2$ and $\|(Su)_m - f_m\|_2^2$ of expression (1.9) are data fidelity ones related to the generation of the hyperspectral and the multispectral image. The operator B is a low-pass filter, D is a downsampling operator and S represents the spectral response of the multispectral sensor. The goal of data fitting terms is the penalization of the deviation from the hyperspectral and the multispectral data g_h and f_m, respectively. The penalization is done throughout the L^2-norm.

The last term $\|\tilde{P}_h u_h - P_h \tilde{g}_h\|_2^2$ of expression (1.9) is a radiometric constraint penalized with the L^2-norm. The goal of this constraint is to force the fused image $\mathbf{u}$ to share the same high spatial frequencies as the multispectral image $\mathbf{f}$, that is,

$$\frac{u_h}{P_h} = \frac{\tilde{g}_h}{\tilde{P}_h}, \ \forall h \in \{1, \dots, H\}, \tag{1.11}$$

which is equivalent to

$$u_h - \tilde{g}_h = \frac{\tilde{g}_h}{\tilde{P}_h}\left(P_h - \tilde{P}_h\right).$$

Note that, we are forcing the high frequencies of each hyperspectral band, $u_h - \tilde{g}_h$, to coincide with those of the multispectral image, $P_h - \tilde{P}_h$, so that spatial details are injected into the fused product. The modulation coefficient $\frac{\tilde{g}_h}{\tilde{P}_h}$ takes the energy levels of the multispectral image and of the corresponding hyperspectral bands into account. The different components of the radiometric constraints are defined as follows

- $P \in \mathbb{R}^{H \times N}$ is a linear combination of the multispectral image such that,

$$P_h = \sum_{m=1}^{M} \alpha_{m,h} f_m, \ \forall h \in \{1, \ldots, H\}. \tag{1.12}$$

- $\tilde{P} \in \mathbb{R}^{H \times N}$ is computed as follows

$$\tilde{P}_h = \sum_{m=1}^{M} \alpha_{m,h} \tilde{f}_m, \ \forall h \in \{1, \ldots, H\}. \tag{1.13}$$

where $\tilde{f}$ is the multispectral image spatially degraded, as in the generation model (1.7), and then brought back to the grid of the multispectral image with bicubic interpolation. This manipulation kills the high spatial frequencies and leaves only the low spatial ones

- $\tilde{g} \in \mathbb{R}^{H \times N}$ is the hyperspectral image brought to the fine grid of the fused image by bicubic interpolation.

The weights $\alpha_{m,h}$ are derived as

$$\alpha_{m,h} = \frac{s_{m,h}}{\sum_m s_{m,h}},$$

where $s_{m,h}$ are the coefficients extracted from the spectral response of the multispectral sensor.

The radiometric constraint is detailed further in the document. Figure 1.10 shows the visual results of the proposed method. We notice that the spectral information as well as the spatial details were recovered and look similar to the reference image. Quantitative results show that the suggested method compares favorably with other state-of-the-art ones.

In Section 3.7, a modification of the variational model (1.9) for hyperspectral and multispectral image fusion was suggested. The new model is presented as

Figure 1.10: Performances of HS-MS fusion with the nonlocal gadient based method on Chikusei dataset $304 \times 304 \times 93$. From top left to right. Reference image, HS image, MS image and the result of the proposed method.

follows

$$\min_{u \in \mathbb{R}^{H \times N}} \sum_{h=1}^{H} \|\nabla_{\omega_h'} u_h\|_1 + \frac{\mu}{2} \sum_{h=1}^{H} \|DBu_h - g_h\|_2^2 +$$

$$\frac{\gamma}{2} \sum_{m=1}^{M} \|(Su)_m - f_m\|_2^2 + \frac{\lambda}{2} \sum_{h=1}^{H} \|\tilde{P}_h u_h - P_h \tilde{g}_h\|_1,$$

where the new similarity weights are computed as follows

$$\omega_{h,i,j}' = \begin{cases} \dfrac{1}{\Gamma_i'} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|\tilde{g}_{h,i+t} - \tilde{g}_{h,j+t}\|_2^2}{h_{sim}^2 (2\nu_c + 1)^2} \right) & \text{if } \|i-j\|_\infty \leqslant \nu_r \ (i \neq j) \\ 0 & \text{else} \end{cases}$$

and

$$\Gamma_i' = \sum_{\{j : \|i-j\|_\infty \leqslant \nu_r\}} \exp\left(-\frac{\|i-j\|_2^2}{h_{spt}^2} - \frac{\sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|\tilde{g}_{h,i+t} - \tilde{g}_{h,j+t}\|_2^2}{h_{sim}^2 (2\nu_c + 1)^2} \right)$$

where

$$\omega_{h,i,i}' = \max\left\{ \omega_{h,i,j} : \|i-j\|_\infty \leqslant \nu_r \text{ and } j \neq i \right\}.$$

The terms implicated in the radiometric constraint P, $\tilde{P}$ and $\tilde{g}$ were the results of transformations of the initial data as it was shown above. Thus, a new noise could have been introduced throughout these transformations. The L^2-norm is efficient in the case of Gaussian noise but not in other types of noise. Therefore, the L^2-norm was maybe not well adapted enough to the kind of noise fabricated during the data transformation process and thus not adapted for the minimization of $\tilde{P}_h u_h - P_h \tilde{g}_h$. Hence, in the new model the radiometric constraint was penalized with the L^1-norm.

In this part, the experimentations were carried out on images acquired by hyperspectral cameras of indoor and outdoor scenes. Figure 1.11 shows the result of the fusion. We notice that the fused image is visually close to the reference image and that the spatial details as well as the spectral ones are recovered throughout the fusion process. The analysis of the quantitative results shows that the modified fusion model compares favorably to the state-of-the-art methods and overcomes the results of the first model in some quality indices.

1.5   State of the art

Image fusion consists of combining relevant information from different available images into one single fused image. The latter is more informative and complete than any of the input images. Moreover, the fused image is supposed to have superior quality and it is not presumed to contain any details that are non-existent in the input images. There are many types of image fusion. In remote sensing and in astronomy, multisensor fusion is used to achieve high spatial and spectral resolutions by combining images from two sensors, one with a high spatial resolution and a low spectral resolution and the other one with a high spectral resolution and a low spatial resolution [28]. In the case of
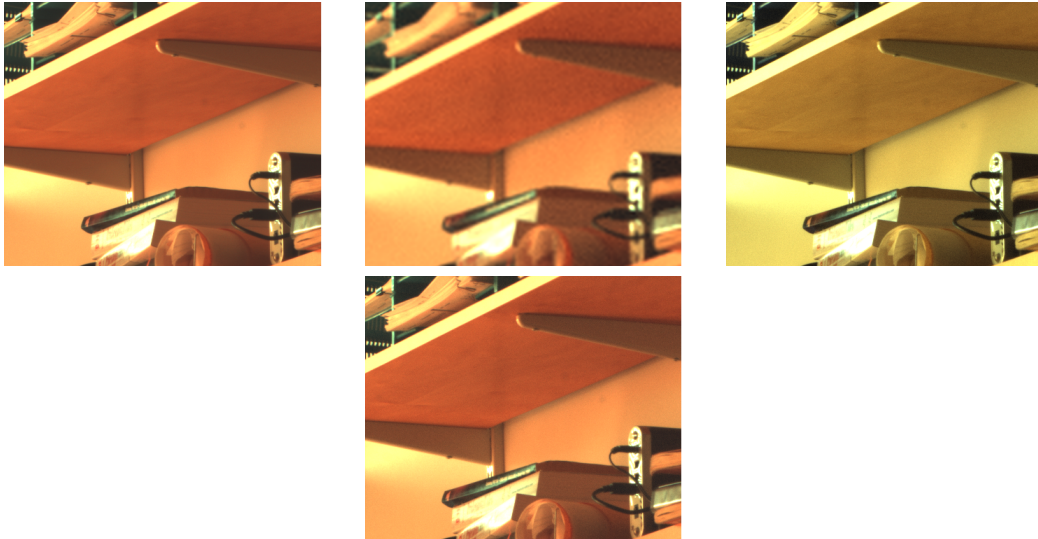
Figure 1.11: Performance of HS-MS fusion, on the image "Bookshelves" of size $512 \times 512 \times 31$, with the nonlocal-gradient based method in the case of the penalization with the L^1 norm of the radiometric constraint. From top left to right. Reference image, HS image, MS image and the suggested method.

multi-view fusion, a set of images of the same scene taken by the same sensor, but from different viewpoints, is fused in order to obtain an image with higher resolution and recover the 3D representation of the scene [29]. One other type of image fusion is the multitemporal one [30]. In the latter, images of the same scene are acquired at different times. This is helpful for finding and evaluating changes in the studied scene, which is quite useful in medical imaging especially when it comes to change detection of organs and tumors. Also, the aim of multitemporal fusion is to obtain a less degraded image of the scene, which is very used in remote sensing for land monitoring or forests exploitation over months or years. Each fusion method is different following the purpose of the fusion and the characteristics of the sensors and the images to be fused. Multi-focus image fusion is another type of merging different information in different images. Its goal is to combine images that focus on different objects or areas. Thus, the fused image has better focus across all the captured scene than

any other input image [31].

This thesis is focused on fusing different resolutions of images that come from different remote sensing sensors. Satellite image fusion has been an active field of research these last years due to the growing availability of data and to the need of gathering information from different imaging sources. In fact, there are many ongoing and upcoming hyperspectral missions that help shed more light about many Earth phenomena [32, 33, 34, 35, 36, 37]. As a matter of fact, satellite sensors with a good GSD (Ground Sampling Distance) factor provide images with a good spatial quality and a good description of the geometric details. Satellites such as Aviris, Spot, Pléiades etc capture images with a high number of spectral bands which provides a good color description of the captured scene. However, due to technical constraints of imaging sensors onboard, Earth observation satellites are always subjects to compromises such as to the one between the SNR and the GSD factors. These tradeoffs lead to acquiring an image with good spatial details and low spectral ones or to acquiring an image with good spectral details and low spatial ones. The lack of relevant spatial or spectral details is sometimes not convenient when good spatial and spectral information is needed on the same image. Table 1.1 shows the spatial and spectral resolutions of some satellite sensors.

Name	AVIRIS (HS)	SPOT-5 (MS)	Pleiades (MS)	WorldView-3 (MS)
Res. (m)	20	10	2	1.24
Nb. bands	224	4	4	8

Table 1.1: Characteristics of some satellite sensors

Pansharpening has been the first case of satellite image fusion and its goal is to merge the high spatial resolution of the panchromatic image and the high spectral resolution of the multispectral image in order to provide one high-resolution image called the pansharpened image. A panchromatic image has

one single spectral band and a high spatial resolution. Both the panchromatic and the multispectral image must cover the same scene. In the ideal case, the spectral range covered by the panchromatic and the multispectral image has to be exactly the same in order to avoid undesired outcomes on the pansharpened image such as color distortion.

Pansharpening methods have been thoroughly explored in the literature and they can be divided into three main classes: Component Substitution (CS), Multiresolution analysis (MRA) and Bayesian methods. The CS methods rely on the transformation of data by using different techniques such as the Itensity Hue Saturation (IHS) method [38, 39, 40], which is another representation of the RGB colors that approaches the human eye color visualization system. In CS methods we also find PCA (Principal Component Analysis) [41, 42] which is a technique that reduces the dimension of data by converting the correlated data into non-correlated one. In the category of CS methods, we also find the Brovey approach which is a pixel-level image fusion technique that includes a RGB color transformation method, moreover, it is effective for merging data that come from different sensors [43, 44]. The Gram-Schmidt orthonormalization [45, 46] and partial replacement adaptive CS [47] are also part of the CS family.

The MRA methods are based on multiresolution decomposition techniques. In this group of methods, we find wavelet-based methods [48, 49, 50, 51], undecimated wavelet transform [52],"à-trous" wavelet transform (ATWT) [53], nonseparable transforms based on wavelets [54] and other transforms that are not based on wavelets [55]. We also find in the MRA family methods that are based on Laplacian pyramid [56, 57, 58], which is a technique based on multi-scale decomposition of an image that undergoes repeated smoothing and subsampling. The difference between different MRA methods resides in the way spatial details are extracted.

The Bayesian approach is related to the use of posterior distribution of the unknown high resolution image with respect to the available observed data. The posterior distribution depends on the likelihood function which is the probability of the observed data given the unknown high resolution image. It also depends on the choice of the prior, because, this latter describes the desired properties in the target image [59, 60, 61, 62] . Choosing a good prior is also crucial because it accounts for the ill-posedness of inverse pansharpening problems which are considered to be a special case of the Bayesian family [63].

CS methods are known for providing fusion results with good spatial quality that preserves the fidelity in the geometric details [46]. However, most of the time, the fusion results of CS methods suffer from spectral distortion which is due to the fact that the panchromatic and the multispectral sensors do not always cover exactly the same spectral domain. MRA methods on the other hand succeed at preserving the spectral details in the fused image, whereas the spatial resolution suffers from problems such as spatial distortion, ringing and staircasing artifacts [19, 64, 65] . However, the spatial quality of the MRA methods can be of a good quality and even compare to the one of the CS family results when the low-pass function used in the multi-scale decomposition matches the MTF (Modulation Transfer Function) of the spectral channels.

Most pansharpening techniques assume that the panchromatic image is a linear combination of the spectral bands of some reference image. The other assumption is about the geometric alignment of the spectral bands of the MS image. The last assumption is not always true because sensors in reality introduce some shifts between the bands that make them not co-registered. In [66], the authors studied the performances of different pansharpening methods with mis-registered data and they came to the conclusion that, when data is not co-registered, CS methods perform better than the MRA ones. In [67], a

pansharpening variational model without any assumptions on the spectral data is suggested. Instead of the linearity assumption, they introduced a radiometric constraint that injects the high spatial frequencies of the panchromatic image into each spectral band of the multispectral image. The minimization problem proposed by the authors for pansharpening is carried out on each channel which deals better with mis-registered and aliased spectral data. For a long time in pansharpening problems, multispectral images were used for their spectral quality, the authors in [20] did a review where they presented hyperspectral pansharpening methods where the spectral information is extracted from the hyperspectral image and not from the multispectral one. For more details about the pansharpening methods, see the papers [65, 67].

With the growing number of satellite data and the availability of hyperspectral and multispectral images covering the same scene and the same spectral domain, many methods throughout the literature adapted pansharpening techniques for HS-MS fusion. The pansharpening techniques presented in [68, 69] were among the first to have been adapted to HS-MS fusion. The performance of both methods depend on the way multispectral data is interpolated in the spectral domain. Chen *et al*. [70] adapted a pansharpening method for HS-MS fusion that consists in dividing the spectrum of the hyperspectral image into many regions. In each region, the HS-MS fusion is carried out with a chosen pansharpening algorithm. In [71], we find a pansharpening-based model which assumes that the value of every pixel of an optical image depends on two factors only: the solar radiation and the spectral reflectance of the land surface. The fused image is thus found by multiplying the hyperspectral image by a ratio computed between a high-resolution image and its low-pass filtered version. The technique presented in [72], which is one of the CS family, is another pansharpening model that was adapted for HS-MS fusion. In this technique, an

intensity component is computed as a weighted combination of the multispectral bands. Moreover, in order to minimise the effect of the spatial distortion, the used weights are computed as regression coefficients between the bands of a high-resolution image and a spatially degraded one. One way to account for missing spectral information in the multispectral image is the resampling of the latter in the spectral domain using RIBSR (Ratio Image-Based Spectral Resampling). Following this HS-MS fusion framework, most pansharpening algorithms can be used. Another framework was suggested by Selva *et al.* called *hypersharpening* where each hyperspectral band is synthesized as a linear combination of multispectral bands in order to produce a high-resolution image. The *hypersharpening* framework proved effective in adapting pansharpening to HS-MS fusion and it was tested on a MRA-based pansharpening method that uses GLP [73]. Eismann *et al.* developed a Bayesian fusion method based on a maximum a posteriori (MAP) estimation and a stochastic mixing model (SMM) of the spectral scene. The aim of these estimations was to develop a cost function that optimizes the estimated target image based on the hyperspectral and the multispectral data [74, 75]. This Bayesian method proved very effective in terms of upgrading the spatial resolution of each hyperspectral band. The MAP and SMM estimations were carried out in the principal component subspace. Using Bayesian estimation and carrying out optimization in a subspace was used in many examples throughout the literature [61, 62, 76].

Many authors used spectral unmixing techniques to carry out HS-MS fusion task. In hyperspectral imaging, each pixel is assumed to be a mixture of different distinct materials (vegetation, soil, water etc) and each one of these materials is present with a certain proportion in the pixel. Thus, in unmixing techniques, materials and their proportions are separated into two different matrices called endmember and abundance matrices. Zhukov *et al.* suggested

a fusion technique where the low-resolution image is unmixed by using the information about the pixel composition from the high-resolution image after classifying the latter with an unsupervised algorithm. The fused image is then induced by sharpening the low-resolution image by assigning the estimated endmembers to the corresponding high-resolution pixels of the classification map [77]. Gross and Schott suggested an unmixing-based method for fusion. The method starts by an unmixing step which gives as an output the endmember fraction images (maps) and then constrained optimization techniques are used to generate a high-resolution image [78].

Using unmixing for fusion purposes has been very used during recent years. It consists in obtaining endmember and abundance matrices from hyperspectral and multispectral data, finally the fused image is the product of both matrices. Berné *et al.* suggested a HS-MS method based on the decomposition of the low-resolution hyperspectral data using Non-Negative Matrix Factorization (NMF) and then non-negative least square regression was used to estimate the abundance-maps from the high-resolution multispectral data. This method was originally destined for applications on mid-infrared data in astronomy [79]. Yokoya *et al.* [80] proposed a fusion method where hyperspectral and multispectral data are alternately unmixed to extract the endmember spectra and the abundance matrices. Finally the fused image is induced by multiplying the two extracted matrices. An unmixing-based fusion technique was suggested by Kawakami *et al.* [81] where spatial sparsity of the hyperspectral data was harnessed and only a few materials were assumed to constitute the composition of a pixel in the hyperspectral image. The latter is factorized into a basis and the sparse coefficients are found throughout a L^1-norm minimization.

Wycoff *et al.* [82] proposed a fusion technique based on the spatial sparsity

of the hyperspectral data. In Lanaras *et al.* a linear mixing model (LMM) was used and hyperspectral and multispectral images were jointly unmixed into endmember and abundance matrices. The fusion problem was solved with a projected gradient scheme [83]. Akhtar *et al.* suggested a fusion scheme based on dictionary learning which was used for the estimation of abundance and endmember matrices [84].

Wei *et al.* [62, 85] proposed a fusion approach based on a variational model. The fusion problem is an ill-posed one and in order to account for the ill-posedness a sparse regularization term was added and it was determined based on the decomposition of the scene on a set of dictionaries. The same authors proposed a Sylvester equation-based fusion model named FUSE (Fast fUsion base on Sylvester Equation) which can be easily generalized and allow for the use of Bayesian estimators. The fusion problem was solved by alternating direction method of multipliers coupled with the block coordinate descent method. This model achieved the same performance as the existing Bayesian fusion frameworks but it proved very effective in terms of reducing the computational time [86]. Simoes *et al.* [61] suggested a total variation-based fusion model. To tackle the ill-posedness of the problem, a non-smooth variational regularizer was added. The resolution of the fusion problem was carried out with SALSA algorithm.

Deep learning has been enjoying a huge success in many applications in computer vision including in HS-MS fusion and in super-resolution. The latter, unlike in HS-MS fusion, aims at inferring a high-resolution image from a single low-resolution image only and not two low-resolution images. Convolutional neural networks (CNN) are among the deep learning models and they consist of an input and an output layer along with hidden connected layers where convolution based operations take place. CNN-based methods for HS-MS

fusion have been suggested in [87, 88]. Also, many super-resolution fusion methods based on CNN have been proposed [89, 90, 91, 92, 93]. The use of deep learning methods in this kind of fusion is relatively recent.

Chapter 2

Optimal transport

Summary

Satellite sensors acquire either hyperspectral images with a low spatial resolution and a high spectral one or multispectral images with high spatial resolution and low spectral one. The aim of HS-MS fusion is to fuse both images in order to infer a high resolution image with a good spatial and spectral resolution. In this chapter, we introduce two HS-MS fusion models based on optimal transport. Both models are the sum of two weighted regularized Wasserstein distances which describe the transport of mass in the spectral and the spatial domain. Both models assume that the low-spatial resolution bands in the hyperspectral image are obtained from the high-resolution image by low-pass filtering followed by subsampling. We also assume that the multispectral image is a spectral degradation of the high-resolution one, and that the spectral response of the multispectral sensor is known.

The first model outperforms state-of-the-art techniques on remote sensing data. Ground truth data is also available which allowed evaluation of the results with objective metrics. In the second model, the information on the similarity of

pixels was included in the cost function. Due to the high number of parameters of the second model, a study of the sensitivity of parameters was carried out on multiple data sets. The results showed that the performance of the second model did as good as the first one but did not outperform it.

This chapter is organized as follows. In Section 2.1 we review a few applications of optimal transport in the literature especially in the image processing field. Section 2.2 is an introduction to optimal transport, it explains the principle of the latter and reviews the different methods used to solve it. Section 2.3 presents a detailed study of discrete optimal transport and sets the framework for the formulation of the HS-MS fusion. In Section 2.4, the HS-MS fusion is formalized and presented as a barycenter problem in the sense of optimal transport. Each part of the fusion model is commented and analyzed. The subject of Section 2.5 is to analyze the cost matrix chosen for the fusion model and to study relevant ways for its computation. Section 2.6 summarizes the fusion algorithm. In Section 2.7, different quantitative methods for evaluating the fusion results are presented and detailed. In Section 2.8, the results of the first fusion method are analyzed. The visual results showed that the suggested method provided an inferred image close to the reference one. As to the quantitative part, the results showed that the proposed model outperforms state-of-the-art methods. The second part of this chapter is presented in Section 2.9. In this latter a modification of the HS-MS fusion model is presented. The novelty in this part is the mixing of nonlocal weights and the position of pixels in the cost matrix. Section 2.10 presents the experiments of the second fusion method. The experimental part shows that the new suggested model performs as good as the first one and that the adding of the nonlocal weights did not improve the previous results. Finally, Section 2.11 presents a conclusion of the whole chapter and summarizes the approaches and their related conclusions.

2.1 Optimal transport and image processing applications

Optimal transport goes back to the 18th century where originally it has emerged to tackle a land leveling problem and it has been enjoying a great success over the last years in many applied fields. In the literature, optimal transport has been used as a powerful tool due to its interesting properties in the image processing field such as in restoration [94], color manipulation [95, 96, 97], reflectance interpolation [98] and denoising [99]. In statistical learning, optimal transport has proved efficient in the manipulation of histograms [100], in image retrieval in computer vision [101], in semi-supervised learning [102] and in domain adaptation [103].

One successful use of optimal transport in image processing is color transfer [104], where the goal is to transfer the colors of a target image to an original image while preserving the geometry of the latter as shown in Figure 2.1. The transfer of the colors is carried out in the color space with optimal transport tools. The result image kept exactly the same geometry as the original one except that now, it has new colors which are the ones of the target image.

Figure 2.1: (Left) The original image. (Middle) The target image with the color information of interest. (Right) The result with the geometry of the original image and the colors of the target image, see [104].

Another application of optimal transport is shape interpolation [105]. Figure 2.2 shows a 3D interpolation of three geometrical objects: a torus, a duck

and a hippopotamus, each with an interpolation coefficient β_i, $i \in \{1,2,3\}$, and $\beta_i \in [0,1]$. We can notice that for each new interpolation coefficients $(\beta_1, \beta_2, \beta_3)$, the new object has its own geometrical shape. Thus, we can conclude that optimal transport is efficient for providing results with pleasant visual geometrical properties.
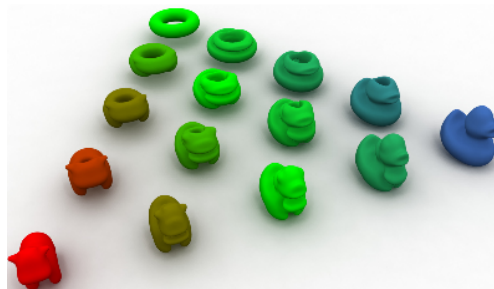
Figure 2.2: A 3D interpolation between a torus, a duck and a hippopotamus, see [105].

From the two examples in Figure 2.1 and Figure 2.2, we can notice that optimal transport has the powerful quality of preserving the geometrical shape of objects and creating objects with new geometrical properties. In this chapter, we suggest to harness the interesting properties of optimal transport and apply them in the case of HS-MS fusion. The idea behind applying optimal transport is to carry out a color transfer from the hyperspectral image while preserving the geometrical details of the multispectral image in order to infer one high-resolution image. To our best knowledge, before the present work, optimal transport has never been applied in the case of HS-MS fusion.

2.2 Introduction to optimal transport

The theory of optimal transport [106], or transportation theory, started a few years before the French revolution with the French mathematician Gaspard Monge in 1781 [107]. The latter submitted a memoir called "*Théorie des déblais*

et des remblais" to the "*Académie des Sciences*" in which he exposed the question of how to move a sand pile from a starting point A to an arrival point B. In mathematical notations, Monge's problem can be presented as follows: let X and Y be two Polish spaces, we denote by $\mathscr{P}(X)$ and $\mathscr{P}(Y)$ the spaces of probability measures on X and Y, respectively. Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ be two probability measures and $c : X \times Y \rightarrow [0, +\infty]$ a cost function. Then, the following problem arises

$$\min_{\mathsf{T}} \left\{ \int_X c(x, \mathsf{T}(x)) \, d\mu(x) : \mathsf{T}_{\#}\mu = \nu \in \mathscr{P}(Y) \right\}, \tag{2.1}$$

where $\mathsf{T} : X \rightarrow Y$ is a map and the measure $\mathsf{T}_{\#}\mu \in \mathscr{P}(Y)$ is called the *pushforward* of μ through T and it is defined by $\mathsf{T}_{\#}\mu(E) = \mu(\mathsf{T}^{-1}(E)), \forall E \in Y$. The probability measures μ and ν describe each a different distribution of mass and c describes the cost payed for moving a unit mass located at $x \in X$ to $y \in Y$. Finding the optimal map allows an efficient transportation of mass from μ to ν that guarantees an overall minimum cost.

To illustrate the main idea of how optimal transport works, let us consider the following example of hot croissants distribution from bakeries to the cafés in Vannes city. As one can see in Figure 2.3, the distribution of bakeries is different from the one of cafés and there are $4! = 24$ possible configurations to move from one distribution to the other. The goal is to supply the cafés with the croissants with the minimum cost of transportation.

There are many possibilities for choosing the transportation cost, once it is fixed, the transport has to be optimized based on it. For example, if the chosen cost to move croissants from one bakery to one café is the distance between these latter, then the cost of transporting the croissants should be the minimum of the sum of costs of all possible distances. Figures 2.4 and 2.5 show two configurations of transport from bakeries to cafés. The way of transporting

Figure 2.3: Illustration of the principle of optimal transport

croissants following the configuration in Figure 2.4 is less expensive, in terms of the sum of distances, than the one in Figure 2.5. In fact, the sum of distances during transport in Figure 2.4 is 23.8 cm whereas the sum of distances in Figure 2.5 is 25.7 cm.
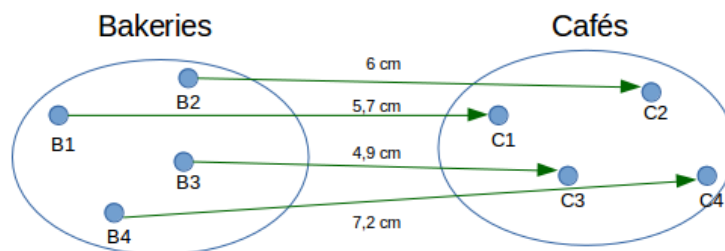
Figure 2.4: Possible optimal configuration
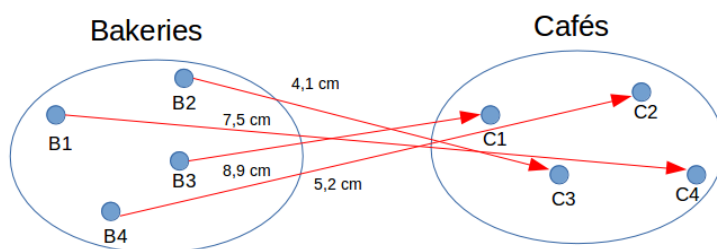
Figure 2.5: Non-optimal configuration

Thus formulated in the example, Monge's problem is a combinatorial one and it is the first operational search problem in which one seeks to have the best solution among multiple choices. In general, the problem of Monge is

not easy to solve due to its non-linearity and to the difficulty of proving the existence of a minimizer and studying it. Monge did not solve the optimal transport problem and the issue about the existence of a minimizer was not even addressed [108]. Years after Monge initiated the optimal transport problem, the Russian mathematician Kantorovich re-studied it and contributed to its modernization by making the connection between Monge's problem and linear programming. The work of Kantorovich made optimal transport problems more addressable where closed form solutions can be found.

Thanks to the work of Kantorovich, Monge's problem was inserted in a suitable framework. The main idea of Kantorovich consisted in relaxing the nature of transportation by describing a new way of movement for the mass sources. In Monges's problem, a mass source x can only be assigned or transported to one and only one location $T(x)$. Kantorovich suggested that, the mass at x can be dispatched to different locations y which is commonly known as "mass splitting". This new description of transport allows for more movement during mass transportation, thus, the transport problem is not described through a map T anymore.

Kantorovitch's problem is formulated as follows

$$\min_{\pi} \left\{ \int_{X \times Y} c(x,y) d\pi(x,y) : \pi \in \Pi(\mu,\nu) \right\}, \tag{2.2}$$

where $\Pi(\mu,\nu)$ is the set of the so-called transportation plans such that

$$\Pi(\mu,\nu) = \{\pi \in \mathscr{P}(X \times Y) : (\text{Proj}_x)_{\#}\pi = \mu, (\text{Proj}_y)_{\#}\pi = \nu\}, \tag{2.3}$$

and Proj_x and Proj_y are the two projections of $X \times Y$ onto X and Y respectively. Thus, by changing the point of view of Monge's problem, Kantorovitch described the transportation of mass via a measure π, also called *transport plan*, on $X \times Y$

satisfying $(\text{Proj}_x)_{\#}\pi = \mu$ and $(\text{Proj}_y)_{\#}\pi = \nu$. Instead of focusing on finding the location $T(x)$ of each mass source x, the objective of π is to precise, for each pair (x, y), how can a mass originally located at x be distributed in other locations y. We note that the value $\pi(A \times B)$ denotes the amount of mass moving from $A \in X$ to $B \in Y$ and the constraints $(\text{Proj}_x)_{\#}\pi = \mu$ and $(\text{Proj}_y)_{\#}\pi = \nu$ mean that we only focus on moving the mass sources distributed according to μ onto ν. The new Kantorovitch formulation of optimal transport made a connection between Monge's problem and linear programming which opened the door to an important amount of applications. One of the essential facts established in Kantorovitch's work is the introduction of a relevant quantity to compare probability measures known nowadays as the p-Wasserstein distance which is defined as follows

$$W_p(\mu, \nu) = \left(\min_{\pi} \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) : \pi \in \Pi(\mu, \nu) \right\} \right)^{\frac{1}{p}}, \tag{2.4}$$

where $p \in [1, \infty)$ is the order of the Wasserstein distance. The quantity W_p is called distance because it satisfies all the axioms of a distance, see [106] for more details. In the remainder of this chapter we work with $p = 1$, and instead of denoting the 1-Wasserstein distance $W_1(\mu, \nu)$, we simply write $W(\mu, \nu)$.

Optimal transport has been used in many fields for different applications such as in image processing [95, 109], remote sensing [110], computer graphics [98], image retrieval in computer vision [101], machine learning [102] and astrophysics [111]. Computationally speaking, the resolution of optimal transport problems can be very costly and it is one of the drawbacks of using optimal transport tools. In fact, Wasserstein distances, which are important metrics in optimal transport, have many interesting properties, but their formulation imposes the resolution of linear problems with costs that can grow incredibly fast, especially when measures with important supports are manipulated. In the

discrete case, when the probability measures have the same support, many resolution methods exist to solve optimal transport problems [112]. The complexity of these methods is $O(N^3)$ and it becomes too costly when N is important such as in image processing problems where N can very rapidly reach one million pixels.

Developing optimal algorithms to tackle optimal transport problems have been the object of a variety of research work. For example, in the case of quadratic costs in transport problems, linearizing Monge-Ampère equation with the Newton solver method can be used to solve optimal transport problems [113]. In [114], the problem of computing the transport is tackled differently by using proximal splitting methods. Computing Wasserstein distances with classic methods such as the interior point method or the simplex one, which are iterative methods, is quite costly too. For example, computing the Wasserstein distance between two probability measures or two histograms with supports of size N is of the order of $O(N^3 \log N)$.

Improving the computational performances of optimal transport algorithms has long been an active field of research. Among the interesting ideas that contributed to the widespread of optimal transport applications there is the entropic regularization. The latter consists in regularizing the Wasserstein distance by substracting the entropic energy times a regularization coefficient. The goal of using the entropic regularization is to provide an approximation to the original linear Wasserstein distance. The idea behind entropic regularization has its roots in the work of Schrodinger [115] and it was mainly used in optimal transport problems for predicting flows of commodities and traffic patterns [116]. Using entropic regularization in these kinds of problems ensures the smoothness of the flows. Thanks to the concavity of the entropic energy (i.e. minus the entropy is convex), the entropic regularization has many computational

properties such as strictly convexifying the transport problem, which ensures the uniqueness of the solution and allows the use of matrix scaling based methods. One of the most famous matrix-scaling based algorithms for solving transport problems there is Sinkhorn-Knopp algorithm [117, 118]. The latter, known for having a linear convergence rate, is based on matrix-vector multiplications and more specifically, on the diagonal scaling of the exponential of minus the cost matrix. Sinkhorn-Knopp's algorithm's matrix-vector multiplications are easily parallelized which makes the computations go faster in addition to the fact that it is easy to implement.

The entropic regularization has the advantage of ensuring the non-sparsity of the solution and helps to stabilize the solution [119]. Furthermore, it has also a connection to Kullback-Leibler (KL) divergences and it allows the use of iterative Bregman projections [120] which are carried out on affine subsets. The relationship between KL divergences and Bregman projections will be highlighted in details in the remainder of the present thesis. One should stress that the interesting computational properties provided by the entropic regularization are independent of the choice of the ground metric (or the cost function) [100].

One of the widespread applications of optimal transport is the Wasserstein barycenter which is an average in the sense of the Wasserstein metric. Wasserstein barycenter was first solved by replacing the Wasserstein metric by a sliced approximation over 1D distribution which allowed the use of a stochastic gradient descent algorithm [121]. The drawback of this method is that it is not easy to generalize for high dimensional and for non-Euclidean metric spaces. The properties of the Wasserstein barycenter in practice were highlighted in [101] and for the theoretical properties the reader can check Villani's book [106]. The regularized Wasserstein barycenter, computed with regularized Wasserstein

distances, is formulated as the solution, over the space of measures, of a convex variational problem and it is composed of the sum of weighted regularized Wasserstein distances, see [122] for a detailed study of the Wasserstein barycenter.

The optimal transport barycenter is used in statistics where it is very practical for computing estimators [123] or in machine learning for computing an average of a family of histograms [124]. Many methods in the literature suggested algorithms for solving the Wasserstein barycenter problem. In [124] the regularized barycenter problem is solved with a subgradient descent based method where the gradients are computed with matrix scaling algorithms. The formulation of the regularized Wasserstein barycenter has many computational advantages like allowing the use of Sinkhorn's algorithm described above as well as Bregman projections under KL divergences [119]. In general, solving the barycenter problem with iterative Bregman projections is faster and easier than using a gradient descent method.

2.3 Discrete optimal transport

Let Ω be an arbitrary space and let $X = \{x_1, \ldots, x_n\}$ be a set of Ω, $x_i \in \Omega$, $\forall i = 1, \ldots, n$. A discrete measure u with weights μ and locations $x_1, \ldots, x_n \in \Omega$ reads

$$u = \sum_{i=1}^{n} \mu_i \delta_{x_i},$$

where μ is a vector of probability masses $(\mu_i)_{i=1}^{n}$ and δ_{x_i} is the Dirac located at position x_i which represents the quantity of mass μ_i concentrated in the location x_i. The measure u is a probability measure if μ belongs to the probability

simplex Σ_n defined as follows

$$\Sigma_n \stackrel{\text{def}}{=} \left\{ \mu \in \mathbb{R}_+^n \; ; \; \sum_{i=1}^n \mu_i = 1 \right\}. \tag{2.5}$$

In optimal transport, the manipulated objects in the continuous or the discrete framework need to be modeled as probability measures. We consider two probability measures u and v with two probability vectors $\mu \in \Sigma_n$ and $v \in \Sigma_m$ respectively. Then, the Wasserstein distance between μ and v is defined as follows

$$W(\mu, v) \stackrel{\text{def}}{=} \min_{\pi \in \Pi(\mu, v)} < \pi, D >,$$

where $\Pi(\mu, v)$ is the set of all couplings that have μ and v as marginals defined as follows

$$\Pi(\mu, v) \stackrel{\text{def}}{=} \{ \pi \in \mathbb{R}_+^{n \times m} | \pi \mathbb{1}_m = \mu, \pi^\mathsf{T} \mathbb{1}_n = v \},$$

with $\forall l \in \mathbb{N}$, $\mathbb{1}_l \stackrel{\text{def}}{=} (1, \dots, 1)^\mathsf{T} \in \mathbb{R}^l$ and $D \in \mathbb{R}^{n \times m}$ is a cost function defined as

$$D = [d(x_i, y_j)]_{ij},$$

where $d(x_i, y_j)$ is a distance between the positions of two Diracs located at x_i and y_j.

2.3.1 Regularized Wasserstein distance

The numerical implementation of optimal transport algorithms has a high computational cost especially in image processing. Introducing the entropic regularization has the advantage of making the minimization problem strictly

convex, which guaranties the uniqueness of the minimum and the fast computation of the latter with matrix-vector based algorithms. The entropic regularization is carried out by penalizing the entropy of the joint coupling π and it is defined as follows

$$W_\gamma(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} <\pi, D> -\gamma E(\pi), \tag{2.6}$$

where γ is the regularization coefficient and $E(\pi)$ is the entropy defined as

$$E(\pi) = -\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{i,j}(\log \pi_{i,j} - 1) - \iota_{\mathbb{R}+}(\pi_{i,j}),$$

where ι is the indicator of $\mathbb{R}_+$ s.t

$$\forall x, \quad \iota_{\mathbb{R}_+}(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R}_+, \\ +\infty & \text{otherwise.} \end{cases}$$

The regularized Wasserstein distance is not a distance in the strict sense, thus, it does not satisfy the axioms of a metric because of the regularization coefficient: $W_\gamma(\mu, \mu) \neq 0$. We can notice that the original non-regularized linear problem can be recovered as $\gamma \rightarrow 0$. However, when γ is too small, the computations might become unstable.

Entropic regularization is linked to the divergence of Kullback-Leibler and then to Sinkhorn's algorithm which leads to simple formulation of the solution of the regularized problem. The resolution of problem (2.6) is equivalent to solving the following minimization problem

$$\min_{\pi \in \Pi(\mu,\nu)} \mathrm{KL}(\pi|\xi), \tag{2.7}$$

where ξ is a matrix computed component-wise and defined as

$$\xi = e^{-\frac{D}{\gamma}}, \tag{2.8}$$

and KL denotes the divergence of Kullback-Leibler defined as follows

$$\mathrm{KL}(\pi|\xi) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{i,j} \log\left(\frac{\pi_{i,j}}{\xi_{i,j}}\right) - \pi_{i,j} + \xi_{i,j}.$$

The entropy and the KL divergence are two concepts that come from the information theory. The entropy is a measure of the average of information delivered from a data source whereas the KL divergence measures how far two distributions deviate from each other. The KL divergence is not a distance because it does not satisfy the axioms of a metric, for example the symmetry is not verified: $\mathrm{KL}(\pi|\xi) \neq \mathrm{KL}(\xi|\pi)$.

Proposition 1. *The problem* (2.6) *has the unique solution [119]:*

$$\forall (i,j) \in [\![1,n]\!] \times [\![1,m]\!], \;\; \pi_{i,j} = a_i \xi_{i,j} b_j,$$

where $a \in \mathbb{R}_+^n$ *and* $b \in \mathbb{R}_+^m$ *are two scaling vectors.*

Proof. Let $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^m$ two Lagrangian multipliers. The Lagrangian of (2.6) is written as

$$L(\pi,\alpha,\beta) = <\pi,D> -\gamma E(\pi) - <\alpha,\pi\mathbb{1}_m - \mu> - <\beta,\pi^\mathsf{T}\mathbb{1}_n - \nu>$$

The first order condition gives

$$\frac{\partial L(\pi, \alpha, \beta)}{\partial \pi_{i,j}} = D_{i,j} + \gamma \log \pi_{i,j} - \alpha_i - \beta_j = 0.$$

Thus, the optimal coupling verifies: $\pi_{i,j} = e^{\alpha_i/\gamma} e^{-D_{i,j}/\gamma} e^{\beta_j/\gamma}$. By expression (2.8), $e^{-\frac{D_{i,j}}{\gamma}} = \xi_{i,j}$ and taking $a_i = e^{\frac{\alpha_i}{\gamma}}$, $b_j = e^{\frac{\beta_j}{\gamma}}$, the optimal coupling is written as $\pi_{i,j} = a_i \xi_{i,j} b_j$. Finally, in matrix notation, the solution of (2.6) can be written as

$$\pi = \text{diag}(a) \xi \text{diag}(b),$$

where $\text{diag}(a)$ and $\text{diag}(b)$ are diagonal matrices where the diagonal entries are the vectors a and b respectively. ∎

A theorem by Sinkhorn [125] states that for any matrix ξ with positive elements, there exists diagonal scaling matrices $\text{diag}(a)$ and $\text{diag}(b)$ with a, $b > 0$ such that $\pi = \text{diag}(a) \xi \text{diag}(b)$ has prescribed row sums and column sums (π satisfies $\pi \mathbb{1}_m = \mu$ and $\pi^{\intercal} \mathbb{1}_n = \nu$). The coupling π can be seen as the scaling of ξ and the scaling vectors a and b may be found by Sinkhorn iterations. At each iteration the coupling is defined as follows

$$\pi^{(i)} = \text{diag}(a^{(i)}) \xi \text{diag}(b^{(i)}),$$

where $a^{(i)}$ and $b^{(i)}$ are the updated scaling vectors throughout Sinkhorn's iterations as follows

$$a^{(i+1)} \stackrel{\text{def}}{=} \frac{\mu}{\xi b^{(i)}} \quad \text{and} \quad b^{(i)} \stackrel{\text{def}}{=} \frac{\nu}{\xi^{\intercal} a^{(i)}}. \tag{2.9}$$

Thanks to the updating steps (2.9), we do not have to manipulate at each iteration the transport coupling variable π which has the size $n \times n$ and it is

simply huge when it comes to image processing application. So instead of updating π at each iteration, it is sufficient to update the two scaling vectors a and b which have the sizes n and m respectively. Besides, by applying these two scaling steps, the computations can be easily vectorized and parallelized and thus accelerate the computational time which is valuable in minimization problems.

2.3.2 Wasserstein barycenter

Computing the barycenter of Wasserstein metrics is one of the important problems in the optimal transport theory. In the discrete case it consists in computing the average of empirical probability measures in the sense of the optimal transport metric. In fact, this average is the minimum of the sum of each Wasserstein distance to each probability measure.

Let $(q_k)_{k=1}^{K}$ be a set of K positive probability vectors such that $q_k \in \Sigma_n$ and let q be the Wasserstein mean of these probability vectors. Then, the Wasserstein barycenter problem is presented as follows

$$\min_{q \in \Sigma_n} \sum_{k=1}^{K} \lambda_k W_{\gamma_k}(q_k, q), \tag{2.10}$$

where W_{γ_k} are regularized Wasserstein distances with γ_k being regularization coefficients, λ_k are weights related to W_{γ_k} such that $\sum_{k=1}^{K} \lambda_k = 1$ and Σ_n is the simplex defined in (2.5) over which the minimization (2.10) is carried out.

The problem (2.10) is subject to constraints which are defined as

$$
\begin{aligned}
C_1 &\overset{\text{def}}{=} \{\boldsymbol{\pi} = (\pi_k)_1^K \in (\mathbb{R}_+^{n \times n})^K; \exists q \in \mathbb{R}^n, \forall k \in \{1, \ldots, K\}, \; \pi_k \mathbb{1} = q\}, \\
C_2 &\overset{\text{def}}{=} \{\boldsymbol{\pi} = (\pi_k)_1^K \in (\mathbb{R}_+^{n \times n})^K; \forall k \in \{1, \ldots, K\}, \; \pi_k^\mathsf{T} \mathbb{1} = q_k\},
\end{aligned}
\tag{2.11}
$$

where π_k^T is the transpose of π_k and $\mathbb{1} = (1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^n$.

The subsets C_1 and C_2 are affine constraints that describe the conditions that the optimal coupling $\boldsymbol{\pi} = (\pi_k)_1^K$ has to verify. The first constraint C_1 imposes that the sum over the columns of each optimal coupling π_k is equal to the Wasserstein mean q which is the sought solution, whereas the sum over the rows has to be equal to each marginal q_k, and that is the purpose of the constraint C_2.

Earlier, a link between regularized Wasserstein distances and KL divergences (Section 2.3.1) was established. Based on this relationship, the problem (2.10) can be rewritten as follows

$$
\min_{\substack{\boldsymbol{\pi} = (\pi_k)_{k=1}^K \\ \text{s.t.} \\ \boldsymbol{\pi} \in C1 \cap C2}} \left\{ \mathrm{KL}_\lambda(\boldsymbol{\pi}|\boldsymbol{\xi}) = \sum_{k=1}^K \lambda_k \mathrm{KL}(\pi_k|\xi_k) \right\}.
$$

The set of probability vectors $(q_k)_{k=1}^K$ as well as the mean q have the same support size (the same number of Diracs), so the matrices D_k and ξ_k have the same size for all Wasserstein distances.

At each iteration, each coupling π_k has the form

$$
\pi_k^{(i)} = \mathrm{diag}(a_k^{(i)}) \xi_k \mathrm{diag}(b_k^{(i)}),
$$

and the scaling vectors are updated as follows

$$a_k^{(i+1)} \stackrel{\text{def}}{=} \frac{q}{\xi b_k^{(i)}} \quad \text{and} \quad b_k^{(i)} \stackrel{\text{def}}{=} \frac{q_k}{\xi^\intercal a_k^{(i)}}.$$

At this point, we know how to update the scaling vectors and to compute the optimal coupling but still, the barycenter given by the Wasserstein mean (q) needs to be computed. The computation of the barycenter is given by the following proposition [119]

Proposition 2. *For* $\boldsymbol{\pi} \stackrel{\text{def}}{=} (\pi_k)_k \in (\mathbb{R}_+^{n \times n})^K$, *the projection* $\overline{\boldsymbol{\pi}} \stackrel{\text{def}}{=} (\overline{\pi}_k)_k = Proj_{C_1}(\boldsymbol{\pi})$ *satisfies*

$$\overline{\pi}_k = diag(q \oslash \pi_k \mathbb{1})\pi_k \quad \text{with} \quad q = \Pi_{k=1}^K (\pi_k \mathbb{1})^{\lambda_k}, \tag{2.12}$$

Where Π *and* $(.)^{\lambda_k}$ *should be understood as element-wise operations and* $\oslash$ *denotes element-wise division.*

Proof. The optimization problem is the following

$$\min_{(\overline{\pi}_k)_k, q} \quad \sum_{k=1}^K \lambda_k KL(\overline{\pi}_k | \pi_k) \tag{2.13}$$

$$\text{s.t.} \quad \overline{\pi}_k \mathbb{1} = q \ \forall k \in \{1, \dots, K\}$$

Let $(\beta_k)_k \in (\mathbb{R}^n)^K$ be Lagrange multipliers. The Lagrangian of problem (2.13) is written

$$L(\overline{\pi}_k, q, \beta_k) = \sum_k \lambda_k KL(\overline{\pi}_k | \pi_k) - <\beta_k, \overline{\pi}_k \mathbb{1} - q>$$

$$= \sum_k \lambda_k \sum_{i,j} (\overline{\pi}_k)_{i,j} \log\left(\frac{(\overline{\pi}_k)_{i,j}}{e(\pi_k)_{i,j}}\right) - <\beta_k, \overline{\pi}_k \mathbb{1} - q>$$

The first order condition with respect to $(\overline{\pi}_k)_{i,j}$ gives

$$\frac{\partial L(\overline{\pi}_k, q, \beta_k)}{\partial \overline{\pi}_k)_{i,j}} = \lambda_k \log\left(\frac{(\overline{\pi}_k)_{i,j}}{(\pi_k)_{i,j}}\right) - (\beta_k)_i = 0$$

which gives

$$(\bar{\pi}_k)_{i,j} = e^{-\frac{(\beta_k)_i}{\lambda_k}} (\pi_k)_{i,j}$$

and in terms of matrix notations we have

$$(\bar{\pi}_k) = \text{diag}(e^{-(\beta_k \oslash \lambda_k)})\pi_k \tag{2.14}$$

where $\oslash$ is element-wise division.

The first order condition with respect to q gives

$$\nabla_q L(\bar{\pi}_k, q, \beta_k) = -\sum_k \beta_k = 0 \tag{2.15}$$

with the result established in (2.14) we have

$$q = \bar{\pi}_k \mathbb{1}$$
$$= \text{diag}\left(e^{-\beta_k \oslash \lambda_k}\right)\pi_k \mathbb{1}$$

define $r_k = \pi_k \mathbb{1}$, we have then

$$q = \text{diag}\left(e^{-\beta_k \oslash \lambda_k}\right)r_k$$
$$= e^{-\beta_k \oslash \lambda_k} \odot r_k \tag{2.16}$$

where $\odot$ is element-wise multiplication.

The first order condition established in (2.15) shows that $\sum_k \beta_k = 0$, thus

$$q = e^{-\beta_k \oslash \lambda_k} \odot r_k \Leftrightarrow q \oslash r_k = (e^{-\beta_k})^{\frac{1}{\lambda_k}}$$
$$\Leftrightarrow (q \oslash r_k)^{\lambda_k} = e^{-\beta_k} \tag{2.17}$$
$$\Rightarrow \Pi_k (q \oslash r_k)^{\lambda_k} = \Pi_k e^{-\beta_k} = e^{-\sum_k \beta_k} = 1$$

The operations $(\cdot)^*$, Π and $\sum$ are element-wise ones. We know that $\sum_k \lambda_k = 1$, we have then

$$\Pi_k (q \oslash r_k)^{\lambda_k} = q^{\sum_k \lambda_k} \oslash \Pi_k r_k^{\lambda_k}$$
$$= q \oslash \Pi_k r_k^{\lambda_k}$$

(2.18)

from (2.17) and (2.18) we have

$$q \oslash \Pi_k r_k^{\lambda_k} = 1 \Leftrightarrow q = \Pi_k r_k^{\lambda_k}$$
$$= \Pi_k (\pi_k \mathbb{1})^{\lambda_k}$$

We have already established in (2.16) that $q = e^{-\beta_k \oslash \lambda_k} \odot r_k$, we conclude then that $e^{-\beta_k \oslash \lambda_k} = q \oslash r_k$ and if we insert it in the relation established in (2.14) we have

$$\overline{\pi}_k = \mathrm{diag}(q \oslash r_k) \pi_k$$
$$= \mathrm{diag}(q \oslash \pi_k \mathbb{1}) \pi_k.$$

Hence we have $\overline{\pi}_k = \mathrm{diag}(q \oslash \pi_k \mathbb{1}) \pi_k$ and $q = \Pi_k (\pi_k \mathbb{1})^{\lambda_k}$

$\blacksquare$

From the previous proposition, we have $q = \Pi_k (\pi_k \mathbb{1})^{\lambda_k}$ and $\pi_k = \mathrm{diag}(a_k) \xi_k \mathrm{diag}(b_k)$, thus, the barycenter q is given for each iteration by the following formula

$$q^{(i)} = \Pi_{k=1}^K (a_k^{(i)} \odot (\xi b_k^{(i)}))^{\lambda_k}.$$

2.4 Formulation of the fusion problem

During the fusion process, spectral and spatial information is transferred from the observed data, the hyperspectral and the multispectral image, to an induced image **u** which is the result of the fusion. Following this idea of

energy transfer from one image to another, the HS-MS fusion is formulated as a Wasserstein barycenter (problem (2.10)) as follows

$$\min_{\mathbf{u} \in \Sigma_u} \left\{ G(\mathbf{u}) = \lambda W_{\gamma_M}(\mathbf{u}, \widetilde{\mathbf{S}}(\mathbf{f})) + (1 - \lambda) W_{\gamma_H}(\mathbf{u}, \widetilde{\mathbf{T}}(\mathbf{g})) \right\}, \tag{2.19}$$

where Σ_u is the simplex is $\mathbb{R}^{H \times N}$ over which the minimization (2.19) is carried out, defined by $\Sigma_N = \{\mathbf{u} \in \mathbb{R}_+^{H \times N}, \sum_{i=1}^{HN} \mathbf{u}_i = 1\}$.

Then, the HS-MS fusion problem (2.19) is formulated as the minimization of the sum of two weighted Wasserstein distances. The minimization is done over $\mathbf{u} \in \mathbb{R}^{H \times N}$ which is the high-resolution hyperspectral image to be produced. In the generic case, the barycenter problem (2.19) is a balanced one which means that each Wasserstein distance is computed between two images that have the same dimensions. In our problem, the images $\mathbf{f} \in \mathbb{R}^{M \times N}$ and $\mathbf{g} \in \mathbb{R}^{H \times N_l}$ are not on the same grid which means that they do not have the same dimensions. In this case, the introduction of inversion operators is needed in order to bring the input images $\mathbf{f}$ and $\mathbf{g}$, that live on a coarse grid, on the same fine grid as $\mathbf{u}$. The operator $\widetilde{\mathbf{S}}$ interpolates spectrally $\mathbf{f} \in \mathbb{R}^{M \times N}$ whereas $\widetilde{\mathbf{T}}$ is a spatial inversion operator that interpolates spatially each pixel of each band of $\mathbf{g} \in \mathbb{R}^{H \times N_l}$. Both operators bring $\mathbf{f}$ and $\mathbf{g}$ to the same dimensions as $\mathbf{u}$, therefore we have $\widetilde{\mathbf{S}}(\mathbf{f}) \in \mathbb{R}^{H \times N}$ and $\widetilde{\mathbf{T}}(\mathbf{g}) \in \mathbb{R}^{H \times N}$. The operation $\widetilde{\mathbf{T}}(\mathbf{g})$ is carried out by applying a classic bicubic interpolation for each pixel in each band of the hyperspectral image $\mathbf{g}$, while $\widetilde{\mathbf{S}}(\mathbf{f})$ is obtained by solving the following least square problem

$$\min_{\widetilde{\mathbf{S}} \in \mathbb{R}^{H \times M}} \|(\mathbf{S} \circ \widetilde{\mathbf{S}})(\mathbf{f}) - \mathbf{f}\|^2. \tag{2.20}$$

In practice, the operator $\mathbf{S} \in \mathbb{R}^{M \times H}$ is a non-square matrix with M rows and H columns which are the numbers of the multispectral and the hyperspectral

bands respectively. The matrix $\mathbf{S}$ is non-invertible and $\mathbf{S}^{-1}$ cannot be taken for an inverse operator. So hijacking the invertibility of $\mathbf{S}$ consists in looking for a pseudo-inverse matrix that approaches the behavior of an inverse matrix and that brings $\mathbf{f}$ back on the finer grid. One of the ways of inverting $\mathbf{f}$ spectrally is through the resolution of the minimization problem (2.20).

As one can notice, the functional (2.19) to be minimized is composed of two regularized Wasserstein distances. The first distance $W_{\gamma_M}(\mathbf{u}, \widetilde{\mathbf{S}}(\mathbf{f}))$ is computed between $\mathbf{u}$ and $\widetilde{\mathbf{S}}(\mathbf{f})$ with the latter containing the valuable spatial information. The role of $W_{\gamma_M}(\mathbf{u}, \widetilde{\mathbf{S}}(\mathbf{f}))$ is the minimization of the transport cost between $\widetilde{\mathbf{S}}(\mathbf{f})$ and $\mathbf{u}$, so in the spirit of optimal transport, this first distance could be interpreted as the transport of spatial information from $\widetilde{\mathbf{S}}(\mathbf{f})$ to $\mathbf{u}$. As to the second distance $W_{\gamma_H}(\mathbf{u}, \widetilde{\mathbf{T}}(\mathbf{g}))$, it represents the transport of the spectral energy from $\widetilde{\mathbf{T}}(\mathbf{g})$ to $\mathbf{u}$. By being the average in the sense of the Wasserstein metric, $\mathbf{u}$ gathers the relevant spectral information of $\widetilde{\mathbf{T}}(\mathbf{g})$ and the geometric details of $\widetilde{\mathbf{S}}(\mathbf{f})$.

In order to balance between the spectral and the spatial information, the trade-off parameter λ is used in the weight for each Wasserstein distance with $\lambda \in [0, 1]$. The weight of each Wasserstein distance gives the possibility to choose whether to have more spatial than spectral details or vice versa. The choice of λ depends on the purpose of the produced image $\mathbf{u}$ and the applications this latter is destined to.

As it was mentioned before, the aim of optimal transport is to transfer energy from a starting point to an arrival one with a minimum cost. The cost to be minimized is expressed in a cost matrix that describes how, for each Wasserstein distance, matter should be redistributed in the new configuration. In the fusion problem, the manipulated data is three-dimensional and it is composed of a two-dimensional spatial component and a one-dimensional spectral one. Thus, the energy transport in each Wasserstein distance is done in a three-dimensional

space which means that the cost matrix should describe a three-dimensional transport.

Let p and q be two different pixels and let (x_p, y_p, z_p) and (x_q, y_q, z_q) be their coordinates respectively. The ground cost paid for a transport from pixel p to pixel q is computed as follows

$$D_k(p, q) = (x_p - x_q)^2 + (y_p - y_q)^2 + \alpha_k(z_p - z_q)^2 \qquad (2.21)$$

where $k \in \{M, H\}$, with M stands for the multispectral component and H for the hyperspectral one.

One can see that the cost of transport between two pixels consists of a spatial distance and a spectral one with a parameter α_k that balances the spectral distance with respect to the spatial one. In practice, the spatial distances are expressed in meters(m) and the spectral ones, which are differences between wavelenghts, are expressed in nanometers ($nm = 10^{-9}$). Thus, in the absence of a parameter to balance these two distances, the contribution of the spectral distance does not have any influence which would mean that the transport occurs only spatially which is not realistic.

One should stress that the distance D_k of (2.21) is computed with the positions of the pixels and it does not depend on their values but only on the grid on which the image is defined. Because of the existence of the parameter α_k, the ground cost distance matrix is not going to be the same for both Wasserstein distances. This could be explained by the fact that the transport following the spectral dimension is going to depend on the type of data. In fact, the hyperspectral image contains more relevant spectral data so in this case, the transport in the spectral domain is obviously going to be favored over the transport in the spatial one.By the contrary, for the multispectral image, the transport in the spatial

domain is going to be more important that the one in the spectral domain.

Once the ground cost is determined by (2.21), the computation of the cost matrix ξ_k from expression (2.8) and for two pixels p and q is given by

$$
\begin{aligned}
\xi_k(p,q) &= \exp\left(-\frac{1}{\gamma_k}((x_p-x_q)^2+(y_p-y_q)^2) - \frac{\alpha_k}{\gamma_k}(z_p-z_q)^2\right) \\
&= \underbrace{\exp\left(-\frac{1}{\gamma_k}((x_p-x_q)^2+(y_p-y_q)^2)\right)}_{\text{2D spatial distance}} \underbrace{\exp\left(-\frac{\alpha_k}{\gamma_k}((z_p-z_q)^2)\right)}_{\text{1D spectral distance}} \\
&= \underbrace{\exp(-\frac{1}{\gamma_k}(x_p-x_q)^2)}_{\text{1D spatial ditance}} \underbrace{\exp(-\frac{1}{\gamma_k}(y_p-y_q)^2)}_{\text{1D spatial distance}} \underbrace{\exp(-\frac{\alpha_k}{\gamma_k}(z_p-z_q)^2)}_{\text{1D spectral distance}},
\end{aligned}
$$

where (x_p, y_p, z_p) and (x_q, y_q, z_q) are the coordinates of p and q respectively.

2.5   Computations of the cost matrix

The matrix ξ contains all the distances of all the pixels with respect to every other pixel in the image. The image is browsed in columns (or in rows) and for each pixel, the distances with respect to all the pixels of the image should be computed following the formula (2.8). For example, in the case of a 3D image of an overall size of N, the first row of ξ contains all the distances between the first pixel with the coordinates $(1,1,1)$ and all the other pixels of the image, thus, the first row of ξ is a vector of size N. Figure 2.6 shows an example of computing the ξ matrix of a 3D image of size $2 \times 2 \times 2$. The overall size of the latter is $N = 8$, thus the matrix ξ corresponding to this image has a size of 8×8.

For a 3D image that has n_x rows, n_y columns and n_z spectral bands, the position of each pixel (x,y,z) is converted to a linear subscript denoted l_{xyz} following the formula:

The initial grid shows cells labeled (1,1,1), (1,2,1), (1,1,2), (1,2,2), (2,1,1), (2,2,1), (2,2,2) with arrow "Construction of ξ".

$$
\begin{array}{cccccccc}
1 & e^{-\frac{1}{\gamma}} & e^{-\frac{1}{\gamma}} & e^{-\frac{2}{\gamma}} & e^{-\frac{\alpha}{\gamma}} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(2+\alpha)} \\
e^{-\frac{1}{\gamma}} & 1 & e^{-\frac{2}{\gamma}} & e^{-\frac{1}{\gamma}} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{\alpha}{\gamma}} & e^{-\frac{1}{\gamma}(2+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} \\
e^{-\frac{1}{\gamma}} & e^{-\frac{2}{\gamma}} & 1 & e^{-\frac{1}{\gamma}} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(2+\alpha)} & e^{-\frac{\alpha}{\gamma}} & e^{-\frac{1}{\gamma}(1+\alpha)} \\
e^{-\frac{2}{\gamma}} & e^{-\frac{1}{\gamma}} & e^{-\frac{1}{\gamma}} & 1 & e^{-\frac{1}{\gamma}(2+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{\alpha}{\gamma}} \\
e^{-\frac{\alpha}{\gamma}} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(2+\alpha)} & 1 & e^{-\frac{1}{\gamma}} & e^{-\frac{1}{\gamma}} & e^{-\frac{2}{\gamma}} \\
e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{\alpha}{\gamma}} & e^{-\frac{1}{\gamma}(2+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}} & 1 & e^{-\frac{2}{\gamma}} & e^{-\frac{1}{\gamma}} \\
e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(2+\alpha)} & e^{-\frac{\alpha}{\gamma}} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}} & e^{-\frac{2}{\gamma}} & 1 & e^{-\frac{1}{\gamma}} \\
e^{-\frac{1}{\gamma}(2+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{1}{\gamma}(1+\alpha)} & e^{-\frac{\alpha}{\gamma}} & e^{-\frac{2}{\gamma}} & e^{-\frac{1}{\gamma}} & e^{-\frac{1}{\gamma}} & 1 \\
\end{array}
$$
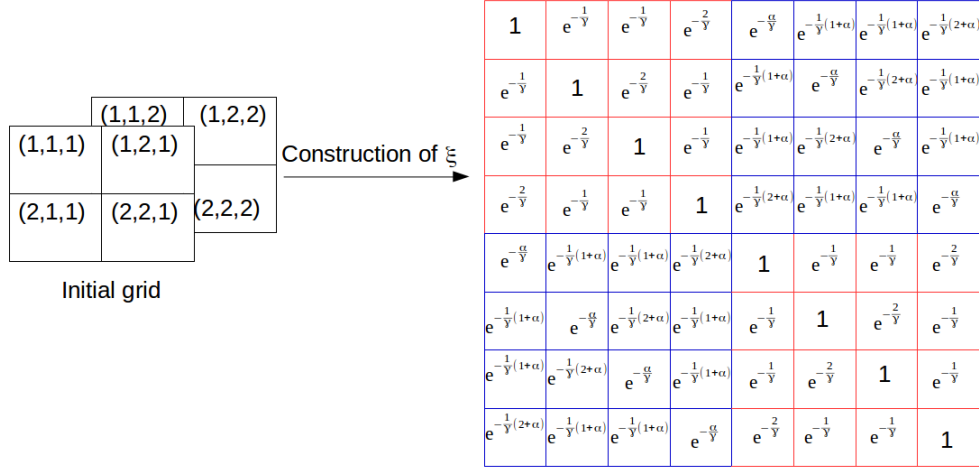
Figure 2.6: The construction of ξ from the grid of a $2 \times 2 \times 2$ image

$$l_{xyz} = x + (y-1)n_x + (z-1)n_y n_x.$$

Thus, all the distances of the pixel (x, y, z) with respect to all the pixels of the image are stored in the row number l_{xyz} of the matrix ξ.

The size of ξ can be very important and it can exceed the RAM capacity storage of a laptop very rapidly which makes it hard to store ξ and even harder to manipulate and use it in the computations. For example, the size of the 3D satellite image in Figure 2.7 is $128 \times 128 \times 93$ which gives an overall number of pixels of $N_t = 128 \times 128 \times 93 = 1523712$. Visually, this image is quite small for any geometric interpretation due to the lack of a high number of pixels. However, the size of ξ in this case is 1523712×1523712 which in terms of GB corresponds to 18574 GB which is simply huge and cannot be carried out on a machine.

Thus, to hijack storage problems, it is important to look for ways to cope with the huge size of ξ that could make its use easy and not cumbersome. In order to illustrate the interesting properties of ξ let us go back to Figure 2.6 which highlights the computations of the matrix ξ for a grid of size $2 \times 2 \times 2$.

Figure 2.7: Hyperspectral image with a size of: $128 \times 128 \times 93$

For the sake of simplicity, the size of the grid was chosen very small so that the matrix ξ can be representable and readable. As one can notice, ξ has interesting properties, it is square, positive, symmetric and block-Toeplitz.

A matrix A is said to be Toeplitz when all its diagonals are constant. If the i,j element of A is denoted by $a_{i,j} = a_{i+1,j+1} = a_{i-j}$, then we have

$$A = (a_{i,j})_{\substack{i \in [\![1,N]\!] \\ j \in [\![1,N]\!]}} = a_{i,j} = a_{i-j}.$$

Then, the matrix representation of A is the following

$$A = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots & a_{-N+1} \\ a_1 & a_0 & a_{-1} & \ddots & \vdots \\ a_2 & a_1 & a_0 & \ddots & a_{-2} \\ \vdots & \ddots & \ddots & \ddots & a_{-1} \\ a_{N-1} & \dots & a_2 & a_1 & a_0 \end{pmatrix}. \tag{2.22}$$

The symmetry of ξ allows the computations of one half (the upper or the lower one) and then the other half of the matrix is deduced by computing the transpose. Moreover, the values of the diagonal are $e^0 = 1$ which saves the trouble of doing unnecessary computations each time. The symmetry of ξ makes the computations go fast but before that, ξ needs to be stored. Then, the fact that the matrix ξ is block-Toeplitz is going to play an important role in making the matrix storable.

2.5.1   In the case of a Toeplitz matrix

Let us recall that the goal is to find interesting properties of the matrix ξ that can help deal with memory issues, due to the huge size of ξ, and carry out conveniently all the computations where the latter is implicated. The idea behind the equivalence between the multiplication and the convolution was highlighted in [105]. In what follows, this idea is explained with some examples and computations.

As it has already been mentioned before, the matrix ξ is block-Toeplitz and interesting computational techniques can be concluded. Before starting to deal with block-Toeplitz matrices straight away, properties of matrix-vector multiplication are going to be studied first when the matrix in question is only Toeplitz.

In some cases, the matrix-vector multiplication can be very consuming in terms of memory space. In the case of a Toeplitz matrix, interesting properties allow replacing the multiplication by convolution, with the advantage of saving only a representative vector extracted from the matrix instead of saving the whole matrix.

In order to highlight the relationship between the convolution and the multiplication in the case of a Toeplitz matrix, let us go back to the $N \times N$ Toeplitz matrix A defined in (2.22). Let P be a $N \times 1$ vector defined as follows

$$P = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{N-1} \end{pmatrix}.$$

Let us now consider the vector G extracted from the matrix A defined as

follows

$$
G = \begin{bmatrix} \overset{G(2N-2)}{a_{N-1}} & \overset{G(2N-3)}{a_{N-2}} & \dots & \overset{G(N)}{a_1} & \overset{G(N-1)}{a_0} & \overset{G(N-2)}{a_{-1}} & \dots & \overset{G(1)}{a_{-N+2}} & \overset{G(0)}{a_{-N+1}} \end{bmatrix}.
$$

As one can notice, the vector G contains the elements of the first column and the first row of the matrix A. These elements are sufficient to reconstruct the result of the multiplication $A \times P$ because A is Toeplitz.

The discrete convolution, at an instant $n \in \mathbb{Z}$, between two real or complex-valued functions G and P defined on the set $\mathbb{Z}$, is given by

$$
y(n) = (G * P)(n) = \sum_{i=0}^{N-1} P(i)G(n-i), \tag{2.23}
$$

where y is the resulting function defined on $\mathbb{Z}$.

The multiplication $A \times P$ gives

$$
A \times P = \begin{pmatrix} a_0p_0 + a_{-1}p_1 + \dots + a_{-N+1}p_{N-1} \\ a_{-1}p_0 + a_0p_1 + \dots + a_{-N+2}p_{N-1} \\ \vdots \\ a_{N-1}p_0 + a_{N-2}p_1 + \dots + a_0p_{N-1} \end{pmatrix} = \begin{pmatrix} y(N-1) \\ y(N) \\ \vdots \\ y(2N-2) \end{pmatrix} \tag{2.24}
$$

As (2.24) shows, each row of the result of the multiplication $A \times P$ can be written as a convolution at precise instants from $N-1$ to $2N-2$. In other words, $A \times P$ can be expressed as the central part of the convolution between Z and P as follows

$$
A \times P = y((N-1) \rightarrow (2N-2))
$$

where the notation $n_i \rightarrow n_j$ should be understood as: from the instant n_i to the

instant n_j.

The evolution of the convolution (2.23) can be highlighted throughout a FIR filter (Finite Impulse Response), as shown in Figure 2.8, at discrete instants n. At the instant $N - 1$ we start having the first row of the multiplication $A \times P$ and at the instant $2N - 2$, the result of the multiplication is complete.
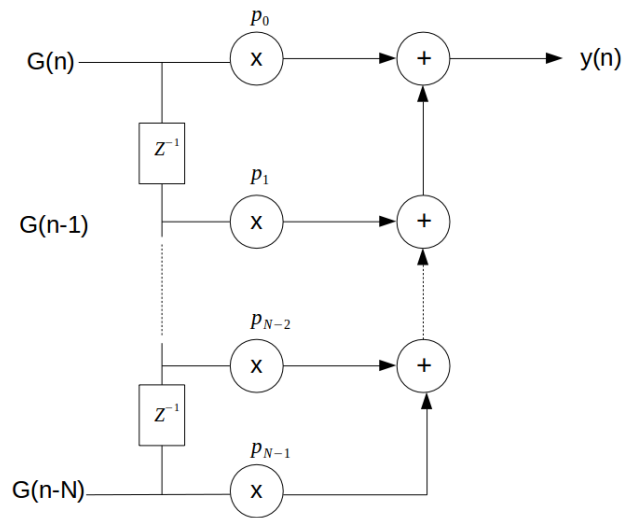
Figure 2.8: A FIR filter

From these first results we can conclude that the multiplication matrix-vector, when the matrix is Toeplitz, can be replaced with the central part of the convolution between a vector extracted from the matrix (as shown above) and the vector originally destined for the multiplication. By using the convolution, there is no need to store the whole matrix but only a vector which saves a considerable amount of memory space. The saving of memory space helps accelerate the computations and prevents from switching on the SWAP memory, which can slow down the computations considerably.

2.5.2 In the case of a block-Toeplitz matrix

The link between the multiplication matrix-vector and the convolution was made in the case of a Toeplitz matrix, except that in our case, the matrix ξ is block-Toeplitz and the multiplication matrix-vector is handled differently.

Let us consider in a general setting an image of size $n_l \times n_c \times b$ where n_l, n_c and b are the number of rows, columns and spectral bands respectively. The cost matrix ξ of this image is computed in a row-major order. It is a block-Toeplitz, symmetric and defined as follows

$$\xi = \begin{pmatrix} \xi_0 & \xi_1 & \xi_2 & \ldots & \xi_{b-1} \\ \xi_1 & \xi_0 & \xi_1 & \ldots & \xi_{b-2} \\ \xi_2 & \xi_1 & \xi_0 & \ldots & \xi_{b-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \xi_{b-1} & \ldots & \xi_2 & \xi_1 & \xi_0 \end{pmatrix}.$$

The matrices ξ_i are themselves block-Toeplitz and symmetric matrices of size $n_l \times n_l$ with $i \in [\![0, b-1]\!]$

Let F be the following block-vector

$$F = \begin{pmatrix} F_0 \\ F_1 \\ \vdots \\ F_{b-1} \end{pmatrix}$$

where F_i are block-vectors of size $n_l \times 1$ and $i \in [\![1, b-1]\!]$.

The multiplication $\xi \times F$ is given as follows

$$\xi \times F = \left(V * \left(\xi_0 * F \right) (0 \to b-1) \right) ((b-1) \to (2b-2))$$

68

where

$$V = \begin{bmatrix} e^{-(b-1)\frac{\alpha}{\gamma}} & e^{-(b-2)\frac{\alpha}{\gamma}} & \dots & e^{-\frac{\alpha}{\gamma}} & 1 & e^{-\frac{\alpha}{\gamma}} & \dots & e^{-(b-2)\frac{\alpha}{\gamma}} & e^{-(b-1)\frac{\alpha}{\gamma}} \end{bmatrix}$$

The details of the computations are given in Appendix A.

2.6 Fusion algorithm

The resolution of the fusion problem as it was formulated in Section 2.4 is carried out by the HS-MS fusion algorithm named HMWB (Hyperspectral and Multispectral Wasserstein Barycenter) presented as follows

---
Algorithm 1: HMWB

Input : $\widetilde{\mathbf{S}}(\mathbf{f}))$, $\widetilde{\mathbf{T}}(\mathbf{g})$, ξ_M, ξ_H

1 **Initialization**: $\mathbf{a}_M^{(0)} = \mathbf{a}_H^{(0)} = \mathbf{b}_M^{(0)} = \mathbf{b}_H^{(0)} = \mathbb{1}_{\mathbb{R}^{H \times N}}$;

2 **for** $i \leftarrow$ *0* **to** *niter* **do**

3   // *Updating the scaling vectors with data (Sinkhorn's iteration)*

4     $\mathbf{b}_M^{(i)} \leftarrow \frac{\widetilde{\mathbf{S}}(\mathbf{f}))}{\xi_M * \mathbf{a}_M^{(i)}}$, $\mathbf{b}_H^{(i)} \leftarrow \frac{\widetilde{\mathbf{T}}(\mathbf{g})}{\xi_H * \mathbf{a}_H^{(i)}}$;

5   // *Computing the barycenter*

6     $\mathbf{u} \leftarrow \exp(\lambda \log(\mathbf{a}_M^{(i)} \odot ((\xi_M * \mathbf{b}_M^{(i)})) + (1-\lambda) \log(\mathbf{a}_H^{(i)} \odot (\xi_H * \mathbf{b}_H^{(i)})))$;

7   // *Updating the scaling vectors with **u** (Sinkhorn's iteration)*

8     $\mathbf{a}_M^{(i+1)} \leftarrow \frac{\mathbf{u}}{(\xi_M * \mathbf{b}_M^{(i)})}$, $\mathbf{a}_H^{(i+1)} \leftarrow \frac{\mathbf{u}}{(\xi_H * \mathbf{b}_H^{(i)})}$;

9 **end**

Output: $\mathbf{u}$ (the fused image)

The operation $\frac{a}{b}$ should be understood as an element-wise division, the operation $*$ refers to the convolution developed in (2.5), ξ_M and ξ_H are the cost matrices of the Wasserstein distances W_{γ_M} and W_{γ_H} respectively, λ is a trade-off

parameter linked to the Wasserstein distances and $\mathbf{a}_M$, $\mathbf{a}_H$, $\mathbf{b}_M$ and $\mathbf{b}_H$ are scaling vectors.

The updating of the scaling vectors $\mathbf{a}_M$, $\mathbf{a}_H$, $\mathbf{b}_M$ and $\mathbf{b}_H$ for the multispectral and the hyperspectral components can be done in parallel. Their computation is quite convenient for it is carried out with element-wise operations (divisions and multiplications) which are parallelized in Matlab.

The fusion model (2.19) contains five parameters that can be adjusted and that allow the evaluation of the optimal performances of the fusion algorithm. In order to assess the best performance of the fusion algorithm, the optimal parameters should to be found. The search of these latter is done by looking for the set of parameters that gives the minimum of problem (2.19) in terms of objective measures. Finding the optimal parameters is also a fair and objective way of comparing the performance of our fusion algorithm with other state-of-the-art fusion methods.

For each parameter, a set of testing values was determined based on manual adjustment to determine roughly an upper and a lower bound for each parameter. Thus, the search for the optimal parameters was carried out with an exhaustive search in the space of possible values. The exhaustive search for optimal parameters was carried out on three data sets: Pavia, Chikusei and Urban, and the optimal values are $\alpha_{M_{opt}} = 10^{-4}$, $\gamma_{M_{opt}} = 10^{-3}$, $\alpha_{H_{opt}} = 10^2$, $\gamma_{H_{opt}} = 1$ and $\lambda_{opt} = 0.1$. From these optimal parameters we can notice that $\alpha_{H_{opt}} >> \alpha_{M_{opt}}$ which means that the spectral distance for the hyperspectral component has a more important weight than the one for the multispectral component. Then, the spectral information in the hyperspectral image is more important than the one in the multispectral image. Thus, $\alpha_{H_{opt}}$ being important means a high cost and a penalized transport in the spectral domain for the hyperspectral component which leads to the conservation of the spectral information.

We notice that $\gamma_{M_{opt}} << \gamma_{H_{opt}}$, which means that the Wasserstein distance linked to the multispectral image is less regularized than the one linked to the hyperspectral image. That could be interpreted by the fact that an important regularization coefficient $\gamma_{M_{opt}}$ would degrade the valuable spatial resolution of the multispectral image, because the latter needs less regularization. By the contrary, the hyperspectral image is already quite regularized spatially because of its generation process, thus, it is more tolerant to greater regularization coefficients. Finally, the coefficient $\lambda = 0.1$ shows that the Wasserstein distance linked to the hyperspectral image has more weight $(1 - \lambda = 0.9)$ than the one linked to the multispectral image. This means that the hyperspectral image contains an important quantity of information that needs to be preserved and that the fused image **u** has to contain mainly the information provided by the hyperspectral image.

2.7   Evaluation methods

The aim of the fusion is to produce a high-resolution hyperspectral image from two low-resolution images. The fusion result can be evaluated visually but it is not a reliable and objective way of measuring the performance of an imaging algorithm. Thus, many objective measures exist and are widely used to measure the performance of a method in an objective way. Let $\mathbf{X} \in \mathbb{R}^{H \times N}$ be a reference image such that $\mathbf{X}=[\mathbf{x}_1, \ldots, \mathbf{x}_H] = [\mathbf{x}^1, \ldots, \mathbf{x}^N]$ where $\mathbf{x}_k \in \mathbb{R}^{1 \times N}$ $(k \in \{1, \ldots, H\})$ is the kth band of $\mathbf{X}$ and $\mathbf{x}^i \in \mathbb{R}^{H \times 1}$ $(i \in \{1, \ldots, N\})$ is the ith pixel of $\mathbf{X}$. Let $\widehat{\mathbf{X}}$ be the estimated image. The quality measures used in this work to evaluate the fusion results are the following

1. **PSNR**: The *PSNR* (Peak Signal-to-Noise Ratio) measures the quality of the spatial reconstruction of each hyperspectral band. The *PSNR* is the ratio

between the maximum power of the kth band of the reference image and the residual error between the kth band of the reference and the estimated image. For the kth band, the *PSNR* is computed as follows

$$\text{PSNR}(\mathbf{x}_k, \widehat{\mathbf{x}}_k) = 10.\log_{10}\left(\frac{\max{(\mathbf{x}_k)^2}}{\|\mathbf{x}_k - \widehat{\mathbf{x}}_k\|_2^2/N}\right),$$

where $\max{(\mathbf{x}_k)}$ is the maximum pixel value of the kth band for the reference image $\mathbf{X}$. The residual error is normalized for each band and thus it is not affected by data values which allows fair comparison between the bands. The final value of the *PSNR* is the average of the *PSNR* value of each band.

2. **SAM**: The *SAM* (Spectral Angle Mapper) [126] measures the quality of the spectral reconstruction by computing the angle between two spectral vectors at each pixel in the reference and in the reconstructed image. For two spectral vectors $\mathbf{x}^i$ and $\widehat{\mathbf{x}}^i$ the *SAM* is measured as follows

$$\text{SAM} = \arccos\left(\frac{<\mathbf{x}^i, \widehat{\mathbf{x}}^i>}{\|\mathbf{x}^i\|_2 \|\widehat{\mathbf{x}}^i\|_2}\right).$$

The values of the *SAM* belong to $(-90, 90]$ and they are measured in degrees, the smaller the absolute value of the *SAM* the weaker the spectral distortion and the higher the spectral quality of the fusion. The final *SAM* value is computed by averaging all the *SAM*s of the pixels of the image.

3. **RMSE**: the *RMSE* (Root Mean Square Error) measures the L^2 error between the original image $\mathbf{X}$ and its estimation $\widehat{\mathbf{X}}$ as follows

$$\text{RMSE}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{\|\widehat{\mathbf{X}} - \mathbf{X}\|_F}{\sqrt{N \times H}},$$

where $\|\mathbf{X}\|_F = \sqrt{\text{trace}(\mathbf{X}^\mathsf{T}\mathbf{X})}$ is the Frobenius norm of $\mathbf{X}$. The ideal value of

the RMSE is 0.

4. **ERGAS**: *ERGAS* (Erreur Relative Globale Adimentionelle de Synthèse) [127] measures the global fusion quality. It is calculated as follows

$$\text{ERGAS}(\mathbf{X}, \widehat{\mathbf{X}}) = 100d \sqrt{\frac{1}{H} \sum_{k=1}^{H} \left(\frac{\text{RMSE}_k}{\mu_k} \right)^2},$$

where d is the ration between the spatial resolution of the multispectral image and the one of the hyperspectral image, $\text{RMSE}_k = \frac{\|\widehat{\mathbf{x}}_k - \mathbf{x}_k\|_F}{\sqrt{N}}$ and μ_k is the mean of the kth band of **X**. The ideal value of ERGAS is 0.

5. **CC**: The Cross Correlation (CC) is defined as follows

$$\text{CC}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{1}{H} \sum_{k=1}^{H} \text{CCS}(\mathbf{x}_k, \widehat{\mathbf{x}}_k)$$

where CCS is the cross correlation between two single-banded images and it is defined as follows

$$\text{CCS}(\mathbf{A}, \mathbf{B}) = \frac{\sum_{j=1}^{N}(\mathbf{A}_j - \mu_A)(\mathbf{B}_j - \mu_B)}{\sqrt{\sum_{j=1}^{N}(\mathbf{A}_j - \mu_A)^2 \sum_{j=1}^{N}(\mathbf{B}_j - \mu_B)^2}}$$

where $\mu_A = \frac{1}{N} \sum_{j=1}^{N} \mathbf{A}_j$ is the mean of **A**. The ideal value of CC is 1.

6. **DD**: The measure of Degree of Distortion between two images **X** and $\widehat{\mathbf{X}}$ is defined as follows

$$\text{DD}(\mathbf{X}, \widehat{\mathbf{X}}) = \frac{1}{N \times H} \|\text{vec}(\mathbf{X}) - \text{vec}(\widehat{\mathbf{X}})\|_1$$

where $\text{vec}(\mathbf{X})$ and $\text{vec}(\widehat{\mathbf{X}})$ represent the vectorization of the image **X** and $\widehat{\mathbf{X}}$), respectively. The ideal value of DD is 0.

7. **Q**: The Q index was suggested by Wang and Bovik [128] to evaluate the similarity between two single-band images. It measures their distortion as the product of loss of correlation, luminance distortion and contrast distortion. The Q index between two single-band images **A** and **B** is defined as follows

$$Q(\mathbf{A}, \mathbf{B}) = \frac{4\sigma_{AB}^2 \mu_A \mu_B}{(\sigma_A^2 + \sigma_B^2)(\mu_A^2 + \mu_B^2)} \tag{2.25}$$

where $(\mu_A, \mu_B, \sigma_A^2, \sigma_B^2)$ are the means and the variances of **A** and **B**, respectively, and σ_{AB} is the covariance of (**A**,**B**). The range of Q is $[-1, 1]$ and the ideal value of Q is 1.

2.8   Experimental results

This section studies the performances of the HMWB (Hyperspectral and Multispectral Wasserstein Barycenter) fusion algorithm. The experiments are carried out with dataset acquired by the ROSIS optical sensor over the urban area of the University of Pavia in Italy with a resolution of 1.3 m. The image was initially composed of 115 bands and only 93 bands were left after the removal of water absorption bands. The spectral range is from 430 to 860 nm. For the experiments, Urban dataset, acquired by the HYDICE imaging sensor with a ground sampling distance of 2 m, is also used. The image has a spatial size of 307x307 pixels and 210 wavelengths ranging from 400 nm to 2500 nm with a spectral resolution of 10 nm. After removing bands due to dense water vapor and atmospheric effects, 162 spectral bands were left.

2.8.1 Simulation scenario

In this section we suggest to infer a high-resolution hyperspectral image from a high-spatial resolution but low-spectral resolution image (MS) and a high-spectral resolution but low-spatial resolution image (HS). The hyperspectral image is generated by applying a cyclic convolution operator that acts on each band followed by subsampling by a factor of 4. The 4-band multispectral image is generated by filtering the reference image with the IKONOS-like reflectance spectral response as shown in Figure 2.9. The hyperspectral and the multispectral images are both contaminated by zero-mean additive Gaussian noises. In the case of the hyperspectral image, the SNR value is taken 35dB for the first 43 bands and SNR=30dB for the remaining 50 bands. In the case of the multispectral bands, the SNR=30dB is chosen for all the bands.
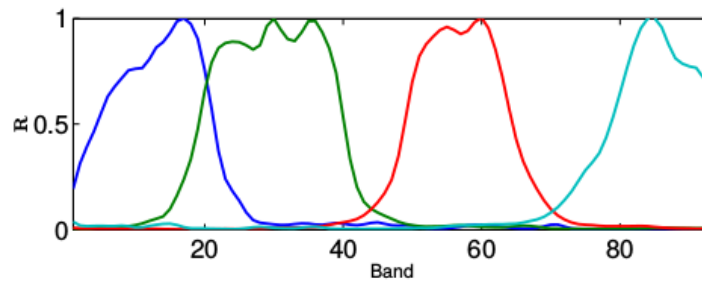
Figure 2.9: IKONOS-like spectral response [62]

In optimal transport, the manipulated images have to be assimilated to probability measures, which means that all their values have to be positive and they should sum to one. Thus, the entry images $\widetilde{\mathbf{S}}(\mathbf{f})$ and $\widetilde{\mathbf{T}}(\mathbf{g})$ defined in (2.19) are normalize to one which means that each value of each image is divided by the sum of the values of the whole image. The values of the output image of the HMWB algorithm sum to one, so in order to compare the performances of the HMWB algorithm with other state-of-the-art methods, the output of each

method is normalized. For the visualization of multi-band images three bands are needed, thus, for the experiments, three arbitrary bands are chosen each time to plot different figures.

2.8.2 Evaluation of results

The fusion model defined in (2.19) does not take into account the removal of noise. In order to lower the effect of the latter, a denoising pre-processing step is applied to the hyperspectral and multispectral images. The denoising methods are chosen in order to provide good results in terms of quality measures (RMSE, SAM etc). For the denoising of the multispectral image, the Wiener filter is used. For the hyperspectral image, a denoising method based on the total variation [129] is chosen. This combination of denoising methods gave better results than when inverting them or imposing the same denoising method to both images. When the Wiener filter is used on the hyperspectral image, it introduces a periodic and piecewise constant structure, which is preserved after the interpolation operation. Therefore, the denoised hyperspectral image contains undesired structures which introduces errors. When the total variation method introduces some sort of diffusion in the image and thus degrades the spatial quality of the latter. Therefore, when this method is applied to the multispectral image, the spatial details are not correctly constructed, which leads to degraded results.

For a fair comparison, all state-of-the-art methods are provided with the same denoised hyperspectral and multispectral dataset.

The experiments are carried out with the Pavia dataset with size $256 \times 256 \times 93$ and Urban dataset with size $196 \times 256 \times 93$. The visual results on Figures 2.10 and 2.13 show that, for both datasets, the HMWB algorithm successfully reconstructed high-resolution hyperspectral images that are visually close to

the reference images. Some results of the state-of-the-art methods suffer from staircasing effects such as MAPSMM and SFIMHS (Figure 2.10), from artifacts like the blue spot on the green vegetation in the GSA method in Figure 2.10 or from bright green surroundings like in the HySure method as shown in Figure 2.13. Unlike these methods, the proposed method provided a result which is visually clean from undesired visual effects in both datasets. Generally, the visualization of the fused image is sufficient to give a first idea about the performance of the fusion algorithm in inferring a good visual result. However, it is important to visualize the errors of the fusion algorithm in order to spot the areas where the fusion could not give satisfying results. That is why, the mapping of the RMSE and the SAM errors were used in order to reveal the distribution of the errors across the bands and the pixels.

Figures 2.11 and 2.14 show the mappings of the RMSE errors and Figures 2.12 and 2.15 show the mappings of the SAM errors for Pavia and Urban datasets respectively. The RMSE images visualize the magnitude of the spatial error at each pixel and the SAM images visualize the spatial distribution of the spectral errors. For the Pavia dataset, we can see that in terms of the RMSE (Figure 2.11) and SAM error mappings (Figure 2.12), the fusion result of the proposed method has fewer errors across the image compared to the other methods, which shows that the spatial and spectral reconstruction at each pixel was correctly done. All state-of-the-art methods show errors in some parts of the image especially where there is the roof of the big building which contains bright and saturated colors. The proposed method however dealt successfully with the saturation of colors in the image. For the Urban dataset, we can notice that in terms of RMSE and SAM errors (Figure 2.14 and Figure 2.15), the errors of the suggested model are lower with respect to the other methods. The error mappings of the results of most state-of-the-art methods show important errors in areas with contours

and fine details which is not the case for the proposed method.

Visual evaluation of the quality of images is a first approach that gives an overall idea about the performance of the fusion algorithm. However, when it comes to conducting objective assessments or comparing with other state-of-the-art methods, visual evaluations become subjective and they are not reliable anymore. That is why many quality indexes were suggested and detailed in Section 2.7 in order to asses the performance of the fusion algorithms. Tables 2.1 and 2.2 show the performances of different fusion methods on Pavia and Urban datasets respectively, evaluated with different quality measures and compared to six methods of the state of the art, the best performances are emphasized in bold type and the second best ones are shown with underlines. From these two tables we can see that the performances of the suggested HMWB algorithm overtake the other methods in all the quality measures.

	RMSE	PSNR	ERGAS	SAM	Q2^n	CC	DD
Reference	0	∞	0	0	1	1	0
CNMF	1.3354	31.7053	1.9882	2.7615	0.9333	0.9559	0.8909
HySure	1.5932	30.1725	2.3947	3.7058	0.9234	0.9422	1.0129
GSA	<u>0.9741</u>	<u>34.4461</u>	<u>1.5409</u>	<u>2.4881</u>	<u>0.9619</u>	<u>0.9772</u>	<u>0.71286</u>
SFIMHS	1.8354	28.9427	2.8866	2.7428	0.8915	0.9125	1.1751
GLPHS	1.7309	29.4522	2.7277	2.7175	0.9022	0.9213	1.1349
MAPSMM	2.1215	27.6848	3.3343	3.0538	0.8423	0.8792	1.3712
HMWB	**0.6846**	**37.5101**	**1.0017**	**1.8691**	**0.9771**	**0.9868**	**0.4738**

Table 2.1: Quality measures of fusion methods on Pavia dataset with size $256 \times 256 \times 93$ corrupted with Gaussian noise. The RMSE and DD values are provided in order of magnitude of 10^{-8}.

2.9 Including the nonlocal means weights

In this part, a new way of carrying out HS-MS fusion is suggested. The novelty is the introduction of nonlocal means (NLM) weights in the cost function.
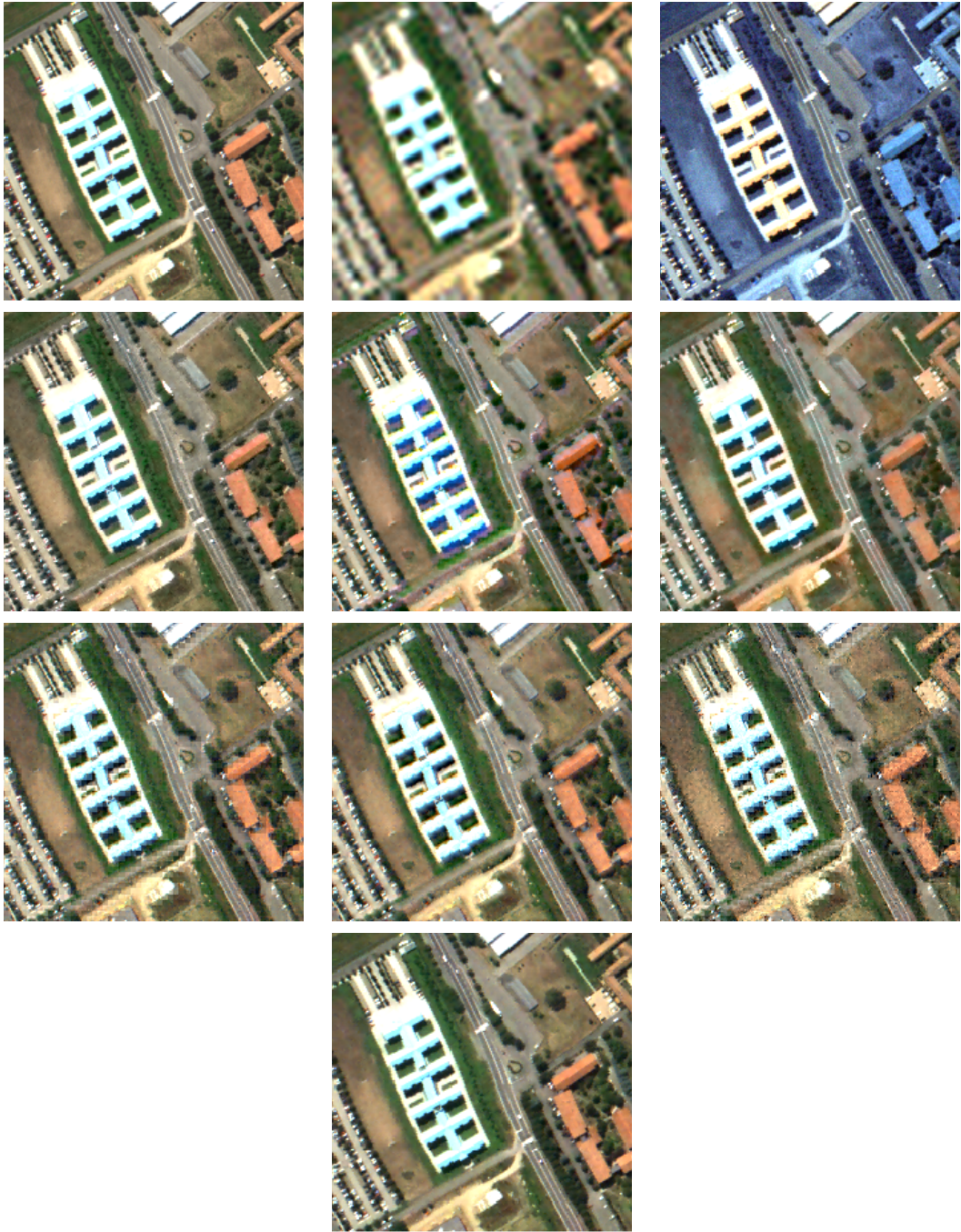
Figure 2.10: Performances of HS-MS fusion of the HMWB algorithm and comparison with other state of the art methods on Pavia dataset of size $256 \times 256 \times 93$. From top left to bottom right: Reference image, HS image, MS image, CNMF [80], HySure [61], GSA [72], SFIMHS [71], GLPHS [73], MAPSMM [75] and the result of the proposed method.
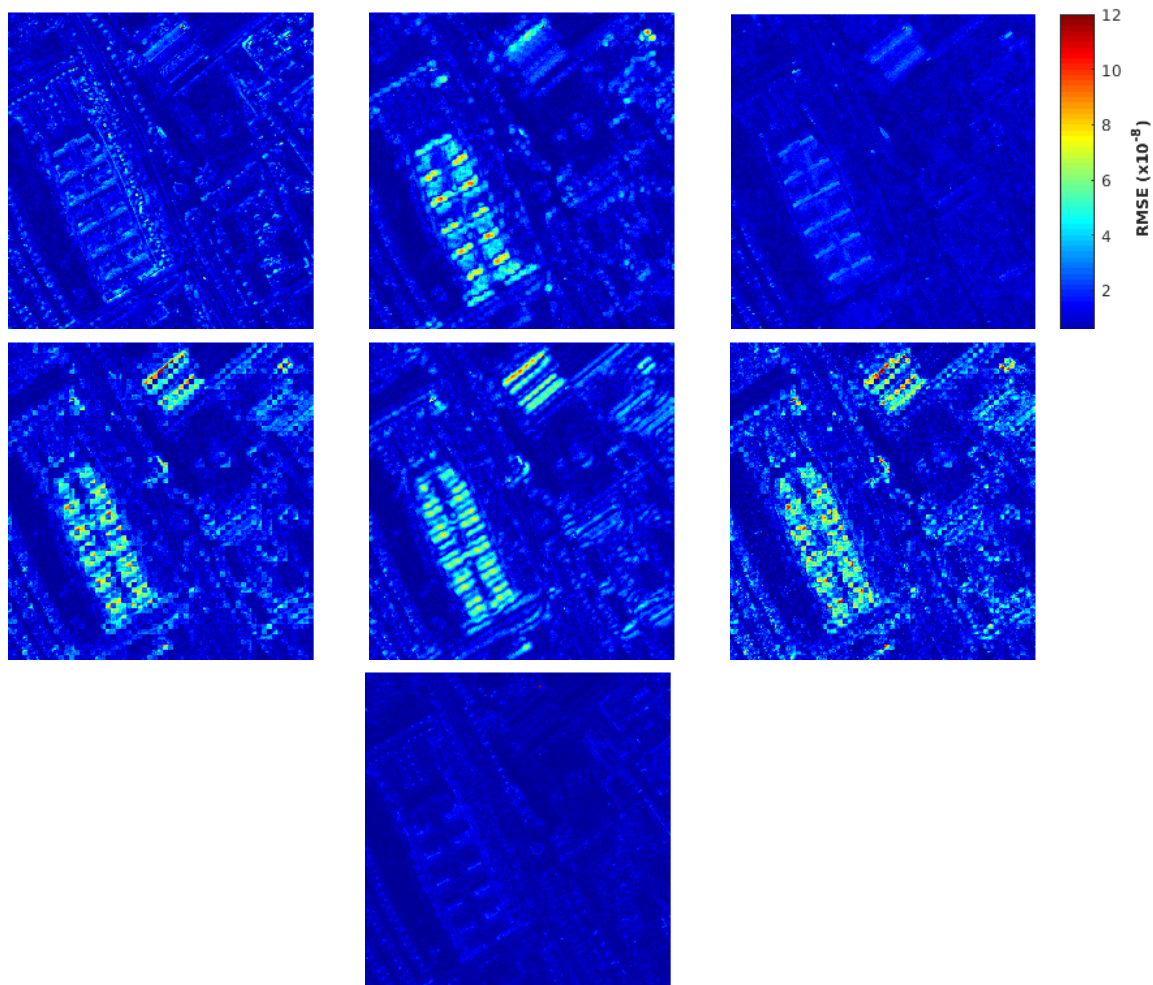
Figure 2.11: Mapping of the RMSE error computed for each pixel for Pavia dataset of size $256 \times 256 \times 93$. From top left to bottom right: CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM and the result of the proposed method.
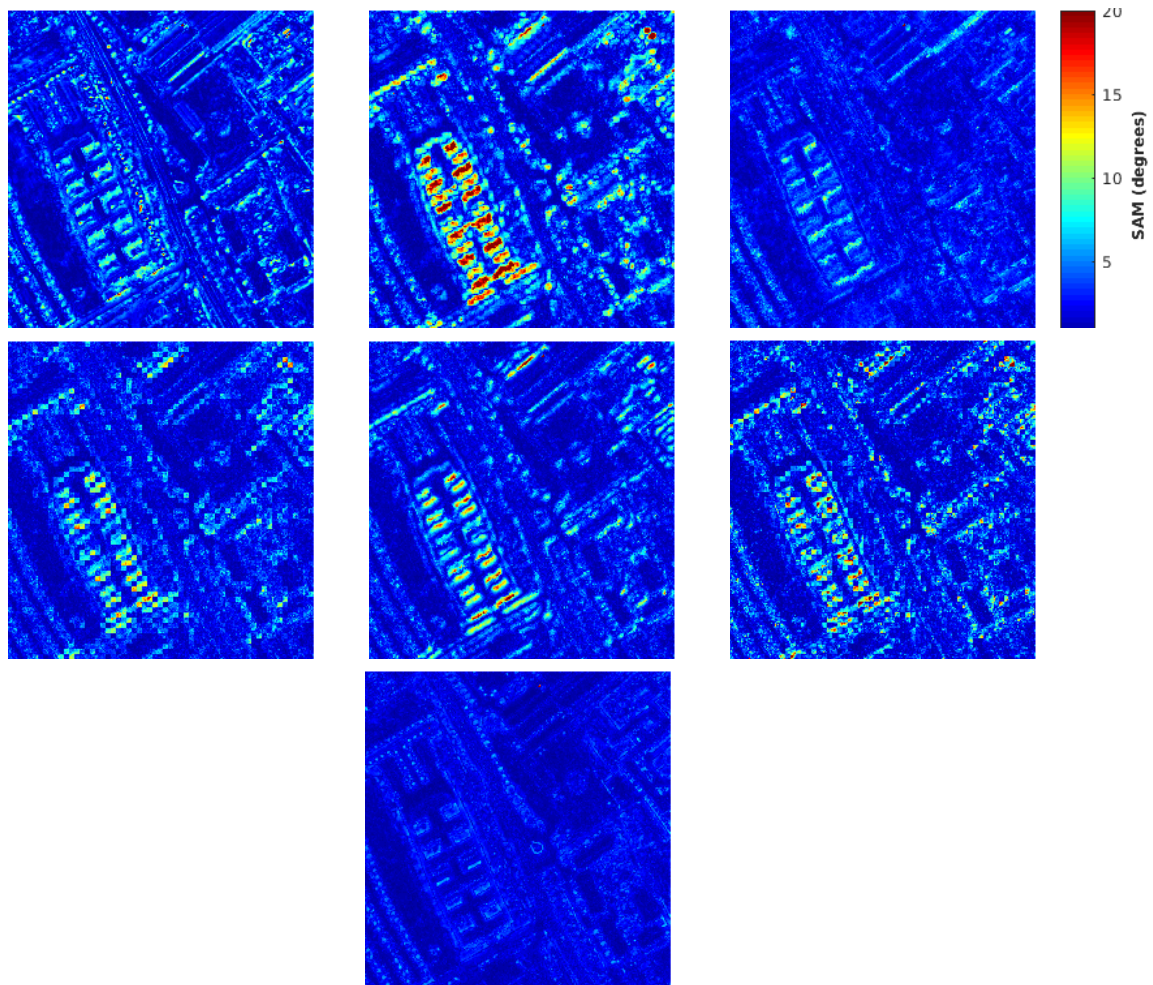
Figure 2.12: Mapping of the SAM error computed for each pixel for Pavia dataset of size $256 \times 256 \times 93$. From top left to bottom right: CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM and the result of the proposed method.
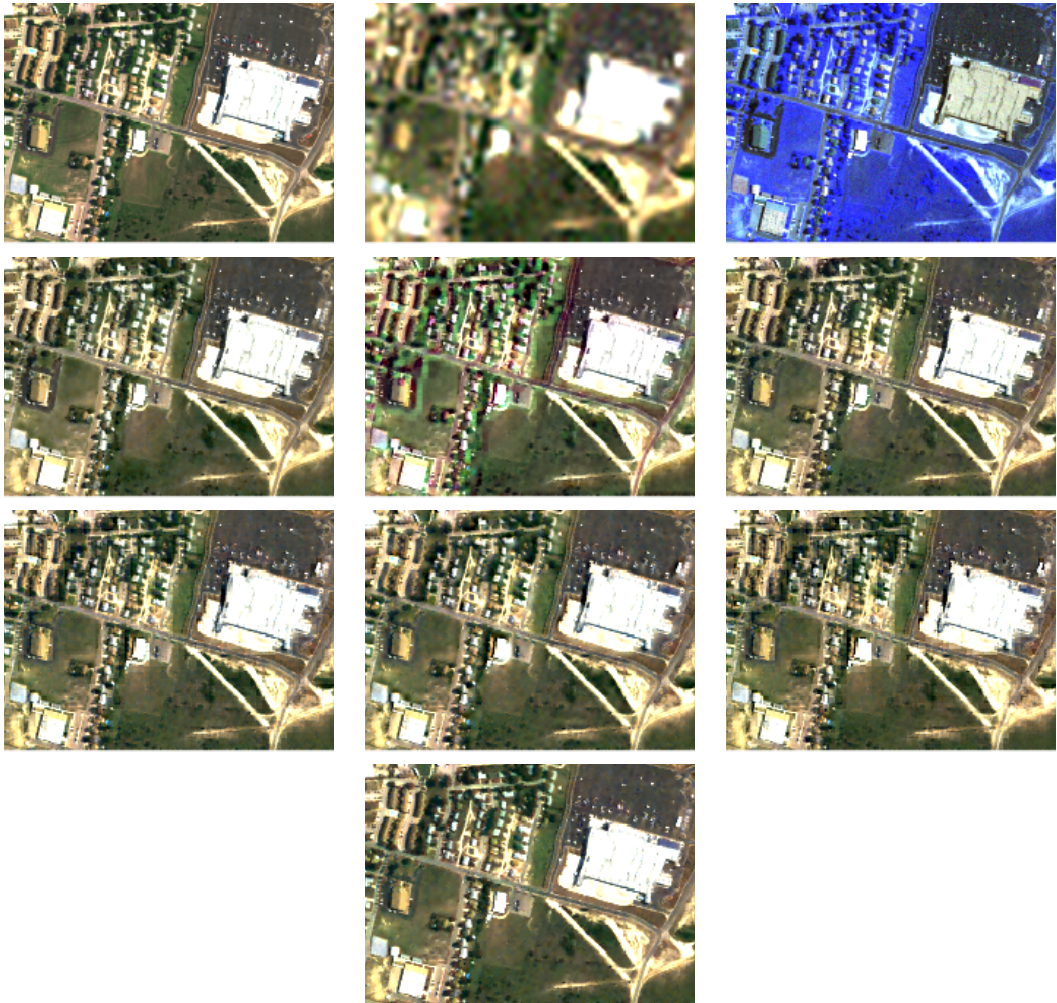
Figure 2.13: Performances of HS-MS fusion of the HMWB algorithm and comparison with other state of the art methods on Urban dataset of size $196 \times 256 \times 93$. From top left to bottom right: Reference image, HS image, MS image, CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM and the result of the proposed method.
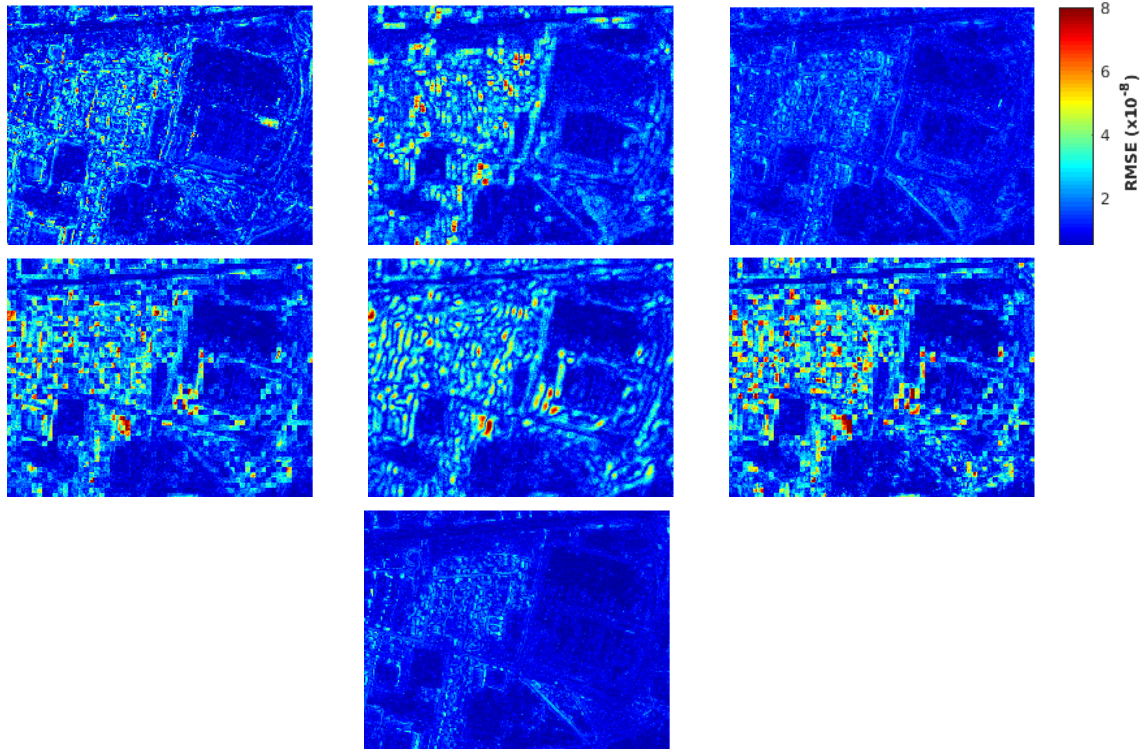
Figure 2.14: Mapping of the RMSE error computed for each pixel for Urban dataset of size $196 \times 256 \times 93$. From top left to bottom right: CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM and the result of the proposed method.

	RMSE	PSNR	ERGAS	SAM	$Q2^n$	CC	DD
Reference	0	∞	0	0	1	1	0
CNMF	1.4276	31.5846	1.5686	<u>2.4071</u>	0.9495	0.9707	0.9888
HySure	1.8618	29.2777	2.1101	3.5705	0.9238	0.9524	1.3074
GSA	<u>1.2036</u>	<u>33.0671</u>	<u>1.3039</u>	2.4963	<u>0.9669</u>	<u>0.9819</u>	<u>0.8419</u>
SFIMHS	2.0421	28.4748	2.2935	2.6032	0.9107	0.9372	1.4053
GLPHS	1.8886	29.1533	2.1251	2.5624	0.9221	0.9454	1.3365
MAPSMM	2.3191	27.3702	2.5999	2.8305	0.8774	0.9163	1.5855
HMWB	**1.0147**	**34.5511**	**1.1106**	**2.0377**	**0.9735**	**0.9855**	**0.6911**

Table 2.2: Quality measures of fusion methods on Urban dataset with size $196 \times 256 \times 93$ corrupted with Gaussian noise. The RMSE and DD values are provided in order of magnitude of 10^{-8}.
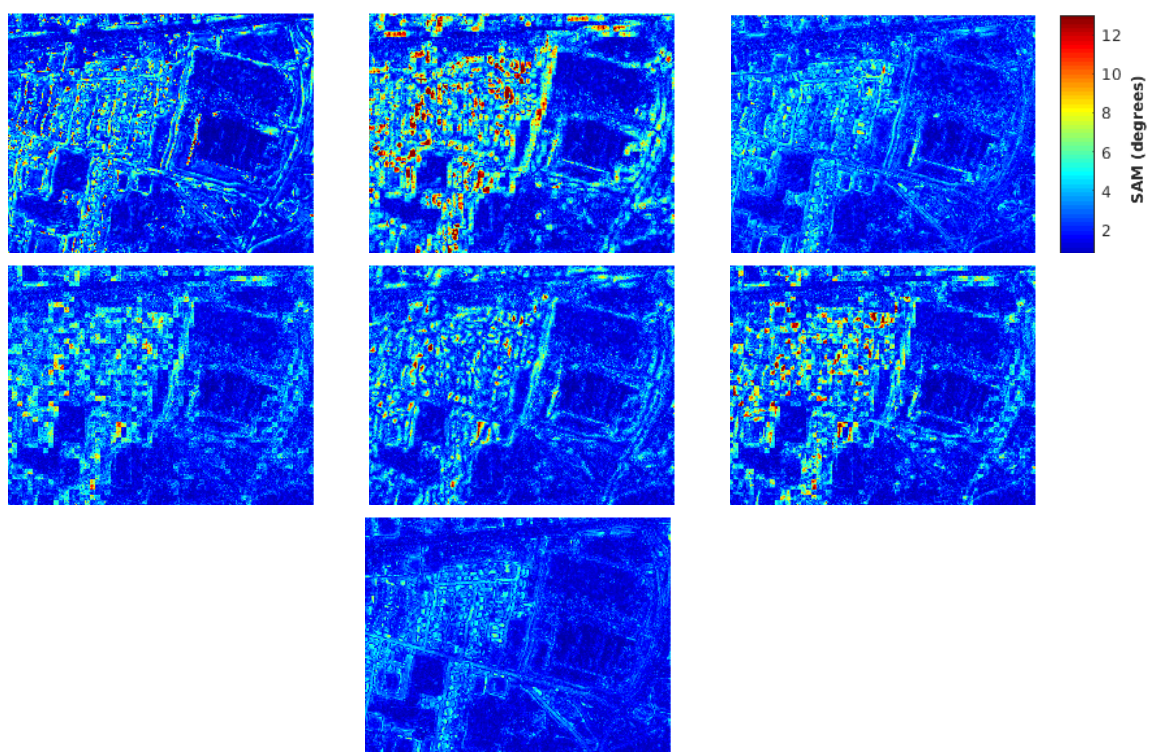
Figure 2.15: Mapping of the SAM error computed for each pixel for Urban dataset of size $196 \times 256 \times 93$. From top left to bottom right: CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM and the result of the proposed method.

This means that the cost function is computed based on the spatial distances between the pixels and on their values as well. The idea of adding nonlocal terms is inspired from the work on nonlocal filtering for image denoising by Buades, Coll and Morel [130]. The authors suggested a NLM algorithm for image denoising by averaging similar pixels in the image and it proved efficient for reducing the noise. The success of the NLM algorithm is related to the computation of the similarity between pixels by comparing patches centered on them and not only comparing their intensity values. Let Ω be an arbitrary space and let u be the image to be denoised. Given a pixel $x \in \Omega$, the filtered or denoised value at x by the NLM algorithm suggested in [130] is

$$
NL[u](x) = \frac{1}{\gamma}(x) \int_{\Omega} \exp\left(-\frac{(\kappa_\rho * |u(x+.) - u(y+.)|^2)(0)}{h^2} \right) u(y) dy, \ \forall x \in \Omega
$$

where κ_ρ is a Gaussian kernel of size ρ and h acts like a filtering parameter that controls the decay of the exponential function and quantifies the speed of decrease of the latter when the dissimilarities between the pixels start increasing. Two pixels x and y are similar when the Gaussian window around x looks like the Gaussian window around $y$. This method removes the noise in the image while preserving the geometric and texture features of the image based on the principle that the images are self-similar.

Nonlocal methods allow the interaction between any two pixels in the image. Unlike the local case where only the closeness between pixels is used and matters, the nonlocal methods search for points in all the image with similar neighborhoods. The nonlocal search allows finding pixels with similar geometry and texture and it it based on the hypothesis on the regularity and self-similarity of the image. Which means that for each pixel in the image, a patch centered on that pixel has many similar patches in the image as shown in Figure 2.16.

Figure 2.16: Similarity of natural images. The different coloured patches indicate the similar parts in the image. We can notice that for each square, many similar squares can be found in different parts of the image.

2.9.1 The new cost function

The previous cost function between two pixels i and j defined in expression (2.8), denoted by $\xi^{\text{dist}}(i,j)$, is computed with the spatial positions of the pixels. The novelty in this part concerns the cost function where the nonlocal means (NLM) weights are included. The new suggested cost function is defined as

$$\xi = \xi^{\text{dist}} \odot \xi^{\text{NLM}} \tag{2.26}$$

where ξ^{NLM} is the NLM cost kernel and $\odot$ is an element-wise multiplication. Following the same storage scheme as the kernel ξ^{dist}, each row of ξ^{NLM} contains the NLM weights of a pixel with respect to all the other pixels in the search perimeter which will be defined later. Let i and j be two pixels, the NLM weight between these two pixels is defined as follows [131]

$$\xi^{\text{NLM}}(i,j) = \exp(-\frac{\|P(i) - P(j)\|_2^2}{h_{\text{sim}}^2}), \tag{2.27}$$

where $P(k)$ is the patch centered on the pixel k and h_{sim} is a filtering parameter. The similarity measure between the patches in $\mathbb{R}^3$ is computed as follows

$$\|P(i) - P(j)\|_2^2 = \frac{1}{d} \sum_{\{z \in \mathbb{Z}^3 : \|z\|_\infty \leqslant d\}} (u(i+z) - u(j+z))^2, \qquad (2.28)$$

where d is the whole size of the patch such that, $d = (2d_s + 1)^2 \times (2d_{s'} + 1)$, with d_s being the radius of the patch in the spatial dimension and $d_{s'}$ is the radius in the spectral dimension. The two parameters d_s and $d_{s'}$ determine if the patch centered on the pixel is a cubic one (covers the spatial and the spectral dimension), a square one (restrained to the spatial dimension) or a "linear" one (covers only the spectral dimension) and u is the image on which the NLM weights are computed. Thus, given two pixels i (x_i, y_i, z_i) and j (x_j, y_j, z_j), ξ is computed as follows

$$\xi(i,j) = \exp\left(-\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) - \frac{\alpha}{\gamma}(z_i - z_j)^2 - \frac{\|P(i) - P(j)\|_2^2}{h_{sim}^2} \right). \quad (2.29)$$

The goal of including the NLM weights is to take into account the values of the pixels during the transport. As we can see in the expression (2.29), the distance between the pixels and the values of these latter are mixed. The idea behind this mixture is to make the decision of transporting the mass between two pixels based on their distances and also on their values. So the new optimal-transport-based fusion can be explained and seen as follows: if two pixels i and j are far from each other, then the distance is going to be quite important. However, if the patches centered on these two pixels are similar and look alike according to the formula (2.28), the energy has to be transferred between these two pixels anyway. If the patches centered on the two pixels are not similar and though the distance between the pixels positions is small, the transport has

to take place anyway. Therefore, the main idea is to de-penalize the transport between pixels that are far and similar and the ones that are close and not very similar.

Mixing the distances between the pixels and their values has been used in many imaging problems such as in the bilateral filtering [132]. The latter inspired many techniques in image processing such as the NLM weights introduced by [133].

2.9.2 Restricting the search window

Changing the cost matrix does not change the algorithm behind the resolution of the fusion problem which is a minimization one. But now that the NLM weights are included, ξ is not Toeplitz anymore, which reduces significantly the tricks that could speed up the multiplications matrix-vector where ξ is implicated. For this reason, other ways of computing and storing ξ were to be found. Since ξ^{dist} is used for the computations of ξ, it has to be stored in a sparse format so that it can fit in the memory space. In short, some values that are very small are discarded because their influence and contribution is extremely small and thus irrelevant.

To make ξ^{dist} sparse, a threshold s was fixed and only the values that were greater or equal to s are kept in ξ^{dist}. However, comparing each time a computed value to "s" before storing it is time consuming and prevent the code from being parallelized. Therefore, restricting the computations in a search window where there is only values greater than s would be very convenient. Nevertheless, one should stress that since the final cost function ξ is the element-wise product of the two matrices ξ^{dist} and ξ^{NLM}, the search window has to be the same for both ξ^{dist} and ξ^{NLM}. Otherwise, the waste of a tremendous amount of time would be caused because of unnecessary computations. Thus, it is better to restrain the

computations of both ξ^{dist} and ξ^{NLM} in the same search window since they are going to undergo element-wise multiplications as specified in expression (2.26).

Thus, with the resolution of simple inequalities, a clever way of restraining the computations to a spatial and spectral window was found as follows

$$
\begin{aligned}
\xi^{\text{dist}}(i,j) = \exp\left(-\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) - \frac{\alpha}{\gamma}(z_i - z_j)^2 \right) &\geqslant s \Leftrightarrow \\
-\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) - \frac{\alpha}{\gamma}(z_i - z_j)^2 &\geqslant \ln(s) \Leftrightarrow \\
(x_i - x_j)^2 + (y_i - y_j)^2 - \alpha(z_i - z_j)^2 &\leqslant -\gamma \ln(s) \Leftrightarrow \\
-\frac{1}{\gamma \ln(s)}\left((x_i - x_j)^2 + (y_i - y_j)^2 + \alpha(z_i - z_j)^2 \right) &\leqslant 1.
\end{aligned}
\tag{2.30}
$$

The expression $\frac{(x_i - x_j)^2}{-\gamma \ln(s)} + \frac{(y_i - y_j)^2}{-\gamma \ln(s)} + \frac{(z_i - z_j)^2}{\frac{-\gamma \ln(s)}{\alpha}} = 1$ is the equation of the ellipsoid centered at (x_i, y_i, z_i) with radiuses (r_x, r_y, r_z) as follows:

$$
\begin{cases}
r_x^2 = -\gamma \ln(s) \Leftrightarrow r_x = \sqrt{-\gamma \ln(s)}, \\
r_y^2 = -\gamma \ln(s) \Leftrightarrow r_y = \sqrt{-\gamma \ln(s)}, \\
r_z^2 = \frac{-\gamma \ln(s)}{\alpha} \Leftrightarrow r_z = \sqrt{\frac{-\gamma \ln(s)}{\alpha}}.
\end{cases}
\tag{2.31}
$$

Thus, for a pixel i with the coordinates (x_i, y_i, z_i), the search zones are $[x_i - r_x, x_i + r_x]$ and $[y_i - r_y, y_i + r_y]$ in the spatial domain and $[z_i - r_z, z_i + r_z]$ in the spectral one. These search domains make sure to keep the values inside the smallest rectangular block that contains the ellipsoid. But in order to stay inside the ellipsoid, it is important to make sure to keep only the values of the rectangular block greater than or equal to s. Figure 2.17 shows a 2D section of the ellipsoid inside the rectangular with the values that should be kept for ξ.

The search perimeter is very important because, on one hand it restrains the computations in a limited area of the image grid. On the other hand, it
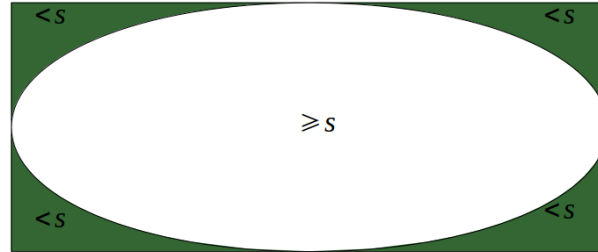
Figure 2.17: Acceptable values in the search perimeter

limits the multiple accesses to the sparse matrix in the memory which are much slower than the accesses to a full matrix.

2.9.3   Computation of the NLM weights

The computations of ξ^{NLM}, as specified in formula (2.27), were carried out in the search window introduced in the Section 2.9.2. The coding of the matrix ξ^{NLM} required six nested loops, and since the coding was done in Matlab, which is known for not getting along with nested loops, the execution of the code was extremely slow. The slowness of the execution made it very hard to test various sets of parameters and assess the performance of the algorithm with the new cost function. Thus, mex engine provided by Matlab was used to recode ξ^{NLM} and carry out the computations in a faster way. In mex engine, the codes are written in the C language that uses the API library, which is a Matlab library that allows coding in C with Matlab commands. Then, the C code is compiled and executed from a Matlab file with a specific command line. Compiling a C code from a Matlab script is very convenient in cases where Matlab and C codes and mixed, which allows using both codes without having to recode everything in Matlab or in C. After converting many parts in Matlab code to C, including ξ^{NLM}, the computations were very fast, as expected, which gave the opportunity of conducting many experiments with different data sets in a

reasonable amount of time.

When only ξ^{dist} was used and before adding the NLM weights to the cost function, there were five parameters (γ_M, γ_H, α_M, α_H, λ) in the fusion model. After adding the NLM weights, three more parameters were added: d_s, $d_{s'}$ and h_{sim}. This means that, in order to evaluate the impact of the new cost function on the performance of the fusion algorithm, these eight parameters should be optimized, which is simply challenging. A classic exhaustive search for optimal parameters was considered with eight nested loops, but the computations never made it to the end because of various technical complications. Another smart way of studying the impact of the eight parameters on the new fusion algorithm was used and it is explained later in the document.

2.9.4 Fusion algorithm

The new HS-MS fusion algorithm is named NLM-HMWB and it is presented below. It is similar to the HMWB algorithm except that the convolution cannot be used anymore with the new cost matrix ξ because it is not Toeplitz. This means that the multiplications matrix-vector cannot be escaped. However, making the matrix ξ sparse by using the threshold s and restraining the search into the ellipsoid simplified the computations.

Algorithm 2: The pseudo-code for the NLM-HMWB algorithm

    Input : $\widetilde{\mathbf{S}}(\mathbf{f})$), $\widetilde{\mathbf{T}}(\mathbf{g})$, ξ_M, ξ_H

1 **Initialization**: $\mathbf{a}_M^{(0)} = \mathbf{a}_H^{(0)} = \mathbf{b}_M^{(0)} = \mathbf{b}_H^{(0)} = \mathbb{1}_{\mathbb{R}^{H \times N}}$;

2 **for** $i \leftarrow 0$ **to** *niter* **do**

3   *// Updating the scaling vectors with data (Sinkhorn's iteration)*

4 $\mathbf{b}_M^{(i)} \leftarrow \dfrac{\widetilde{\mathbf{S}}(\mathbf{f}))}{\xi_M^{\mathsf{T}} \times \mathbf{a}_M^{(i)}}$, $\mathbf{b}_H^{(i)} \leftarrow \dfrac{\widetilde{\mathbf{T}}(\mathbf{g})}{\xi_H^{\mathsf{T}} \times \mathbf{a}_H^{(i)}}$;

5   *// Computing the barycenter*

6 $\mathbf{u} \leftarrow \exp(\lambda \log(\mathbf{a}_M^{(i)} \odot ((\xi_M \times \mathbf{b}_M^{(i)})) + (1-\lambda)\log(\mathbf{a}_H^{(i)} \odot (\xi_H \times \mathbf{b}_H^{(i)})))$;

7   *// Updating the scaling vectors with **u** (Sinkhorn's iteration)*

8 $\mathbf{a}_M^{(i+1)} \leftarrow \dfrac{\mathbf{u}}{(\xi_M \times \mathbf{b}_M^{(i)})}$, $\mathbf{a}_H^{(i+1)} \leftarrow \dfrac{\mathbf{u}}{(\xi_H \times \mathbf{b}_H^{(i)})}$;

9 **end**

    Output: **u** (the fused image)

2.10 Experimental results

2.10.1 Simulation scenario

The experiments were run on three data sets: Pavia, Chikusei and Botswana as shown in Figure 2.18. Pavia was introduced in the Section 2.8. Chikusei dataset [134] was captured by the Headwall Hyperspec-VNIR-C imaging sensor over the agricultural and urban area of Chikusei in Japan. The original data consists of images of size 2517×2335 pixels with a ground sampling distance of 2.5 m and it is formed by 128 bands that range from 363 nm to 1018 nm. The number of bands of the Chikusei dataset was also reduced to 93 after removing noisy and water absorption bands. As to the Botswana dataset, it was acquired by the NASA EO-1 satellite over the Okavango Delta, Botswana. The Hyperion sensor on EO-1 acquired data with a GSD of 30 m over 7.7 km strip in 242 bands

covering the 400-2500 nm portion of the spectrum with a spectral resolution of 10 nm. Uncalibrated and noisy bands that cover water absorption features were removed and 145 bands were left. The preprocessing of the data was performed by the UT center for Space Research.

The generation of hyperspectral and multispectral data was carried out as in Section 2.8.1 with the same sensor. All three data sets (Figure 2.18) have the same size $128 \times 128 \times 93$. The generated hyperspectral and multispectral images from each one of these data sets have been denoised just like in Section 2.8.2.

For the experiments in this part, the number of parameters is eight. Thus, in order to determine the best performance of the new fusion method and compare fairly with state-of-the-art methods, optimal parameters should be found. However, it is difficult to carry out exhaustive search for optimal parameters in the space of all possible parameters ($\mathbb{R}^8$) with eight nested loops. Thus, a new way of looking for optimal parameters was determined in order to reduce the complexity of search. The five parameters of the fusion model when only ξ^{dist} is used were fixed to the optimal ones introduced in Section 2.6. Then, the three parameters linked to the NLM cost matrix were varying. The aim of this search method is to study the sensitivity of the NLM parameters and adapt the search of optimal parameters gradually. The experiments were run for logspaced values of h_{sim} in the interval $[10^{-12}, 10^4]$. The values of h_{sim} are tested for different patch sizes: when only the pixel value is used, 2D patches of sizes 3×3 and 5×5 and 3D patches of sizes $3 \times 3 \times 3$ and $5 \times 5 \times 5$. The algorithm describing the search for the optimal parameters is described in Appendix A, Section A.2. The performances are compared to the case where there is no NLM term contribution, which means where $\xi = \xi^{\text{dist}}$.
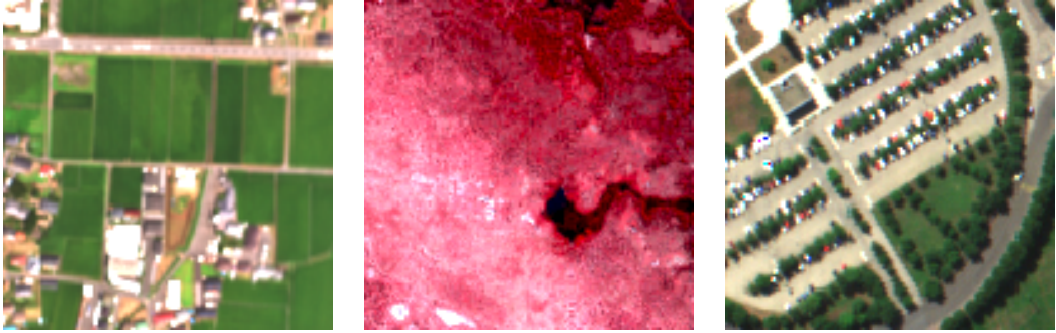
Figure 2.18: Data sets used for the experiments. From left to right. Chikusei, Botswana and Pavia. The size of the three data is 128x128x93

2.10.2 Evaluation of results

The search for the NLM weights takes place in a search area defined by an ellipsoid presented in Section 2.9.2. Based on the fixed optimal parameters defined in Section 2.6 and the fixed threshold $s = 10^{-9}$, the radiuses of the ellipsoid are:

- Hyperspectral component

  - $r_x = 4$

  - $r_y = 4$

  - $r_z = 0$

- Multispectral component

  - $r_x = 0$

  - $r_y = 0$

  - $r_z = 14$

The values of the radiuses show that the search for the hyperspectral component is done spatially in a search window of size 9×9 pixels. The search for the multispectral component is done on the spectral line that has a length of 29 pixels. These search areas are consistent with the principle of transport. In the hyperspectral component we want to keep the rich spectral information and transport only the spatial one, thus the transport has to be done spatially. When it comes to the multispectral component, the transport has to be done spectrally because we want to preserve the valuable spatial information and upgrade the

spectral one.

Figures 2.19, 2.20 and 2.21 show the performance of the sensitivity study of the NLM parameters for the three data sets Chikusei, Botswana and Pavia. Each one of these images display the performance in terms of RMSE, SAM and CC measures. From these figures we can notice that, in all the quality measures and for all data sets, the performance of the fusion algorithm starts with high error which decreases when h_{sim} increases. We also notice that for a specific value of h_{sim}, the performance of the algorithm with NLM terms becomes exactly the same as the performance of the fusion algorithm without the NLM terms.

For the second experiment, the threshold s was reduced to $s = 10^{-3}$ in order to reduce the radiuses of the ellipsoid. A smaller ellipsoid means a restricted search area for NLM weights. The new radiuses of the ellipsoid are

- Hyperspectral component

  - $r_x = 2$,

  - $r_y = 2$,

  - $r_z = 0$,

- Multispectral component

  - $r_x = 0$,

  - $r_y = 0$,

  - $r_z = 8$.

With these new radiuses, the search for the hyperspectral component will take place in a spatial window of size 5×5 pixels and for the multispectral component, the search will occur on the spectral line of lenght 17 pixels. The experiment was carried out on the Chikusei data set. Figure 2.22 shows the performance of the fusion algorithm with the new threshold $s = 10^{-3}$. We can notice that even with smaller search windows for the NLM weights, the performance of the fusion algorithm never gets better than in the absence of the NLM weights.

The use of the NLM weights in the cost matrix was based on the NLM algorithm for image denoising [130]. The hyperspectral and the multispectral

Figure 2.19: Chikusei data set. Sensitivity of the NLM parameters measured with RMSE, SAM and CC quality indices.

Figure 2.20: Botswana data set. Sensitivity of the NLM parameters measured with RMSE, SAM and CC quality indices.
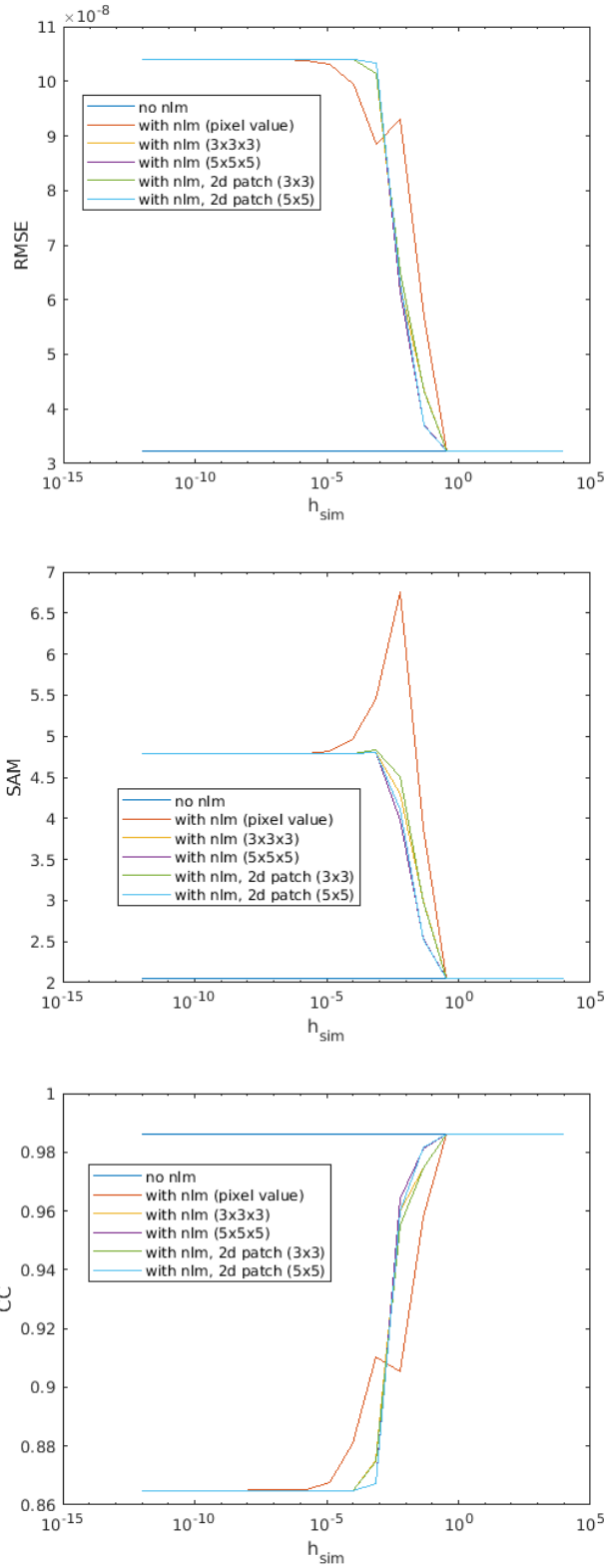
Figure 2.21: Pavia data set. Sensitivity of the NLM parameters measured with RMSE, SAM and CC quality indices.
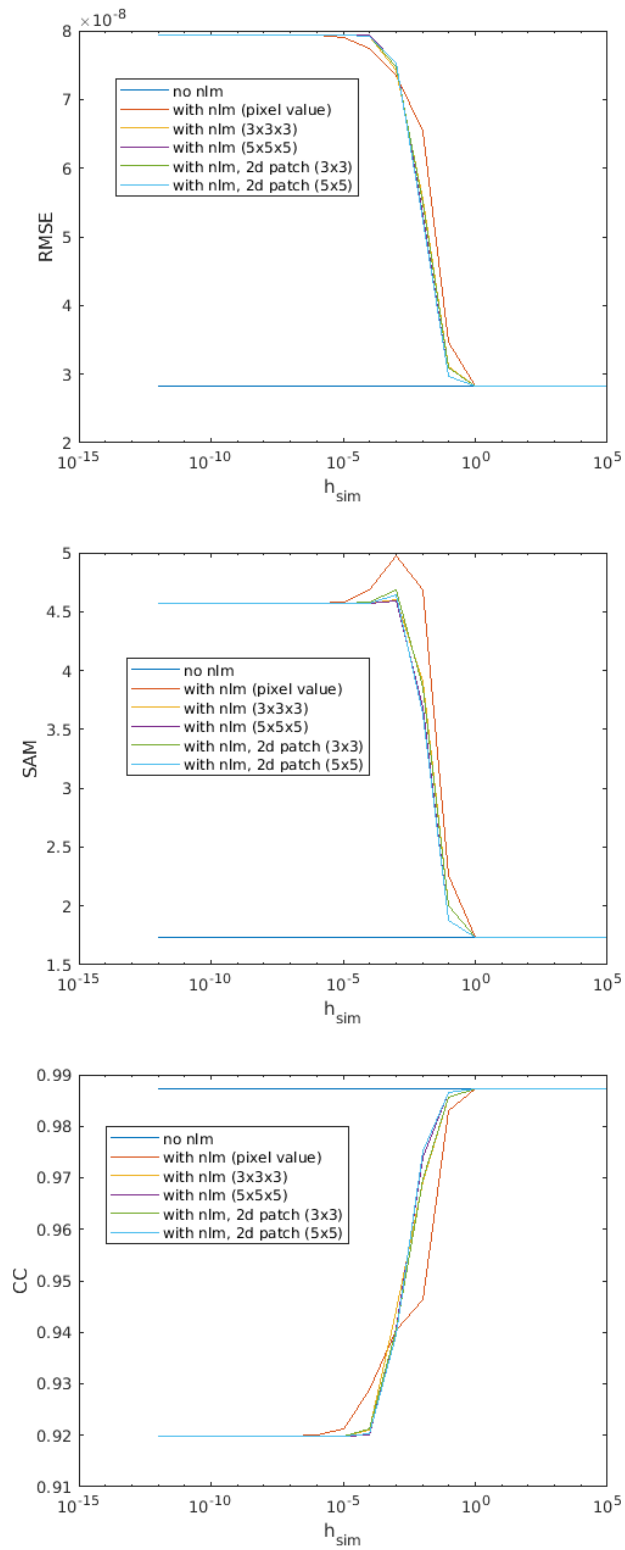
Figure 2.22: Chikusei data set. Sensitivity of the NLM parameters measured with RMSE, SAM and CC quality indices for a threshold of $s = 10^{-3}$.

images are denoised before the fusion. Thus, the denoising properties of the NLM weights could be harnessed to skip the denoising step for the hyperspectral and multispectral images before fusion. The aim of the next experiment is to check if the use of the NLM weights in the cost matrix will lower the effect of noise in the case of noisy hyperspectral and multispectral images. The goal is also to verify if the use of the NLM weights on noisy data will give better results than without the denoising pre-processing step. Figure 2.23 shows the performance of the fusion algorithm without the denoising step. We can notice that we have the same behavior as in the previous experiments: the fusion starts with high errors which decrease when h_{sim} increases and at a specific value of h_{sim}, the performance of the fusion with the NLM weights becomes exactly the same as without the NLM weights.

Figure 2.24 shows the result of the fusion algorithm with and without the denoising step in the case where $\xi = \xi^{dist}$ (no NLM weights). We can notice that the visual result of the fusion is quite noisy when there is no denoising step. Table 2.3 shows the quantitative results of the fusion in the case $\xi = \xi^{dist}$ corresponding to Figure 2.24. We notice that in terms of quality indices, the denoising step plays an important role in minimizing the errors. According to Figure 2.23, the performance of the fusion with the NLM weights do not overtake the result without NLM weight. The best result of the fusion with the NLM weights is obtained when h_{sim} reaches high values and it is the same result when no NLM weights are included. Thus, we conclude that the presence of the NLM weights in the cost matrix does not contribute to removing the noise and does not improve the fusion results.
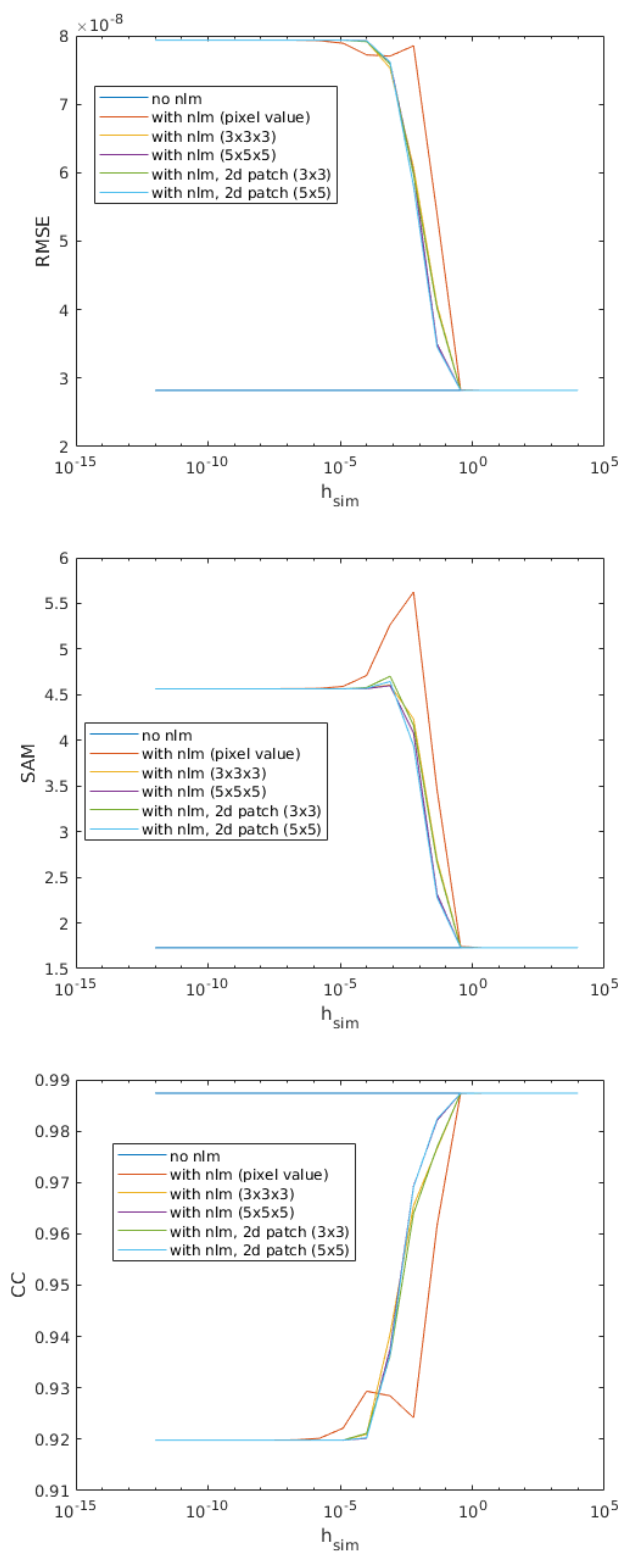
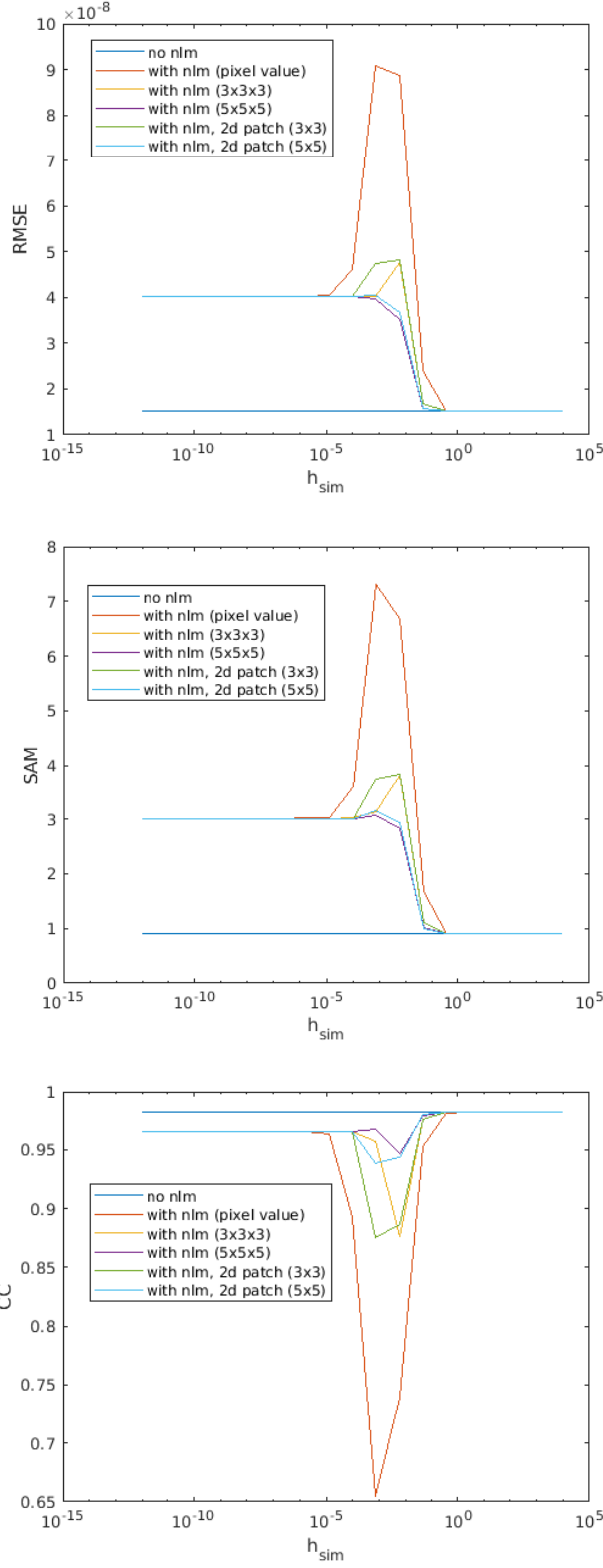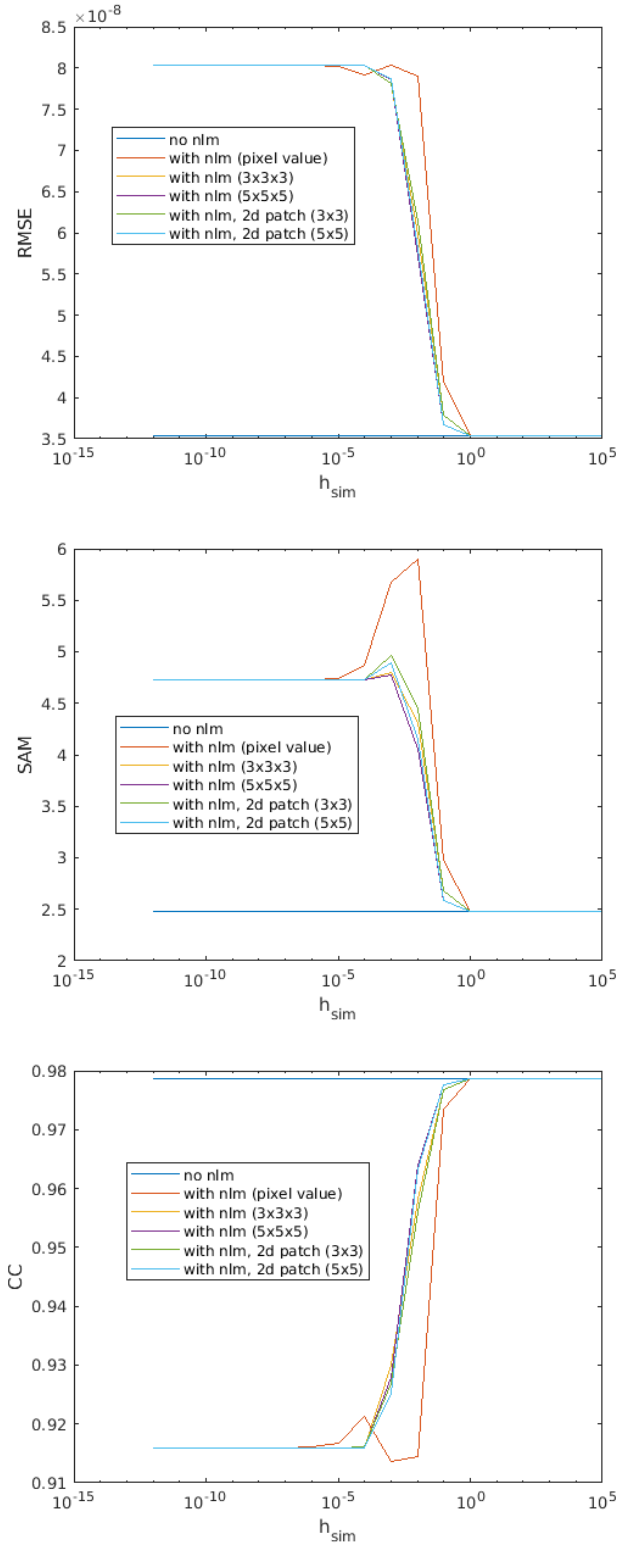Figure 2.23: Chikusei data set. Sensitivity of the NLM parameters measured with RMSE, SAM and CC quality indices for a threshold of $s = 10^{-3}$ and whitout denoising the hyperspectral and the multispectral images.

Figure 2.24: Result of fusion on the Chikusei data set. Left: fusion with denoising. Right: fusion without denoising. Both results are in the case where $\xi = \xi^{\text{dist}}$.

	RMSE	SAM	CC
Reference	0	0	1
Denoising ($\xi = \xi^{\text{dist}}$)	2.8224	1.7318	0.98736
No denoising ($\xi = \xi^{\text{dist}}$)	3.539	2.4737	0.97872

Table 2.3: Quality measures of fusion with and without denoising in the case where $\xi = \xi^{\text{dist}}$. The RMSE are provided in order of magnitude of 10^{-8}.

2.10.3 Analyses of the results

The previous experiments showed that the use of the NLM weights in the cost matrix makes the fusion performance start with high errors in terms of RMSE, SAM and CC for small values of h_{sim}. Then, these errors start decreasing when h_{sim} becomes bigger. The performance of the fusion with the NLM weights reaches the performance of the fusion without NLM without overtaking it.

When h_{sim} is very small, the value of h_{sim}^2 is even smaller which means that $\frac{\|P(i)-P(j)\|_2^2}{h_{\text{sim}}^2}$ has a high value. When the latter increases, the value of the following quantity, that corresponds to the cost of transport, also increases

$$\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) + \frac{\alpha}{\gamma}(z_i - z_j)^2 + \frac{\|P(i) - P(j)\|_2^2}{h_{\text{sim}}^2}.$$

This means that moving the mass in this case is not favored because the transport cost is high. This explains why, when h_{sim} is very small, the fusion is less efficient than in the absence of the NLM weights.

The experiments also showed that the bigger h_{sim} the better the fusion performance with the NLM weights and the closer it gets to the results in the absence of the NLM weights. When h_{sim} takes big values, h_{sim}^2 is even bigger which makes the fraction $\frac{\|P(i)-P(j)\|_2^2}{h_{sim}^2}$ take small values. Therefore, the value of the term $\frac{1}{\gamma}((x_i - x_j)^2 + (y_i - y_j)^2) + \frac{\alpha}{\gamma}(z_i - z_j)^2$ is the one that is mainly implicated in the transport cost, which brings the fusion algorithm to the case where the NLM weigths are absent. Thus, we can deduce that when h_{sim} takes big values, the fusion performance with the NLM weights gets closer to the case where the NLM weights are absent until both performances become exactly the same as shown by the experiments.

Before introducing the NLM weights, the cost matrix was defined by the distance between the pixels positions. This fact implies that, when the pixels are far away from each other, the distance between them is important and thus the transport between these pixels is not favored. But in the case where the pixels are close to each other, the distance between the pixels is small and in this case the transport may take place. Thus, the idea behind the introduction of the NLM weights was to encourage and force the mass move in the case where pixels are far away from each other and are similar. In the case where the patches are similar, the distance $\|P(i) - P(j)\|_2^2$ is small, thus the fraction $\frac{\|P(i)-P(j)\|_2^2}{h_{sim}^2}$ is small too. This means that the term about the distance between the positions of the pixels is preponderant and counts more than the NLM one. Moreover, when the spatial distance between two pixels increases, the probability of finding similar patches is quite low. We can conclude that, either the pixels are far from each other or not, if their patches are similar, the use of the NLM weights does not

much contribute to the transport and has less influence than the spatial distance term.

Finally, the experiments showed that the cost matrix with the NLM weights does not contribute to the reduction of the noise. Because the NLM weights that were used for denoising in the literature acted directly on the image to be denoised. In our case, the NLM terms are included in the cost matrix that is used for the update of the scaling vectors, thus, the image is indirectly affected by the NLM terms.

Throughout these experiments, we saw that the use of the NLM weights does not contribute to upgrading the results of the fusion either in the presence or in the absence of noise. For all the previous experiments, the images were assumed to be registered without any misalignment problems. However, this assumption does not always hold in the real-life conditions. In fact, during the acquisition process of images, misalignment, also called misregistration or geometric noise, affects data covering the same scene. Thus, using the NLM weights in this case might downplay the misalignment effect. Most of HS-MS state-of-the-art fusion methods assume that the hyperspectral and the multispectral images are perfectly aligned. Thus, it is likely that, most of these methods do not take into account the misregistration problems and that the results of the fusion will suffer from geometric noise. Therefore, imposing to the transport to take place, even when the pixels are far from each other spatially but similar in terms of patchs, might decrease the misregistration effect. It is no secret that the misalignment will highly likely not be completely fixed and that the fused image will contain some geometric artifacts. However, the goal is to downplay the geometric noise with respect to the case of transport without NLM weights and with respect to other state-of-the-art methods. This line of research seems promising and will be developed in future work. It might also

open new opportunities for the mixing of NLM weights and optimal transport.

2.11   Concluding remarks

In this chapter, two HS-MS fusion techniques based on optimal transport were developed. In the first part, the introduced fusion model minimized the sum of two weighted and regularized Wasserstein distances. These latter describe the spectral and spatial information to be injected in the final image. The solution of the minimization problem is the barycenter in the sense of the Wasserstein distance, computed with the Hyperspectral and Multispectral Wasserstein Barycenter (HMWB) algorithm developed in Section 2.6.

The Pavia and Urban datasets were used to evaluate the fusion methods. For both datasets, the qualitative results showed that the fused image is visually close to the reference one. The Root Mean Square Error (RMSE) and the Spectral Angle Mapper (SAM) mapping images illustrated the distribution of the spatial and spectral errors. These error mappings showed that errors are either homogeneously small across the image or concentrated in parts characterized with individual and small blocks. Furthermore, the fusion performance of the suggested method was also assessed with various quality measures and compared to six state-of-the-art methods. For both datasets, the suggested method outperformed the other methods in all the quality indices. Overall, visual and numerical results showed that the optimal transport based fusion was able to produce fusion results that compare favorably to the state-of-the-art methods.

In the second part, a new way of carrying out the fusion was suggested. The NLM weights were included in the cost matrix so that during the transport, not only the distance between the pixels is taken into account but also the values of

similar pixels. The goal of including the NLM weights in the cost matrix was to improve the quality of transport and to take into account the additive noise that affect the input data. Due to the high number of parameters implicated in the fusion algorithm (eight), a sensitivity study was conducted in order to identify the behavior of the fusion results with respect to the parameters. Different experiments were carried out on three data sets: Chikusei, Botswana and Pavia. The results of these experiments showed that the NLM weights do not contribute to improving the results of the fusion. In fact, the figures showed that the best results the algorithm with the NLM weights can reach, are the ones without NLM weights. However, the NLM weights might help fix misregistration problems in the case where the hyperspectral and multispectral image are not supposed to be perfectly aligned. The contribution of the NLM weights in downplaying the effect of misalignment will be explored in future work.

Chapter 3

Nonlocal gradient based fusion models

3.1   Introduction to pansharpening methods

Satellite acquire multispectral images at high spatial resolution and subsample the hyperspectral channels by a factor of 2 or 4. The reduction of resolution in satellite images is due to technical constraints of sensors onboard and to the transmission bandwidth. The fusion of hyperspectral and multispectral images produces a high-resolution hyperspectral image with finer spatial and spectral details than any of the entry images.

In this chapter, we introduce nonlocal variational models for HS-MS fusion. The proposed models incorporate a nonlocal regularization term and assume that the low-resolution bands are obtained from the high-resolution ones by low-pass filtering followed by subsampling. We also include a constraint that forces the final image to share the high frequencies of the multispectral image. The suggested fusion models do not need any assumptions on the co-registration

of the bands since each band is fused independently.

Variational methods for image processing problems have been used to tackle many imaging problems and they have enjoyed an explosive success [135, 136, 137, 138, 139] since the pioneering work of Mumfomrd-Shah [140] and Rudin-Osher-Fatemi (ROF) [141]. The goal of variational approaches is to determine an unknown variable that satisfies given constraints. These constraints are formulated as an energy functional and the unknown variable is the solution of the minimization of the energy functional.

The majority of image processing models are formulated as the minimization of ill-posed inverse problems. In order to account for the ill-posedness of these problems, a regularizer is carefully chosen and added to the functional to be minimized. Most of the time, classical variational techniques describe regularity in terms of local relationships of nearby pixels by computing the gradient or the Laplacian. Among the variational techniques, the total variation [141] is the most significant one. Total variation is famous for preserving discontinuities in the image for assigning the same energy cost to sharp and smooth transitions.

The first variational fusion problem was a pansharpening one called $P + XS$ and suggested by Ballester *et al.* [142]. The abreviation $P + XS$ stands for panchromatic and multispectral images and the aim is to merge a panchromatic image that has a high spatial resolution and a low spectral one with a multispectral image that has a low spatial resolution and a high spectral one. In $P + XS$, the formulation of the pansharpening problem consists of the minimization of an energy functional composed of three terms. A regularization term that accounts for the ill-posedness of the fusion problem and forces the spectral channels to share the same geometry as the panchromatic image. The other two terms penalize the deviation from the panchromatic and the multispectral image. The $P + XS$ formulation of the pansharpening problem is presented as follows.

Let Ω be an open and bounded domain in $\mathbb{R}^d$ where $d \geqslant 2$. We denote by $\mathbf{u} : \Omega \to \mathbb{R}^M$, with $\mathbf{u} = (u_1(x), \ldots, u_M(x))$ for any $x \in \Omega$, the high-resolution multispectral image, with M spectral bands, we seek to estimate. Each image $u_k : \Omega \to \mathbb{R}$ represents the intensity values corresponding to the k th band of $\mathbf{u}$. Assume that we are given a high-spatial-resolution panchromatic image $P : \Omega \to \mathbb{R}$ and a low-spatial-resolution multispectral image $\mathbf{u}^S : \Omega_S \to \mathbb{R}$ such that $\mathbf{u}^S(x) = (u_1^S(x), \ldots, u_M^S(x))$, for any $x \in \Omega_S$ where Ω_S is a sampling grid $\Omega_S \subseteq \Omega$.

By supposing that the geometry of the spectral channels is contained in the topographic map of the panchromatic image, the purpose is to reconstruct a high-resolution multispectral image from the two given data P and $\mathbf{u}^S$.

The panchromatic image is supposed to be a linear combination of the bands of the high-resolution multispectral image $\mathbf{u}$ and it is given by

$$P(x) = \sum_{k=1}^{M} \alpha_k u_k, \ \forall x \in \Omega$$

where $\{\alpha_k\}_{k=1}^M$, such that $\alpha_k \geqslant 0$ for all $k \in \{1, \ldots, M\}$, are mixing coefficients for computing the total spectral energy from the channels of the high-resolution multispectral image. In a variational framework, the aim is to minimize

$$\int_\Omega \left(\sum_{k=1}^{M} \alpha_k u_k(x) - P(x) \right)^2 dx, \tag{3.1}$$

We note that the panchromatic image and the spectral channels u_k of the image to be found $\mathbf{u}$ are assumed to be aligned.

In this model, the channels of the multispectral image $\mathbf{u}^S$ are assumed to be obtained from the high-resolution multispectral image $\mathbf{u}$ by low-pass filtering followed by sub-sampling. Thus, we have

$$u_k^S = (K_k * u_k)^{\downarrow s}, \ \forall k \in \{1, \ldots, M\},$$

where $\downarrow_s$ stands for the sub-sampling operator by a factor s (for most satellites $s = 4$), and K_k denotes the impulse response of the low-pass filter corresponding to channel $k \in \{1, \dots, M\}$.

We need to assume that it is possible to evaluate $K_k * u_k$ at any point of Ω_S. For this, K_k is considered to be the kernel of a convolution mapping $L^2(\Omega)$ into $C(\overline{\Omega})$, that is

$$K_k * v(y) = \int_\Omega K_k(y - x) v(x) dx, \ \forall k \in \{1, \dots, M\}, \ \forall v \in L^2(\Omega).$$

The writing of the above constraint in a variational framework is as follows

$$\int_\Omega \Pi_S \cdot (K_k * u_k - u_k^\Omega(x))^2 dx, \ \forall k \in \{1, \dots, M\} \tag{3.2}$$

where $\Pi_S = \sum_{x \in \Omega_S} \delta_x$ is a Dirac's comb defined by the sampling grid Ω_S and $\mathbf{u}^\Omega : \Omega \to \mathbb{R}^M$, with $\mathbf{u}^\Omega(x) = (u_1^\Omega(x), \dots, u_M^\Omega(x))$ for any $x \in \Omega$, is an arbitrary extension of $\mathbf{u}^S$ as a continuous function from the sampling grid Ω_S to the whole domain Ω. Note that the integral of a sum of Dirac's is unambiguous as we assume that no point of Ω_S belongs to the boundary of Ω. Moreover, the integrand term is multiplied by Π_S and the expression (3.2) does not depend on the particular chosen extension from Ω_S to Ω.

Before talking about the last term in the $P + XS$ model, some concepts need to be introduced. Given a function w that belongs to $L^1(\Omega)$, the total variation of w in Ω is defined as follows

$$V(w, \Omega) = \sup \left\{ \int_\Omega w(x) \operatorname{div} \phi(x) dx : \phi \in C_0^\infty(\Omega, \mathbb{R}^2), |\phi| \leqslant 1 \right\},$$

The space of functions of bounded variation (BV functions) can then be defined as

$$BV(\Omega) = \left\{ w \in L^1(\Omega) : V(w, \Omega) < +\infty \right\}.$$

In order to account for the ill-posedness of the pansharpening problem, the following term was suggested

$$\sum_{k=1}^{M} \gamma_k \int_{\Omega} |\theta^{\perp} \cdot \nabla u_k(x)|^2 dx.$$

This term imposes that each channel at the high resolution follows the geometry of the panchromatic image. The coefficients $\gamma_k > 0$ allow the control of the relative weight assigned to each channel, since the authors chose not the privilege any channel over the other ones, they assigned $\gamma_{k\,k=1}^{M} = 1$. The variable θ represents a vector field $\theta : \Omega \to \mathbb{R}^2$ such that

$$\forall x \in \Omega, \quad \theta(x) = \begin{cases} \frac{\nabla u_k(x)}{|\nabla u_k(x)|} & \text{if } \nabla u_k(x) \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta^{\perp}$ is its counterclockwise rotation of angle $\frac{\pi}{2}$ which represents the unit tangents of the level lines of u_k, and ∇u_k is the gradient of u_k which is a vector-valued measure with finite total variation.

The high-resolution multispectral image is computed by minimizing the following energy functional

$$\sum_{k=1}^{M} \gamma_k \int_{\Omega} |\theta^{\perp} \cdot \nabla u_k(x)|^2 dx + \lambda \int_{\Omega} \left(\sum_{k=1}^{M} \alpha_k u_k(x) - P(x) \right)^2$$
$$+ \mu \sum_{k=1}^{M} \int_{\Omega} \Pi_s \left(K_s * u_k(x) - u_k^{\Omega}(x) \right)^2,$$

subject to $0 \leqslant u_k \leqslant M_k$, where

$$M_k = \max_{(i,j) \in \Omega_S} \max \left(\frac{P(i,j)}{\alpha_k}, u_k^S(i,j) \right),$$

and $u_k \in BV(\Omega)$, $\gamma_k, \lambda, \mu \geqslant 0$ (in practice all these parameters are taken equal to

111

1).

The release of the $P + XS$ model inspired the scientific community and many interesting pansharpening algorithms came out. Duran *et al.* [143] suggested a variational technique for pansharpening that was inspired by the $P + XS$ problem and the nonlocal-means algorithm for image denoising [130] presented in (2.9). Duran *et al.* kept the idea of the $P + XS$ model but they introduced a nonlocal regularizer. The latter forces the fused image to share the same geometric regularities as the panchromatic one by harnessing its self-similarities. The panchromatic image was used for extracting the relationship between the patches that describe and define the geometry of the fused image. The optimization is performed on each band independently which is convenient in the case where misregistered and non-aligned data are manipulated. In this way, the propagation of errors due to ill-acquisition is minimized. Morevover, this technique contributed to reducing the color artifacts significantly.

The nonlocal regularization term that takes advantage of the self-similarity of the panchromatic image is presented as follows

$$\iint_{\Omega \times \Omega} (u_k(y) - u_k(x))^2 \omega_P(x,y) \, dy \, dx, \ \forall k \in \{1, \dots, M\},$$

where $\omega_P : \Omega \times \Omega \to \mathbb{R}$ is a weight function. The latter is computed with the panchromatic image and it is used to extract the relationships between the patches. The weight function ω_P describes the geometry of the desired reconstructed image and it is defined as follows

$$\omega_P(x,y) = \frac{1}{\Upsilon(x)} \exp\left(- \frac{d_\rho(P(x), P(y))}{h^2} \right), \tag{3.3a}$$

where

$$\Upsilon(x) = \int_\Omega \exp\left(- \frac{d_\rho(P(x), P(y))}{h^2} \right) dy, \ \forall x \in \Omega \tag{3.3b}$$

is a normalization factor and

$$d_\rho(P(x), P(y)) = \int_\Omega K_\rho(t)|P(x+t) - P(y+t)|^2 dt, \tag{3.3c}$$

is the distance between the patches centered around x and y. In this setting, K_ρ is a Gaussian kernel and $h > 0$ is a filtering parameter. The filtering parameter h controls the decay of the exponential function and thus the decay of the weights. The weights ω_p defined in (3.3) satisfy the condition $0 < \omega_p(x, y) \leqslant 1$ and $\int_\Omega \omega_p(x, y) dy = 1$, but because of the normalization defined in (3.3b), the symmetry of the weights does not hold anymore.

In this pansharpening model the authors assumed that the panchromatic image is a linear combination of the high-resolution multispectral bands. The pansharpened image is found by minimizing the following energy term

$$
\begin{aligned}
&\frac{1}{2} \sum_{k=1}^{M} \iint_{\widetilde{\Omega} \times \widetilde{\Omega}} (u_k(y) - u_k(x))^2 \omega_{P_k}(x, y) \, dy \, dx \\
&+ \frac{\lambda}{2} \int_\Omega \left(\sum_{k=1}^{M} \alpha_k u_k(x) - P(x) \right)^2 \\
&+ \frac{\mu}{2} \sum_{k=1}^{M} \int_\Omega \Pi_s \left(K_s * u_k(x) - u_k^\Omega(x) \right)^2,
\end{aligned}
\tag{3.4}
$$

where $\lambda \geqslant 0$ and $\mu \geqslant 0$ are trade-off parameters that control the contribution of the fidelity terms with respect to the nonlocal regularization term. The domain $\widetilde{\Omega}$ is a nonlocal domain defined as $\widetilde{\Omega} = \Omega \cup \Gamma$ where Γ is a nonlocal boundary such that $\Gamma \subset \mathbb{R}^d \setminus \Omega$.

Duran *et al.* [67] suggested another variational model for pansharpening with a new term. Let $P_k^s : \Omega_S \to \mathbb{R}$ be the downsampled version of the panchromatic image P_k. Let $\tilde{P}_k : \Omega \to \mathbb{R}$ and $\tilde{u}_k : \Omega \to \mathbb{R}$ be the extensions by bicubic interpolation of P_k^s and u_k^s to the whole domain Ω. The goal of the new term is

the preservation of the geometrical properties of the observed scene. First, the ratio between each low-resolution spectral component and the downsampled panchromatic image is computed. Then, this ratio is imposed to be similar to the ratio of the original panchromatic and each band of the fused product. Finally, the following constraint, also called radiometric constraint is imposed

$$\frac{u_k(x)}{P_k(x)} = \frac{\tilde{u}_k(x)}{\tilde{P}_k(x)}, \ \forall x \in \Omega, \ \forall k \in \{1,\dots,M\}. \tag{3.5}$$

The condition (3.5) assumes that each band P_k is geometrically aligned with the corresponding kth spectral band in the unknown image. The variational framework of (3.5) is the following integral expression

$$\int_\Omega \Big(u_k(x)\tilde{P}_k(x) - \tilde{u}_k(x)P_k(x)\Big)^2, \ \forall x \in \Omega, \ \forall k \in \{1,\dots,M\}. \tag{3.6}$$

The pansharpening model is presented as

$$\begin{aligned}
&\frac{1}{2}\sum_{k=1}^M \iint_{\Omega\times\Omega} (u_k(y) - u_k(x))^2 \omega_{P_k}(x,y)\,dy\,dx \\
&+ \frac{\mu s^2}{2}\sum_{k=1}^M \int_\Omega \Pi_s\Big(K_s * u_k(x) - u_k^\Omega(x)\Big)^2 dx \\
&+ \frac{\delta}{2\|P_k\|^2}\sum_{k=1}^M \int_\Omega \Big(u_k(x)\tilde{P}_k(x) - \tilde{u}_k(x)P_k(x)\Big)^2 dx,
\end{aligned} \tag{3.7}$$

where $\mu \geqslant 0$ and $\delta \geqslant 0$ are respectively normalized by the sampling factor s and the mean value of the panchromatic image, $\|P_k\| = \sqrt{\frac{1}{|\Omega|}\int_\Omega (P_k(x))^2 dx}$. These parameters define the contribution of each term to the whole energy.

We notice that, in the pansharpening model (3.7), the constraint (3.1) linked to the generation of the panchromatic image and used in (3.4), was replaced by the radiometric constraint (3.5). The latter, proved efficient in preserving spatial details, reducing color distortions and avoiding aliasing artifacts.

3.2 L^2-based nonlocal model

In this chapter, a new way of carrying out the HS-MS fusion is suggested based on a new approach that is completely different from the OT model studied in Chapter 2. From the generation model (1.7), we can see that recovering an estimation of **u** is equivalent to conducting an inverse operation. Depending on the operators D, B and S that are, by the way, not invertible, one can see that the solution cannot be unique. Therefore, the fusion problem is ill-posed because it does not satisfy the conditions of Hadamard for well-posedness which, we recall, are:

- The existence of the solution,

- The uniqueness of the solution,

- The solution has to have a continuous change of behavior with the initial conditions.

The new fusion problem, based on the idea of (3.4) presented in [67], is seen as the minimization of an energy functional composed of four terms. Each one of these terms has a crucial role in the fusion result because it describes and imposes constraints that **u** should verify. The new suggested energy functional is described as follows

$$
\min_{u \in \mathbb{R}^{H \times N}} \sum_{h=1}^{H} \|\nabla_{\omega_h} u_h\|_1 + \frac{\mu}{2} \sum_{h=1}^{H} \|DBu_h - g_h\|_2^2 +
$$
$$
\frac{\gamma}{2} \sum_{m=1}^{M} \|(Su)_m - f_m\|_2^2 + \frac{\lambda}{2} \sum_{h=1}^{H} \|\tilde{P}_h u_h - P_h \tilde{g}_h\|_2^2. \tag{3.8}
$$

As discussed earlier, the HS-MS fusion is an ill-posed inverse problem. In this case, regularizing the problem is mandatory in order to make it well-posed and recover a unique minimum which is the desired fused image **u**. Thus,

choosing a regularizer that accounts for this ill-posedness is crucial and it has to be chosen carefully.

In the energy functional (3.8), the first term is a regularizer which is the nonlocal gradient $\nabla_{\omega_h} u_h \in \mathbb{R}^{N \times N}$ penalized with the combination of the L^1-norm and the L^2-norm. The gradient is computed for each pixel such that $\nabla_{\omega_h} u_{h,i} = \left((\nabla_{\omega_h} u_{h,i})_1, \ldots, (\nabla_{\omega_h} u_{h,i})_N \right)$ and it is defined as follows

$$(\nabla_{\omega_h} u_{h,i})_j = \sqrt{\omega_{h,i,j}} \left(u_{h,j} - u_{h,i} \right), \tag{3.9}$$

where $\omega_{h,i,j}$ is a similarity measure that will be defined in details later in the document. Unlike the classical gradient that interacts locally with the pixels, the nonlocal gradient allows the computation of the gradient between any two pixels in the image times $\omega_{h,i,j}$. As to the penalization of the gradient at each pixel, it is computed as follows $\|\nabla_{\omega_h} u_h\|_1 = \sum_i |\nabla_{\omega_h} u_{h,i}|$ where in this case $|\cdot|$ denotes the Euclidean norm.

The terms $\|DBu_h - g_h\|_2^2$ and $\|(Su)_m - f_m\|_2^2$ are related to the image generation model (1.7) and they are called data fitting terms. The goal of the data fitting terms is the penalization of the deviation from the multispectral and hyperspectral data f_m and g_h respectively. The penalization is done throughout the L^2-norm.

The last term $\|\tilde{P}_h u_h - P_h \tilde{g}_h\|_2^2$ is a radiometric constraint penalized with the L^2-norm and it aimes at forcing the fused image **u** to share the same high spatial frequencies as the multispectral image **f** where

- $P_h \in \mathbb{R}^{H \times N}$ is a linear combination of the multispectral image such that

$$P_h = \sum_{m=1}^{M} \alpha_{m,h} f_m, \ \forall h \in \{1, \ldots, H\}. \tag{3.10}$$

- $\tilde{P}_h \in \mathbb{R}^{H \times N}$ is computed as

$$\tilde{P}_h = \sum_{m=1}^{M} \alpha_{m,h} \tilde{f}_m, \ \forall h \in \{1, \ldots, H\}. \tag{3.11}$$

- $\tilde{g} \in \mathbb{R}^{H \times N}$ is the hyperspectral image brought to the fine grid of the fused image by bicubic interpolation.

The coefficients $\alpha_{m,h}$ are defined as follows

$$\alpha_{m,h} = \frac{s_{m,h}}{\sum_m s_{m,h}}, \tag{3.12}$$

where $s_{m,h}$ are the coefficients extracted from the spectral response defined in (1.7).

As one can notice, $\tilde{P}_h$ is computed as in (3.10) but with $\tilde{f}_m$. The latter is the result of the spatial degradation as in the generation model (1.7) and then, it is brought back to the grid of the multispectral image with bicubic interpolation. This manipulation kills the high spatial frequencies and leaves only the low spatial ones. The radiometric constraint is detailed further in the document.

3.2.1 Similarity weights

The similarity weights are used in the computations of the nonlocal gradient as shown in expression (3.9). They are computed as a mixture of spatial distances, computed with the positions of the pixels, and nonlocal weights that compare the similarity between the patches centered on the pixels. The similarity weight

corresponding to a pixel with the coordinates (i,j,h) is defined as follows

$$
\omega_{h,i,j} = \begin{cases} \dfrac{1}{\Gamma_i} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{spt}^2} - \dfrac{\sum_{m=1}^{M} s_{m,h} \sum_{\{t\in\mathbb{Z}^2:\|t\|_\infty \leqslant \nu_c\}} \|f_{m,i+t} - f_{m,j+t}\|_2^2}{h_{sim}^2 (2\nu_c+1)^2 \sum_{m=1}^{M} s_{m,h}}\right) & \text{if} \\[2em] & \|i-j\|_\infty \leqslant \nu_r, \\[1em] & (i \neq j) \\[1em] 0 & \text{else,} \end{cases}
$$

$$(3.13a)$$

$$
\Gamma_i = \sum_{\{j:\|i-j\|_\infty \leqslant \nu_r\}} \exp\left(-\frac{\|i-j\|_2^2}{h_{spt}^2} - \frac{\sum_{m=1}^{M} s_{m,h} \sum_{\{t\in\mathbb{Z}^2:\|t\|_\infty \leqslant \nu_c\}} \|f_{m,i+t} - f_{m,j+t}\|_2^2}{h_{sim}^2 (2\nu_c+1)^2 \sum_{m=1}^{M} s_{m,h}}\right) \tag{3.13b}
$$

$$
\omega_{h,i,i} = \max\left\{\omega_{h,i,j} : \|i-j\|_\infty \leqslant \nu_r \text{ and } j \neq i\right\} \tag{3.13c}
$$

where Γ_i is a normalization factor.

The search for similar pixels in the whole image takes a high computational time and requires a big storage capacity. Moreover, it does not make sense to search for similar patches in the whole image because, when the search oversteps the surrounding of the central pixel, it is hard to find similar patches. Thus, the computations prove useless and time and memory consuming. For this matter, a size $\nu_r \in \mathbb{Z}^+$ for the search window was imposed where the search for similar pixels takes place. The parameter $\nu_c \in \mathbb{Z}^+$ is the size of the patch centered on the pixel (i,j,h). The parameters $h_{spt} > 0$ and $h_{sim} > 0$ act like filtering ones for the spatial distance and for the measure of similarity between the patches, respectively. These parameters control the spread of the weights, moreover, they quantify the speed of decrease of the weights when the dissimilarity between the patches increases.

The weight of the central pixel with respect to itself is quite important and in order to avoid giving the central pixel an excessive weight, $\omega_{h,i,j}$ is set to the

maximum weight of the other pixels with respect to the central one as illustrated in expression (3.13c). To summarize, in order to find the pixels that are similar to the pixel (i,j,h) (the central one) in the sense of definition (3.13), a search window of size v_r and a patch of size v_c are centered on (i,j,h). Next, another patch of size v_c is centered on each pixel (i',j',h') in the search window and then the weights are computed according to (3.13) between the central pixel and all the pixels in the search window.

As one can notice, the computation of the weights in (3.13) were chosen to be computed on the multispectral image $\mathbf{f} \in \mathbb{R}^{M \times N}$ because of the high spatial resolution that this latter contains. However, the weights should be computed for each pixel of the fused image $\mathbf{u} \in \mathbb{R}^{H \times N}$, except that $\mathbf{f}$ contains only M bands which are much fewer than H. Thus, the spectral response S was chosen in order to interpolate the weights and extend them to the H number of bands. Other ways of interpolating the weights might be found. But, the choice of the spectral response S to carry out this mission was considered legitimate because, this same response was used to generate the multispectral image from the high resolution one and thus S is a link between both images.

The spectral response S can be represented under the form of a matrix with M rows and H columns as follows

$$
S = \begin{pmatrix}
s_{1,1} & s_{1,2} & s_{1,3} & \cdots & s_{1,H} \\
s_{2,1} & s_{2,2} & s_{2,3} & \cdots & s_{2,H} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
s_{M,1} & s_{M,2} & s_{M,3} & \cdots & s_{M,H}
\end{pmatrix}.
$$

For each band $h \in \{1,\dots,H\}$, there are M corresponding coefficients $s_{m,h}$ in the multispectral grid that are used in the computations of the similarity weights (3.13).

For example, in the case where $M = 4$, $H = 93$ and $x = (1, 1, 20)$ is the central pixel, a search window is centered on the latter. Next, for the computation of similar patches, the four bands of the multispectral image corresponding to the spatial position $(1, 1)$ are used. The final similar patches are found by averaging the four values on the four bands of the multispectral image. The averaging is done with the coefficients (3.12) of the 20th column of the S matrix computed as illustrated in Figure 3.1. The final weights are then computed by adding the distances between the pixels as shown in definition (3.13).
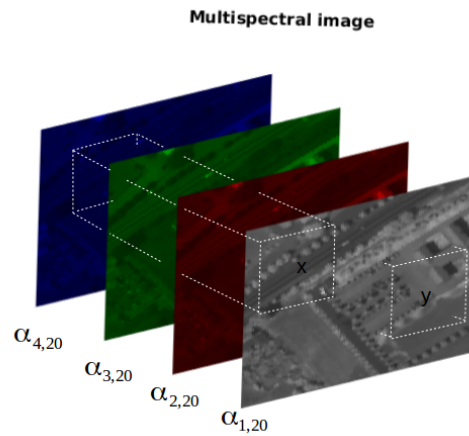
Figure 3.1: Computations of the similarity weights of the pixel $x = (1, 1, 20)$ with the multispectral image

The goal behind choosing the multispectral image **f** to compute the similarity weights for the regularization term, is to copy the geometric information contained in **f** into the final fused image **u**. In other terms, **u** is forced to share the same geometric configuration as the multispectral image, because this latter contains the finest spatial details.

3.2.2 Radiometric constraint

The radiometric constraint used in [67] was adapted in our case in order to inject the high spatial frequencies of the multispectral image into the desired fused image. The multispectral image is used because it contains the relevant spatial information and the fine geometric details of the scene captured by the multispectral sensor. The injection of spatial details is done throughout the following constraint

$$\frac{u_h}{P_h} = \frac{\tilde{g}_h}{\tilde{P}_h} \tag{3.14}$$

and the variational formulation of the latter corresponds to the fourth constraint and energy term in the energy functional (3.8). Note that the constraint (3.14) is equivalent to the following formula [67]:

$$u_h - \tilde{g}_h = \frac{\tilde{g}_h}{\tilde{P}_h}\left(P_h - \tilde{P}_h\right), \forall h \in \{1,\dots,H\} \tag{3.15}$$

The data P_h contains the high and the low spatial frequencies, as to $\tilde{P}_h$, it contains only the low frequencies. Thus, by carrying out the difference $P_h - \tilde{P}_h$, only the high spatial frequencies derived from the multispectral data are kept. After computing the difference $u_h - \tilde{g}_h$, only the unknown high spatial frequencies of the fused image are kept because $\tilde{\mathbf{g}}$ contains only the low spatial frequencies. The modulation coefficient $\frac{\tilde{g}_h}{\tilde{P}_h}$ can be different for each band h and it takes into account the energy level of the multispectral and the hyperspectral image. Thus, by adding the radiometric constraint, the high spatial frequencies of the multispectral image are forced to coincide with the high frequencies of the fused image $\mathbf{u}$.

As one can notice, in the constraints (3.10) and (3.11) the coefficients $s_{m,h}$ are used. The reason behind the use of these coefficients is the same one as for the computation of the similarity weights, as explained in Section 3.2.1, which

is the extension of data from M bands to H bands. Thus, $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are spectral interpolations of $\mathbf{f}$ and $\tilde{\mathbf{f}}$ throughout weighted linear combinations with the coefficients $s_{m,h}$.

3.3 Chambolle-Pock algorithm

The resolution of the minimization problem (3.8) is done with the first-order primal-dual algorithm of Chambolle and Pock [144] which can be applied to many convex optimization problems. The algorithm is presented as follows.

Let X and Y be two finite-dimensional real vector spaces endowed with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \| = \langle \cdot, \cdot \rangle^{\frac{1}{2}}$. Let $K : X \to Y$ be a general continuous linear operator. Let $F : Y \to [0, +\infty]$ and $G : X \to [0, +\infty]$ be a proper, convex, and lower-semicontinous functionals. Consider the nonlinear primal problem

$$\min_{u \in X} F(Ku) + G(u),$$

and its corresponding saddle-point problem

$$\min_{u \in X} \max_{y \in Y} \langle Ku, y \rangle + G(x) - F^*(y), \tag{3.16}$$

where F^* being itself the convex conjugate of F.

Problem (3.16) has at least one solution that satisfies

$$K\hat{x} \in \partial F^*(\hat{y}),$$

$$-(K^*\hat{y}) \in \partial G(\hat{x}),$$

where ∂F^* and ∂G are the subgradients of the convex functions F^* and G, respectively, and K^* is the adjoint of the linear operator K.

The use of Chambolle and Pock's algorithm requires the use of the proximity operator which is a generalization of the projection operator. The proximal operator of a closed proper convex function v is defined as follows

$$\text{prox}_{\tau v}(x) = \underset{y}{\text{argmin}}\{v(y) + \frac{1}{2\tau}\|x - y\|^2\}, \tag{3.17}$$

where $\tau > 0$ is a scaling parameter that controls the extent of the movement with which the proximal operator converges to the minimum of v. Since problem (3.17) is strongly convex, the solution is unique. The definition (3.17) indicates that $\text{prox}_{\tau v}(x)$ is a point that compromises between minimizing v and being near x. In the case where v is the indicator function I_C such that

$$\forall x, \quad I_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise}, \end{cases} \tag{3.18}$$

where C is a closed convex set, the proximal operator defined in (3.17) becomes

$$\text{Proj}_C(x) = \underset{y \in C}{\text{argmin}} \|x - y\|^2,$$

which corresponds to the Euclidean projection on C and thus, the proximity operator can be seen as a generalized projections. For differentiable functions and small values of τ, the proximal mapping behaves as a kind of gradient step, that is, $\text{prox}_{\tau v}(x) \simeq x - \tau \nabla_v(x)$. Moreover, the proximal operator of a function can be related to its Legendre-Fenchel transform (or convex conjugate) throughout *Moreau's identity* as follows

$$x = \text{prox}_v(x) + \tau \text{prox}_{\frac{1}{\tau} v^*}(\frac{x}{\tau}).$$

The algorithm steps to solve the saddle-point problem (3.16) proposed by

Chambolle and Pock are the following

$$\begin{cases} y^{(i+1)} = \text{prox}_{\sigma F^*}(y^n + \sigma K\bar{x}^{(i)}), \\ x^{(i+1)} = \text{prox}_{\tau G}(x^{(i)} - \tau K^* y^{(i+1)}), \\ \bar{x}^{(i+1)} = x^{(i+1)} + \theta(x^{(i+1)} - x^{(i)}). \end{cases} \tag{3.19}$$

The steps of the algorithm (3.19) consist in alternating a gradient descent in the dual variable y and a gradient descent in the primal variable x. The step-size parameters $\tau > 0$ and $\sigma > 0$ are chosen such that $\tau\sigma\|K\|^2 < 1$. The parameter $\theta \in [0,1]$ controls the over-relaxation in x. The proximity operators have a closed-form representation and can be solved with a high-precision.

To apply the latter, the problem (3.8) has to be written under a saddle point problem (a min / max form). To do that, the first three terms of the functional to be minimized are dualized with the use of known results drawn from the convex analysis theory [145] using

$$\frac{\alpha}{2}\|x\|_2^2 = \max_y \langle x, y \rangle - \frac{1}{2\alpha}\|y\|_2^2,$$

and

$$\|x\|_1 = \max_y \langle x, y \rangle - \delta_P(y),$$

where δ_P is the indicator function defined as

$$\delta_P(y) = \begin{cases} 0 & \text{if } y \in P, \\ +\infty & \text{otherwise,} \end{cases}$$

with P the set defined by

$$P = \{y \in Y : \|y\|_\infty \leqslant 1\}.$$

The dualization of the different terms of the functional is as follows:

- $\sum_{h=1}^{H} \|\nabla_{\omega_h} u_h\|_1 = \max_{(p_1,\ldots,p_H)} \sum_{h=1}^{H} \left(\langle \nabla_{\omega_h} u_h, p_h \rangle - \delta_{\mathcal{P}_h}(p_h) \right)$

- $\frac{\mu}{2} \sum_{h=1}^{H} \|DBu_h - g_h\|_2^2 = \max_{(q_1,\ldots,q_H)} \sum_{h=1}^{H} \left(\langle DBu_h - g_h, q_h \rangle - \frac{1}{2\mu} \|q_h\|_2^2 \right)$

- $\frac{\gamma}{2} \sum_{m=1}^{M} \|(Su)_m - f_m\|_2^2 = \max_{(r_1,\ldots,r_M)} \sum_{m=1}^{M} \left(\langle (Su)_m - f_m, r_m \rangle - \frac{1}{2\gamma} \|r_m\|_2^2 \right)$

where $\mathbf{u} \in \mathbb{R}^{H \times N}$ is the primal variable and the unknown high-resolution image. As to $\mathbf{p} \in \mathbb{R}^{H \times N \times (2v_r+1)^2}$, it is the dual variable related to the nonlocal regularization term. The two terms $\mathbf{q} \in \mathbb{R}^{H \times N_l}$ and $\mathbf{r} \in \mathbb{R}^{M \times N}$ are the two dualized data-fidelity terms, related to the hyperspectral and the multispectral data respectively, arising from the observation model. Furthermore, $\delta_{\mathcal{P}_h}$ is the indicator function of

$$\mathcal{P}_h = \{p_h \in \mathbb{R}^{N \times (2v_r+1)^2} : \|p_h\|_\infty \leqslant 1\},$$

where $\|p_h\|_\infty = \max_i |p_h(i)|$ with $|p_h(i)| = \sqrt{\sum_{j=1}^{(2v_r+1)^2} p_h(i,j)^2}$ and $i \in [\![1, N]\!]$ denotes the number of the pixel in the hth band. Note that the set $\mathcal{P}_h$ is the cartesian product of L^2 balls.

Thus, the corresponding saddle-point writing of problem (3.8) can be written in the following form

$$\begin{aligned}
\min_{\mathbf{u}} \max_{\mathbf{p,q,r}} \sum_{h=1}^{H} &\left(\langle \nabla_{\omega_h} u_h, p_h \rangle - \delta_{\mathcal{P}_h}(p_h) \right) + \sum_{h=1}^{H} \left(\langle DBu_h - g_h, q_h \rangle - \frac{1}{2\mu} \|q_h\|_2^2 \right) \\
&+ \sum_{m=1}^{M} \left(\langle (Su)_m - f_m, r_m \rangle - \frac{1}{2\gamma} \|r_m\|_2^2 \right) + \frac{\lambda}{2} \sum_{h=1}^{H} \|\tilde{P}_h u_h - P_h \tilde{g}_h\|_2^2.
\end{aligned} \tag{3.20}$$

Following the general formulation of the saddle-point problem [144], we can do the following identifications:

- $K = \begin{pmatrix} \nabla_\omega \\ DB \\ S \end{pmatrix}$,

- $F^*(\mathbf{p}, \mathbf{q}, \mathbf{r}) = \delta_{\mathcal{P}}(\mathbf{p}) + \langle \mathbf{g}, \mathbf{q} \rangle + \langle \mathbf{f}, \mathbf{r} \rangle + \frac{1}{2\mu} \|\mathbf{q}\|_2^2 + \frac{1}{2\gamma} \|\mathbf{r}\|_2^2$ where $\delta_{\mathcal{P}}(\mathbf{p}) = \sum_{h=1}^{H} \delta_{\mathcal{P}_h}(p_h)$,

- $G(\mathbf{u}) = \frac{\lambda}{2} \|\tilde{\mathbf{P}}\mathbf{u} - \mathbf{P}\tilde{\mathbf{g}}\|_2^2$,

where G and F^* are proper convex lower-semicontinous (l.s.c.) functions with F^* itself the convex conjugate of a l.s.c. function F.

The steps of the primal-dual algorithm [144] are the following

- $(\mathbf{p}^{n+1}, \mathbf{q}^{n+1}, \mathbf{r}^{n+1}) = \text{prox}_{\sigma F^*}(\mathbf{p}^n + \sigma \nabla_\omega \bar{\mathbf{u}}^n, \mathbf{q}^n + \sigma DB\bar{\mathbf{u}}^n, \mathbf{g}^n + \sigma S\bar{\mathbf{u}}^n)$,

- $\mathbf{u}^{n+1} = \text{prox}_{\tau G}(\mathbf{u}^n + \tau \nabla_\omega^* \mathbf{p}^{n+1} - \tau (DB)^* \mathbf{q}^{n+1} - \tau S^* \mathbf{r}^{n+1})$,

- $\bar{\mathbf{u}}^{n+1} = 2\mathbf{u}^{n+1} - \mathbf{u}^n$,

where ∇_ω^*, $(DB)^*$ and S^* are the adjoint operators of ∇_ω, (DB) and S, respectively, with

$$\nabla_\omega^* = -\text{div}, \quad (DB)^* = B^\mathsf{T} D^\mathsf{T}, \quad S^* = S^\mathsf{T}$$

3.4 Computation of the proximal operators

In this part, the update of the dual variables is done through the computation of the proximal operator of σF^* as

$$(\mathbf{p}, \mathbf{q}, \mathbf{r}) = \text{prox}_{\sigma F^*}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{r}}),$$

where $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{r}})$ are the dual variables to be updated and $(\mathbf{p}, \mathbf{q}, \mathbf{r})$ are the updated ones. The details of the computations are given below.

126

3.4.1 Proximal operator of σF^*

- Update of the regularization term-related dual variable $\mathbf{p}$ is carried out as follows

$$\mathbf{p} = \underset{\mathbf{p}}{\arg\min} \frac{1}{2\sigma} \|\mathbf{p} - \tilde{\mathbf{p}}\|_2^2 + \delta_{\mathcal{P}}(\mathbf{p}) \Leftrightarrow \underset{\|p_h\|_\infty \leqslant 1}{\arg\min} \frac{1}{2\sigma} \|p_h - \tilde{p}_h\|_2^2. \qquad (3.21)$$

Equation (3.21) is the Euclidean projection of the convex subset $\mathcal{P}_h$, which we recall is the product of L^2 balls. We have then

$$\underset{\|p_h\|_\infty \leqslant 1}{\arg\min} \frac{1}{2\sigma} \|p_h - \tilde{p}_h\|_2^2 \Leftrightarrow p_h = \begin{cases} \tilde{p}_h & \text{if} \quad \|\tilde{p}_h\|_\infty \leqslant 1, \\ \frac{\tilde{p}_h}{|\tilde{p}_h|} & \text{if} \quad \|\tilde{p}_h\|_\infty > 1 \end{cases}$$

$$\Leftrightarrow p_h(x) = \frac{\tilde{p}_h(x)}{\max(1, |\tilde{p}_h(x)|)}.$$

.

- Update of the dual variable $\mathbf{q}$ related to the hyperspectral image $\mathbf{g}$ as

$$\mathbf{q} = \underset{\mathbf{q}}{\arg\min} \frac{1}{2\sigma} \|\mathbf{q} - \tilde{\mathbf{q}}\|_2^2 + \langle \mathbf{g}, \mathbf{q} \rangle + \frac{1}{\mu} \|\mathbf{q}\|_2^2 \Leftrightarrow q_h = \frac{\tilde{q}_h - \sigma g_h}{1 + \frac{\sigma}{\mu}}.$$

- Update of the dual variable $\mathbf{r}$ related to the multispectral image $\mathbf{f}$ as

$$\mathbf{r} = \underset{\mathbf{r}}{\arg\min} \frac{1}{2\sigma} \|\mathbf{r} - \tilde{\mathbf{r}}\|_2^2 + \langle \mathbf{f}, \mathbf{r} \rangle + \frac{1}{\gamma} \|\mathbf{r}\|_2^2 \Leftrightarrow r_m = \frac{\tilde{f}_m - \sigma f_m}{1 + \frac{\sigma}{\gamma}}.$$

3.4.2 Proximal operator of τG

The update of the primal variable $\mathbf{u}$ is done through the computation of the proximal operator of τG as follows

$$\mathbf{u} = \text{prox}_{\tau G}(\tilde{\mathbf{u}}) \Leftrightarrow \mathbf{u} = \underset{\mathbf{u}}{\text{argmin}} \frac{1}{2\tau} \|\mathbf{u} - \tilde{\mathbf{u}}\|_2^2 + \frac{\lambda}{2} \|\tilde{\mathbf{P}}\mathbf{u} - \mathbf{Pg}\|_2^2$$

$$\Leftrightarrow u_h = \frac{\tilde{u} + \tau\lambda\tilde{P}_h P_h g_h}{1 + \tau\lambda\tilde{P}_h^2}.$$

3.5 The optimization algorithm

The resolution of the fusion problem proposed in (3.8) is solved with the first-order primal-dual algorithm of Chambolle and Pock defined in (3.3).

Algorithm 3: The primal-dual algorithm for HS-MS fusion

Input : $\mathbf{f}$, $\mathbf{g}$, D, B, S, $\mathbf{P}$, $\tilde{\mathbf{P}}$, ∇_ω, τ and σ

1 **for** $i \leftarrow 0$ **to** *convergence* **do**

2   *//Update of the primal variable $\mathbf{u}$ and the dual variables $\mathbf{q}$, $\mathbf{q}$ and $\mathbf{r}$*

3    $p_h^{i+1} = \frac{p_h^i(x) + \sigma\nabla_\omega\bar{u}_h^i}{\max(1, |p_h^i(x) + \sigma\nabla_\omega\bar{u}_h^i|)}$;

4    $q_h^{i+1} = \frac{q_h^i + \sigma(DB\bar{u}_h^i - g_h)}{1 + \frac{\sigma}{\mu}}$;

5    $r_m^{i+1} = \frac{g_l^i + \sigma((S\bar{u}^i)_m - f_m)}{1 + \frac{\sigma}{\gamma}}$;

6    $u_h^{i+1} = \frac{u_h^i + \tau\text{div}_\omega p_h^{i+1} - \tau(B^\intercal D^\intercal q^{i+1})_h - \tau(S^\intercal r^{i+1})_h + \tau\lambda\tilde{P}_h P_h \tilde{g}_h}{1 + \tau\lambda\tilde{P}_h^2}$

7    $\bar{u}_h^{i+1} = 2u_h^{i+1} - u_h^i$

8 **end**

Output: $\mathbf{u}$ (the fused image)

As one can notice, in the resolution algorithm (3), there is one primal variable and three dual variables. For each one of these variables, the update is done for each channel $h \in \{1, \ldots, H\}$. That means that the fusion resolution is carried out for each spectral channel independently.

3.6 Experimental results

In this section, the performance of the fusion algorithm that solves problem (3.8) is studied. Two datasets were used for the experiments, Urban dataset described in Section 2.8 and Chikusei dataset described in Section 2.10.1.

Figure 3.2 shows the visual results of the fusion algorithm on Chikusei dataset. All fused images except the one of the proposed method are affected by noise and aliasing artifacts. The proposed method is able to fuse spatial and spectral information while removing noise at the same time, thus leading to the best visual appearance of the final product. Figures 3.3 and 3.4 show the RMSE and SAM error mappings of the fusion result of the suggested method and the results of the state-of-the-art methods. On both figures, we can see that, visually, the distribution of error in the suggested method is less strong than in the other images. On one hand, the RMSE mapping of our method shows that the areas where there is texture and where patterns are repetitive are the ones where error is the weakest. This result is normal given that the image contains many regular structures which is a suited environment for the nonlocal weights. This can be explained by the fact that, each pixel can have many similar neighbors which is convenient to reconstruct the geometry of the image and reduce noise. We can also notice that, the error is important in isolated spots that do not have similar patterns in the rest of the image, which increases the probability of error. On the other hand, the SAM mapping shows that the error is more located in areas with contours such as roads. This could be explained by the fact that our model is a variational one with a regularization term, the latter applies some kind of diffusion which might penalize edges and sharp changes in the image.

Table 3.1 shows the quantitative results of fusion of the proposed method and other state-of-the-art methods. The results are shown for Chikusei dataset with a SNR of 35dB. From Table 3.1 we can conclude that the proposed method

outperforms all state-of-the-art methods in all the quality indices.

Figure 3.2: Performances of HS-MS fusion with the nonlocal gadient based method and comparison with other state of the art methods on Chikusei dataset $304 \times 304 \times 93$. From top left to bottom. Reference image, HS image, MS image, CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB and the result of the proposed method.

Figure 3.5 shows the results on Urban dataset. In this case, a SNR of 30dB was fixed for the generation of the multispectral and the hyperspectral data,

Figure 3.3: Mapping of the RMSE error computed for each pixel for Chikusei dataset $304 \times 304 \times 93$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB and the result of the proposed method.

Figure 3.4: Mapping of the SAM error computed for each pixel for Chikusei dataset $304 \times 304 \times 93$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB and the result of the proposed method.

	RMSE	PSNR	ERGAS	SAM	$Q2^n$	CC	DD
Reference	0	∞	0	0	1	1	
CNMF	6.1187	37.8164	2.6182	1.5766	0.9614	0.9879	4.0512
HySure	6.1941	37.7096	3.0421	2.0346	0.9425	0.9825	4.2055
GSA	5.0398	39.5011	1.9896	1.8289	0.9474	0.9893	3.3648
SFIMHS	13.7861	30.7601	4.8313	2.8329	0.8858	0.9444	8.4121
GLPHS	13.0311	31.2502	4.3882	2.3949	0.8931	0.9521	8.2641
MAPSMM	16.6531	29.1189	5.7478	2.7116	0.8459	0.9197	9.9148
HMWB	<u>4.1998</u>	<u>41.0851</u>	**1.8715**	<u>1.5297</u>	<u>0.9531</u>	<u>0.9923</u>	<u>2.9915</u>
Proposed	**3.7719**	**42.0191**	<u>1.9286</u>	**1.3101**	**0.9698**	**0.9931**	**2.2352**

Table 3.1: Quantitative quality evaluation of each fused product on Chikusei dataset with the size $304 \times 304 \times 93$ and SNR=35dB.

which means that the level of noise is higher than for Chikusei data set but still moderate. On Figures 3.6 and 3.7 we have the RMSE and SAM error mappings. We notice that visually, even with the presence of more noise, the suggested method performs better than the rest of the methods where some suffer from important errors. Similarly to Chikusei dataset, the RMSE mapping shows that the errors are rather located in areas that lack regular and repetitive patterns whereas the SAM mapping emphasizes the lack of details in contours.

Based on the analysis above, we notice that our method is not affected by noise like some of the state-of-the-art methods. Thus, the suggested model is more robust to noise that corrupts the hyperspectral and the multispectral dataset.

Table 3.2 displays quantitative results corresponding to the Urban dataset. The proposed model is less affected by noise unlike most state-of-the-art methods.

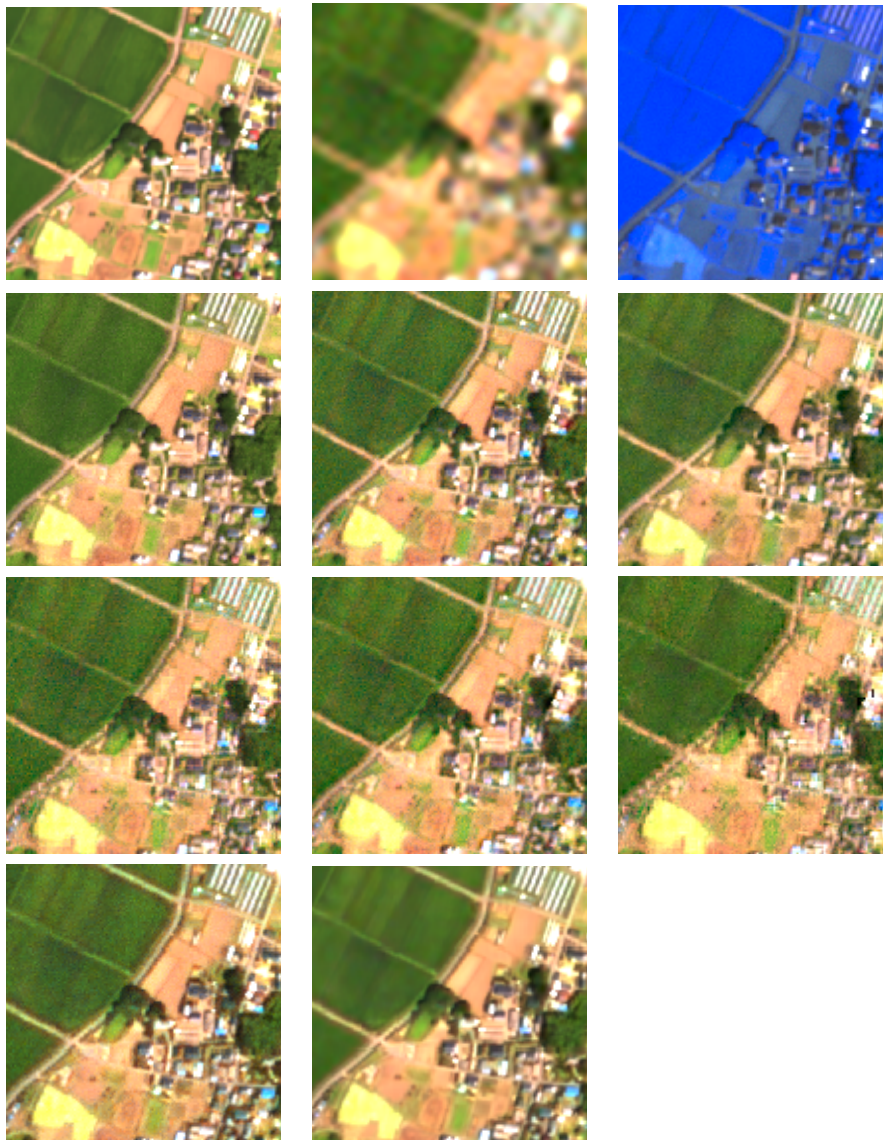Figure 3.5: Performances of HS-MS fusion with the nonlocal-gadient-based method and comparison with other state of the art methods on Urban dataset $128 \times 128 \times 93$. From top left to bottom. Reference image, HS image, MS image, CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB and the result of the proposed method.
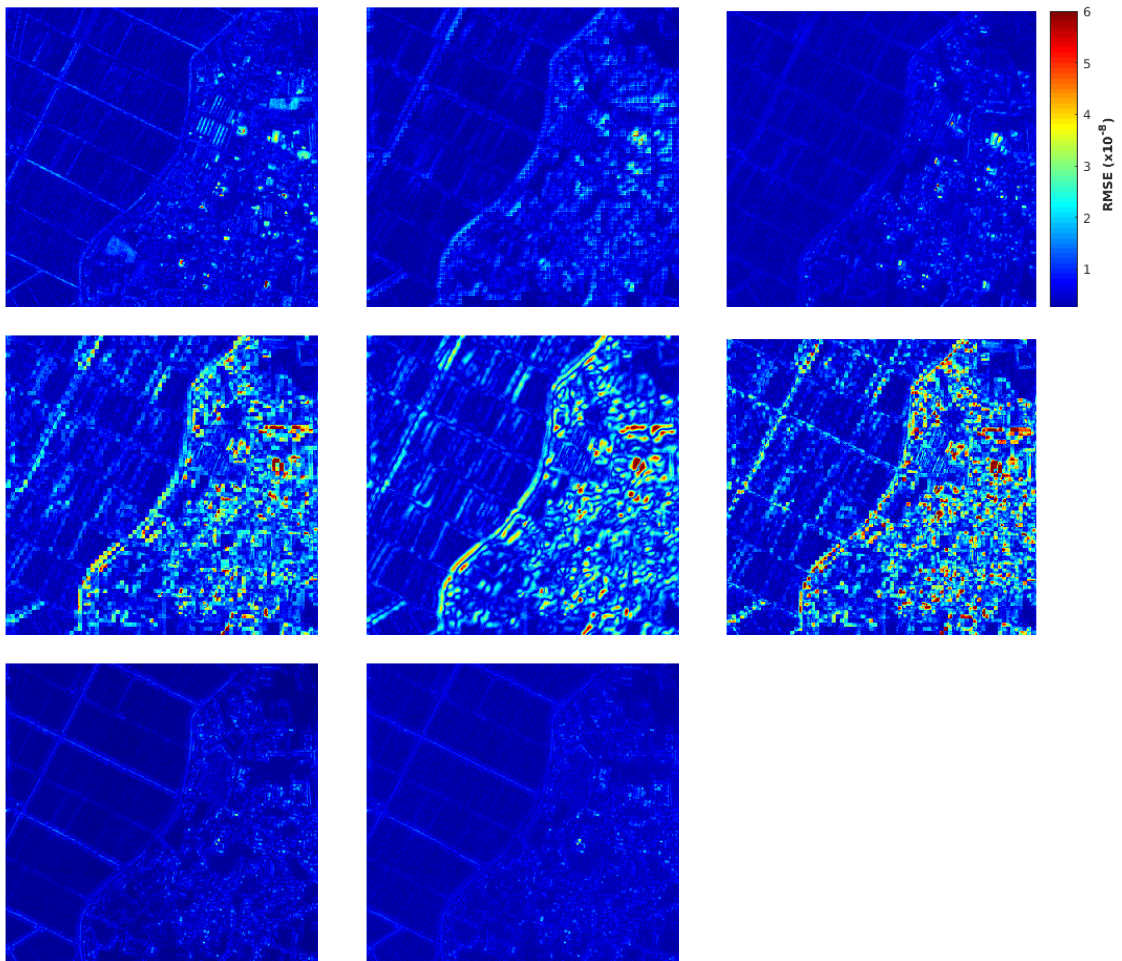
Figure 3.6: Mapping of the RMSE error computed for each pixel for Chikusei dataset $128 \times 128 \times 93$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB and the result of the proposed method.
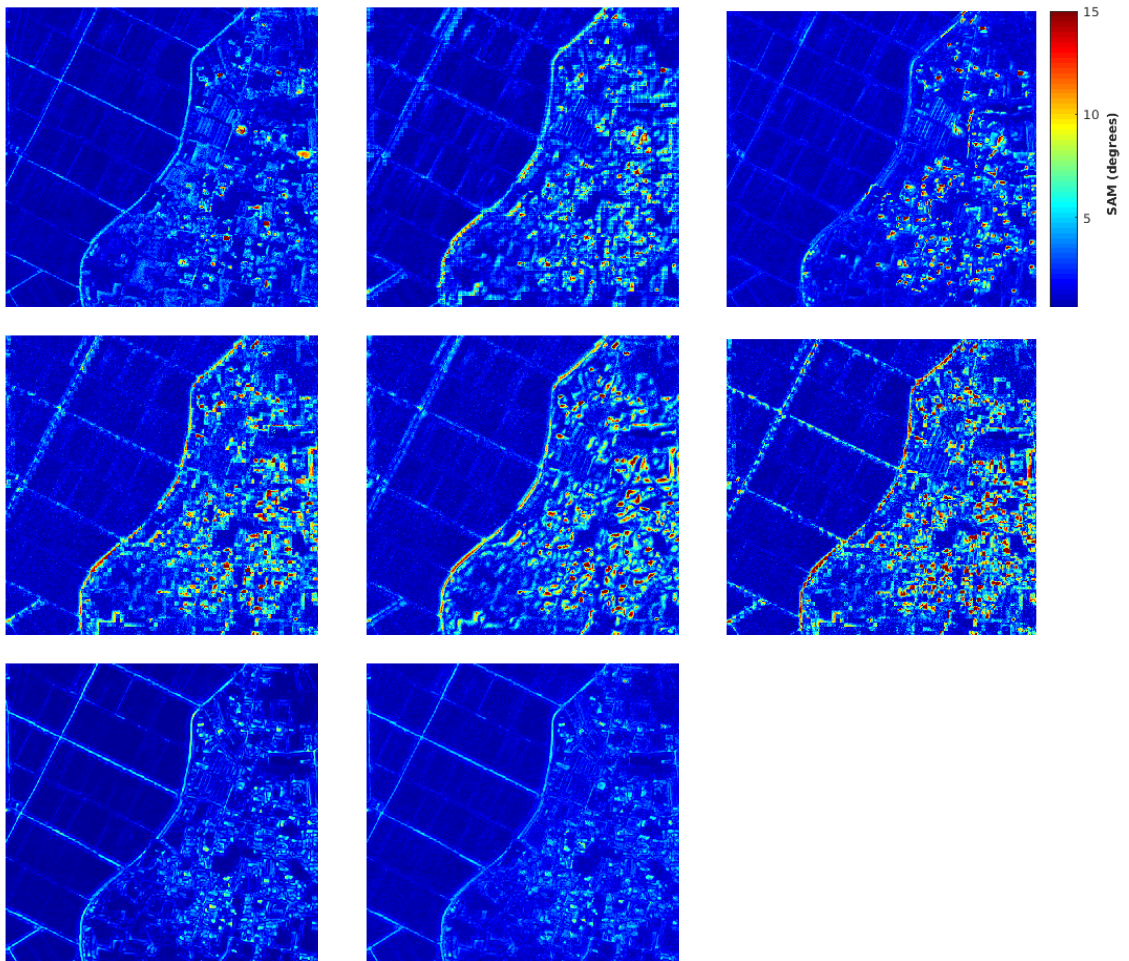
	RMSE	PSNR	ERGAS	SAM	$Q2^n$	CC	DD
Reference	0	∞	0	0	1	1	0
CNMF	6.0361	28.0728	2.0809	2.1354	0.9291	0.9433	4.0331
HySure	4.7931	30.0751	1.7971	2.9656	0.9486	0.9601	3.5448
GSA	5.2798	29.2354	1.8115	3.0526	0.9473	0.9585	3.7186
SFIMHS	6.9903	26.7973	2.5167	2.4926	0.8949	0.9141	5.0309
GLPHS	6.7556	27.0941	2.4272	2.4663	0.9005	0.9223	4.9163
MAPSMM	8.9021	24.6974	3.1723	3.2062	0.8156	0.8609	6.3091
HMWB	_4.0242_	_31.5941_	_1.4591_	_2.0955_	_0.9638_	_0.9733_	_2.9603_
Proposed	**3.6611**	**32.4153**	**1.3115**	**2.0271**	**0.9708**	**0.9798**	**2.5921**

Table 3.2: Quantitative quality evaluation of each fused product on Urban dataset with the size $128 \times 128 \times 93$ and SNR=30dB.
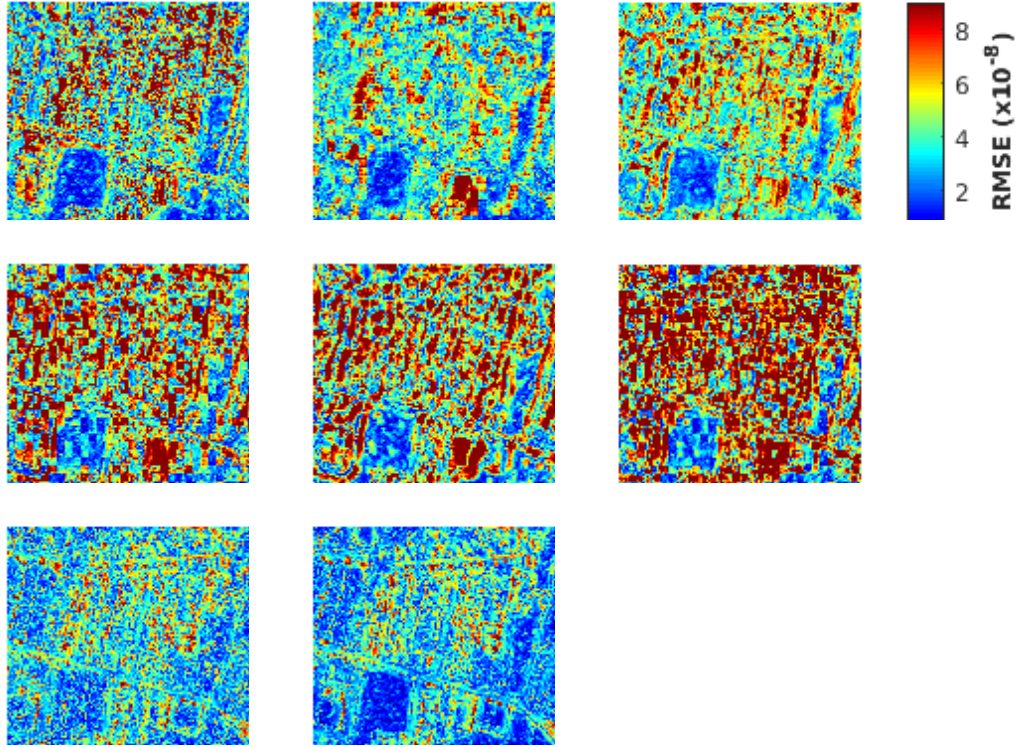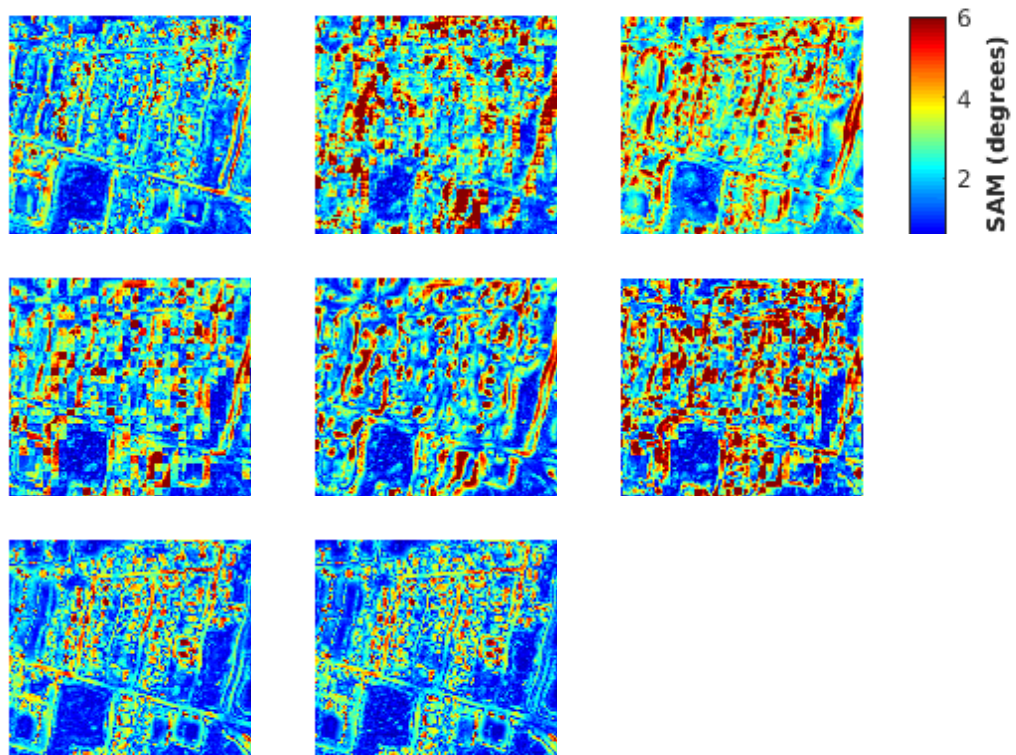
Figure 3.7: Mapping of the SAM error computed for each pixel for Chikusei dataset $128 \times 128 \times 93$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB and the result of the proposed method.

3.7 Using the L^1 norm in the radiometric term

In this chapter, we present a new variational fusion problem that is an upgraded version of the previous model introduced in Section 3.2. The introduced modifications in the new variational model aim at improving the previous fusion results. These modifications are based on observations and conclusions drawn from the previous variational model.

$$
\min_{u\in\mathbb{R}^{H\times N}} \sum_{h=1}^{H} \|\nabla_{\omega'_h} u_h\|_1 + \frac{\mu}{2}\sum_{h=1}^{H} \|DBu_h - g_h\|_2^2 +
$$
$$
\frac{\gamma}{2}\sum_{m=1}^{M} \|(Su)_m - f_m\|_2^2 + \frac{\lambda}{2}\sum_{h=1}^{H} \|\tilde{P}_h u_h - P_h\tilde{g}_h\|_1. \tag{3.22}
$$

The main modifications introduced in the fusion model (3.22) are the penalization with the L^1 norm of the radiometric constraint and the introduction of a new way of computing the similarity weights.

3.7.1 L^1 penalization of the radiometric constraint

Throughout different contributions in variational problems applied to image processing, the L^1-norm has proved efficient and even more efficient than the L^2-norm in many cases. In fact, in the literature it was shown that the L^1-norm prevents the edges from being smooth [146]. Some researchers also argued that both outliers and impulse noise are better removed when the residual control terms as well as the regularizer are in the L^1 norm [147, 148]. The authors in [149] combined the L^1 and L^2 norm in the regularization term as well as in the data fidelity term. They showed afterwards that the results are better than when only the L^1 norm or L^2 norm are used [149].

On one hand, the L^2-norm is quite effective in dealing with the Gaussian noise, however, it was shown by theoretical and numerical studies that the

L^2-norm is not effective for non-Gaussian additive noise and it has a tendency to amplify the effect of outliers in the final image [150]. On the other hand, it was demonstrated that the L^1-norm is quite suitable for non-Gaussian additive noise such as the impulsive and the Laplacian one [148, 151].

The terms implicated in the radiometric constraint $\mathbf{P}$, $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{g}}$ were the results of transformations of the initial data as it was already discussed before, hence, it is hard to predict and to know the noise introduced by these transformations. Therefore, the L^2-norm was maybe not adapted to the kind of noise fabricated during the data transformation process and thus not adapted for the minimization of $\tilde{P}_h u_h - P_h \tilde{g}_h$.

The introduction of the L^1-norm is not without consequences, because, unlike the L^2-norm, it is not differentiable. The non-differentiability of the L^1-norm causes algorithmic and numerical difficulties that might impact the speed of execution of the optimization code. Due to the introduction of the L^1-norm, the computation of the proximity operator and then the steps of updating the variables are not the same as with the L^2-norm.

3.7.2 New similarity weights

In the new variational model for fusion, a new way of computing the weights was introduced in definition (3.23) where this time the weights are computed on $\tilde{\mathbf{g}} \in \mathbb{R}^{H \times N}$. As a reminder, the latter is the hyperspectral image $\mathbf{g} \in \mathbb{R}^{H \times N_l}$ brought to the high resolution by bicubic interpolation.

For computational reasons, only the most similar pixels are kept for the computation of the gradient and for this, the previous way of computing the weights (3.13) is used for classification purposes. Let N_{sim} be the number of similar pixels wished to be kept. First, the weights are computed and classified following (3.13), and only the classification is kept because it is used afterwards.

Then, the final weights are computed with the formula (3.23) and classified following the order of the classification of the weights computed with (3.13).

Classifying the weights with the formula (3.13) and computing the final ones with (3.23), is a way to mix the information contained in both the hyperspectral and the multispectral image. In fact, the classification on the multispectral component gives the most similar pixels in terms of geometric details, then, this information is used to compute weights with the same pixels but with a richer spectral information.

$$
\omega'_{h,i,j} = \begin{cases} \dfrac{1}{\Gamma'_i} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{\text{spt}}^2} - \dfrac{\sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|\tilde{g}_{h,i+t} - \tilde{g}_{h,j+t}\|_2^2}{h_{\text{sim}}^2 (2\nu_c + 1)^2} \right) & \text{if } \|i-j\|_\infty \leqslant \nu_r \ (i \neq j) \\ 0 & \text{else} \end{cases}
$$
(3.23a)

$$
\Gamma'_i = \sum_{\{j : \|i-j\|_\infty \leqslant \nu_r\}} \exp\left(-\dfrac{\|i-j\|_2^2}{h_{\text{spt}}^2} - \dfrac{\sum_{\{t \in \mathbb{Z}^2 : \|t\|_\infty \leqslant \nu_c\}} \|\tilde{g}_{h,i+t} - \tilde{g}_{h,j+t}\|_2^2}{h_{\text{sim}}^2 (2\nu_c + 1)^2} \right)
$$
(3.23b)

$$
\omega'_{h,i,i} = \max\left\{ \omega_{h,i,j} : \|i-j\|_\infty \leqslant \nu_r \text{ and } j \neq i \right\}
$$
(3.23c)

3.7.3 Proximal operator of the radiometric constraint

In the case of the L^1-norm, the computation of the proximal operator related to the radiometric constraint is different from the one in the case of the L^2-norm. The proximal operator is computed as follows

$$
\begin{aligned}
\mathbf{u} = \operatorname{prox}_{\tau G}(\tilde{\mathbf{u}}) &\Leftrightarrow \mathbf{u} = \operatorname*{argmin}_{\mathbf{u}} \frac{1}{2\tau} \|\mathbf{u} - \tilde{\mathbf{u}}\|_2^2 + \frac{\lambda}{2} \|\tilde{\mathbf{P}}\mathbf{u} - \mathbf{P}\mathbf{g}\|_1 \\
&\Leftrightarrow u_h = \operatorname*{argmin}_{u_h} \frac{1}{2\tau} \|u_h - \tilde{u}_h\|_2^2 + \frac{\lambda}{2} \|\tilde{P}_h u_h - P_h g_h\|_1
\end{aligned}
$$
(3.24)

The L^1- norm is non-differentiable, then, there are three cases to be discussed:

i) if $\tilde{P}_h u_h - P_h g_h < 0$, from (3.24) we conclude

$$0 \in \frac{1}{\tau}(u_h - \tilde{u}_h) - \frac{\lambda}{2}\tilde{P}_h \Rightarrow u_h = \tilde{u}_h + \frac{\tau\lambda}{2}\tilde{P}_h \tag{3.25}$$

ii) if $\tilde{P}_h u_h - P_h g_h > 0$, we have

$$0 \in \frac{1}{\tau}(u_h - \tilde{u}_h) + \frac{\lambda}{2}\tilde{P}_h \Rightarrow u_h = \tilde{u}_h - \frac{\tau\lambda}{2}\tilde{P}_h \tag{3.26}$$

iii) if $\tilde{P}_h u_h - P_h g_h = 0$, we have

$$u_h = \frac{P_h \tilde{g}_h}{\tilde{P}_h} \tag{3.27}$$

3.7.4 The optimization algorithm

the new steps of the optimization algorithm that takes into account the fact that the radiometric constraint is penalized with the L^1-norm are the following

Algorithm 4: The primal-dual algorithm (L1-norm)

Input : $\mathbf{f}$, $\mathbf{g}$, $\tilde{\mathbf{g}}$, D, B, S, $\mathbf{P}$, $\tilde{\mathbf{P}}$, ∇_ω, τ and σ

1 for $i \leftarrow 0$ **to** *convergence* **do**

2   **for** $h \leftarrow 0$ **to** H **do**

3   *//Update of the primal variable* u_h *and the dual variables* q_h, q_h *and* r_h

4 $p_h^{i+1} = \dfrac{p_h^i(x) + \sigma \nabla_\omega \bar{u}_h^i}{\max(1, |p_h^i(x) + \sigma \nabla_\omega \bar{u}_h^i|)}$

5 $q_h^{i+1} = \dfrac{q_h^i + \sigma(DB\bar{u}_h^i - g_h)}{1 + \frac{\sigma}{\mu}}$

6 $r_m^{i+1} = \dfrac{g_l^i + \sigma((S\bar{u}^i)_m - f_m)}{1 + \frac{\sigma}{\gamma}}$

          • if $\tilde{P}_h u_h - P_h \tilde{g}_h < 0$

$$u_h^{i+1} = u_h^i + \tau \mathrm{div}_\omega p_h^{i+1} - \tau(B^\mathsf{T}D^\mathsf{T}q^{i+1})_h - \tau(S^\mathsf{T}r^{i+1})_h + \frac{\tau\lambda}{2}\tilde{P}_h \qquad (3.28)$$

          • if $\tilde{P}_h u_h - P_h \tilde{g}_h > 0$

$$u_h^{i+1} = u_h^i + \tau \mathrm{div}_\omega p_h^{i+1} - \tau(B^\mathsf{T}D^\mathsf{T}q^{i+1})_h - \tau(S^\mathsf{T}r^{i+1})_h - \frac{\tau\lambda}{2}\tilde{P}_h \qquad (3.29)$$

          • if $\tilde{P}_h u_h - P_h \tilde{g}_h = 0$

$$u_h^{i+1} = \frac{P_h \tilde{g}_h}{\tilde{P}_h} \qquad (3.30)$$

       $\bar{u}_h^{i+1} = 2u_h^{i+1} - u_h^i$

7   **end**

8 end

Output: $\mathbf{u}$ (the fused image)

3.8 Experimental results for the L^1 case

The experiments in this part are focused on images from the dataset Harvard [152]. The latter contains images taken by a hyperspectral camera of indoor and outdoor scenes. Each image has a spatial resolution of 1392×1040 with thirty-one spectral bands, but in this part, the images were cropped to a spatial resolution of 512×512. The reason behind choosing a non-satellite dataset is to test if the fusion algorithm is able to adapt to data that does not necessarily come from a satellite. The experiments were done with C++ code and the hyperspectral and multispectral images were generated following the generation model (1.7). The spectral response used for the generation of the multispectral image is the one of the Nikon D700 camera as shown in Figure 3.8. The results of the fusion in the case of the L^1 norm are compared to six state-of-the-art methods and to the fusion methods suggested in this thesis.
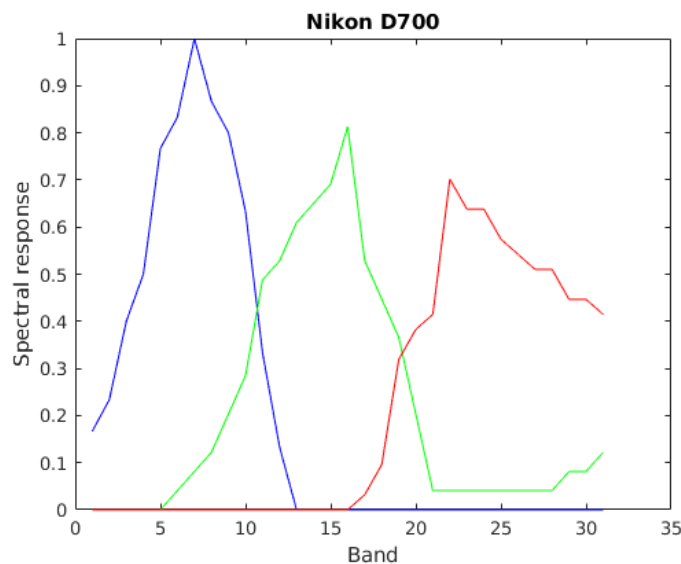

Figure 3.8: The spectral response of the camera Nikon D700

The experiments were carried out on two images "Bicycles" and "Bookshelves" from the Harvard dataset with SNR=35dB. Figure 3.9 shows the visual results

of fusion on the image "Bicycles". We can notice that the fused images of the state-of-the-art suffer either from blurring effects or color artifacts, see the color spot behind the blackbox in GSA image and the left side of the handlebar for the rest of the methods. The result provided by the L^1 norm looks very similar to the ground truth and is artifacts free. Note that, visually, the L^2 norm gave a result very similar to the one of the L^1 norm. On Figures 3.10 and 3.11, we have the mapping of the RMSE and SAM errors of the fusion results given in Figure 3.9. We can see that, visually, the RMSE error mapping corresponding to the L^1 norm is quite weak with respect to many methods and looks similar to the one of the L^2 norm. We can notice that the errors are mainly located in areas with saturation. As to the SAM mapping, we can see that the performance of the L^1 norm compares favorably to the rest of the methods. However, visually speaking, the SAM error of the L^2 norm seems to be weaker than the one provided by the L^1 norm.

Table 3.3 shows the numerical results of the different fusion products corresponding to the image "Bicycles". We can notice that, the L^1 and L^2 norms overcome all the other methods in all the quality indices. The L^2 norm still takes the lead in a few quality measures. However, the performance of the L^1 norm is better than the L^2's in some measures.

On Figure 3.12 we have the results of the fusion on the second image from Harvard dataset: "Bookshelves". We notice that the visual results provided by the L^1 and the L^2 norm are less noisy and blurry than the rest of the methods. Figures 3.13 and 3.14 show the mapping of the RMSE and SAM errors respectively. Visually speaking, the RMSE error corresponding to the L^1 norm is weaker than the rest of the presented methods including the L^2 norm. As to the SAM error, the performance of the L^1 norm compares favorably to the other methods, however, we can see a few red spots that show that the error

Figure 3.9: Performances of HS-MS fusion with the nonlocal-gadient-based method and comparison with other state of the art methods on Harvard dataset, image "Bicycles" of size $512 \times 512 \times 31$. From top left to bottom. Reference image, HS image, MS image, CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB, L^2 norm and L^1.

Figure 3.10: Mapping of the RMSE error computed for each pixel on Harvard dataset, image "Bicycles" of size $512 \times 512 \times 31$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB, L^2 norm and L^1 norm.

Figure 3.11: Mapping of the SAM error computed for each pixel on Harvard dataset, image "Bicycles" of size $512 \times 512 \times 31$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB, L^2 norm and L^1 norm.

	RMSE	PSNR	ERGAS	SAM	$Q2^n$	CC	DD
Reference	0	∞	0	0	1	1	0
CNMF	22,8271	42.5237	5.0822	2.1087	0.9343	0.9881	9.1502
HySure	10.8181	49.0092	2.7622	2.9111	0.9636	0.9964	6.3197
GSA	14.7971	46.2887	2.9249	3.2588	0.9577	0.9949	7.8039
SFIMHS	40.1161	37.6257	8.2207	2.3244	0.8936	0.9637	15.0971
GLPHS	36.3011	38.4936	7.4117	2.6596	0.9036	0.9704	15.3171
MAPSMM	45.6341	36.5066	9.2112	2.9852	0.8533	0.9536	17.8621
HMWB	10.3256	48.1241	2.6822	2.9062	0.9585	0.9971	6.3037
L^2 norm	**6.1401**	**53.9286**	<u>1.7511</u>	**1.6393**	<u>0.9755</u>	<u>0.9983</u>	**3.8729**
L^1 norm	<u>6.7306</u>	<u>53.1321</u>	**1.6705**	<u>1.7894</u>	**0.9763**	**0.9984**	<u>3.9155</u>

Table 3.3: Quantitative quality evaluation of each fused product on Harvard dataset, image "Bicycles" of size $512 \times 512 \times 31$ and SNR=35dB. RMSE and DD values are expressed in a magnitude of 10^{-9}.

becomes important. The visual results show that the SAM error is weaker for the L^2 norm than for the L^1 norm.

Table 3.4 shows the quantitative results corresponding to the image "Bookshelves". We can see that the fusion results provided by the L^1 and the L^2 norms perform better than the other methods in most of the quality measures. We can also notice that, for the "Bookshelves" image, the L^1 norm overtakes the L^2 norm in terms of spatial quality measures such as RMSE, PSNR and CC whereas the L^2 norm takes the lead when it comes to the spectral measure SAM. The quantitative results confirm the visual observations carried out on the RMSE and the SAM mappings.

We can conclude that for both images "Bookshelves" and "Bicycles", the L^2 norm provide good results in terms of spectral quality whereas the L^1 performs better in recovering the spatial details. For the image "Bookshelves" that contains more constant structures, the L^1 norm did well in recovering a fused result with a minimum amount of noise than the rest of the methods including the L^2 norm.
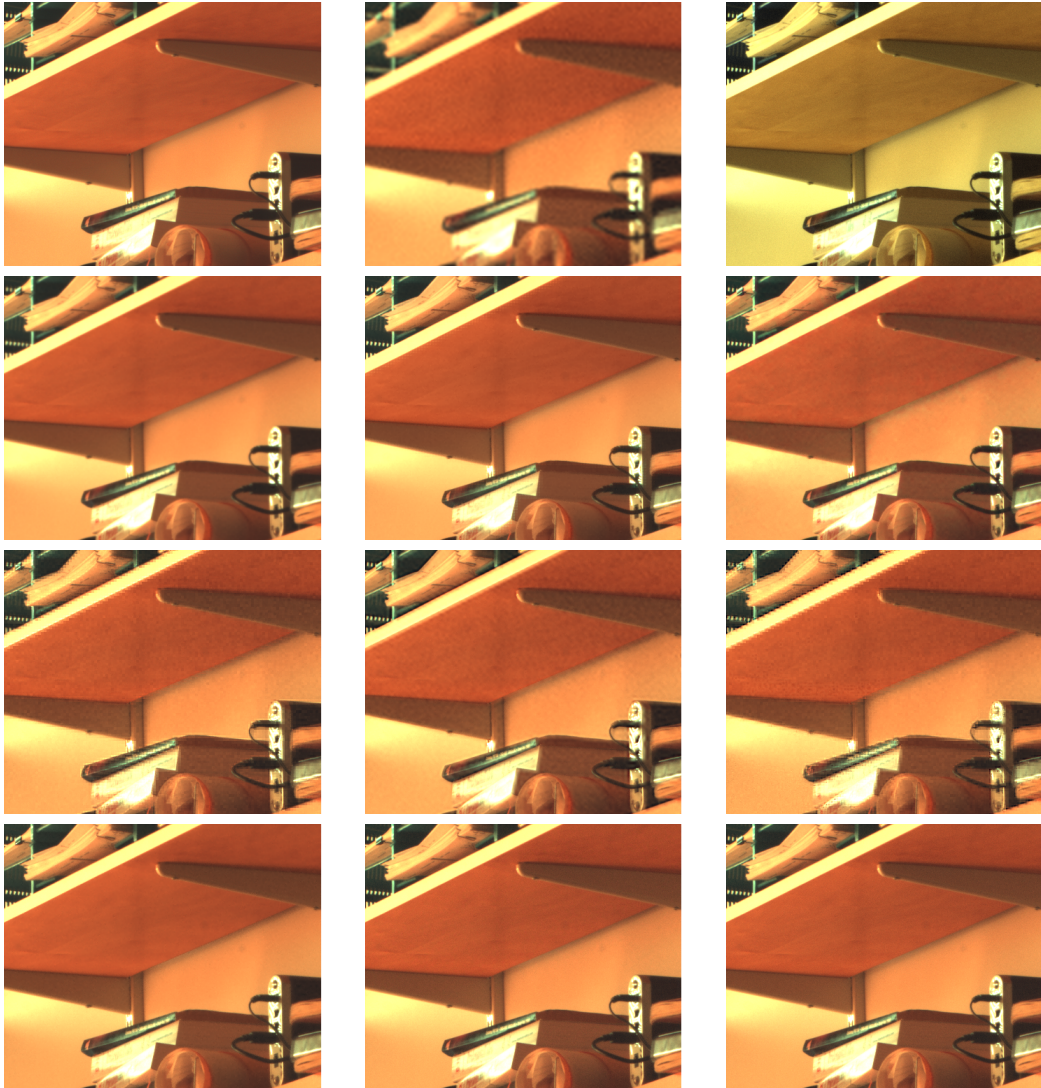
Figure 3.12: Performances of HS-MS fusion with the nonlocal-gradient based method and comparison with other state of the art methods on Harvard dataset, image "Bookshelves" of size $512 \times 512 \times 31$. From top left to bottom. Reference image, HS image, MS image, CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB, L^2 norm and L^1 norm.
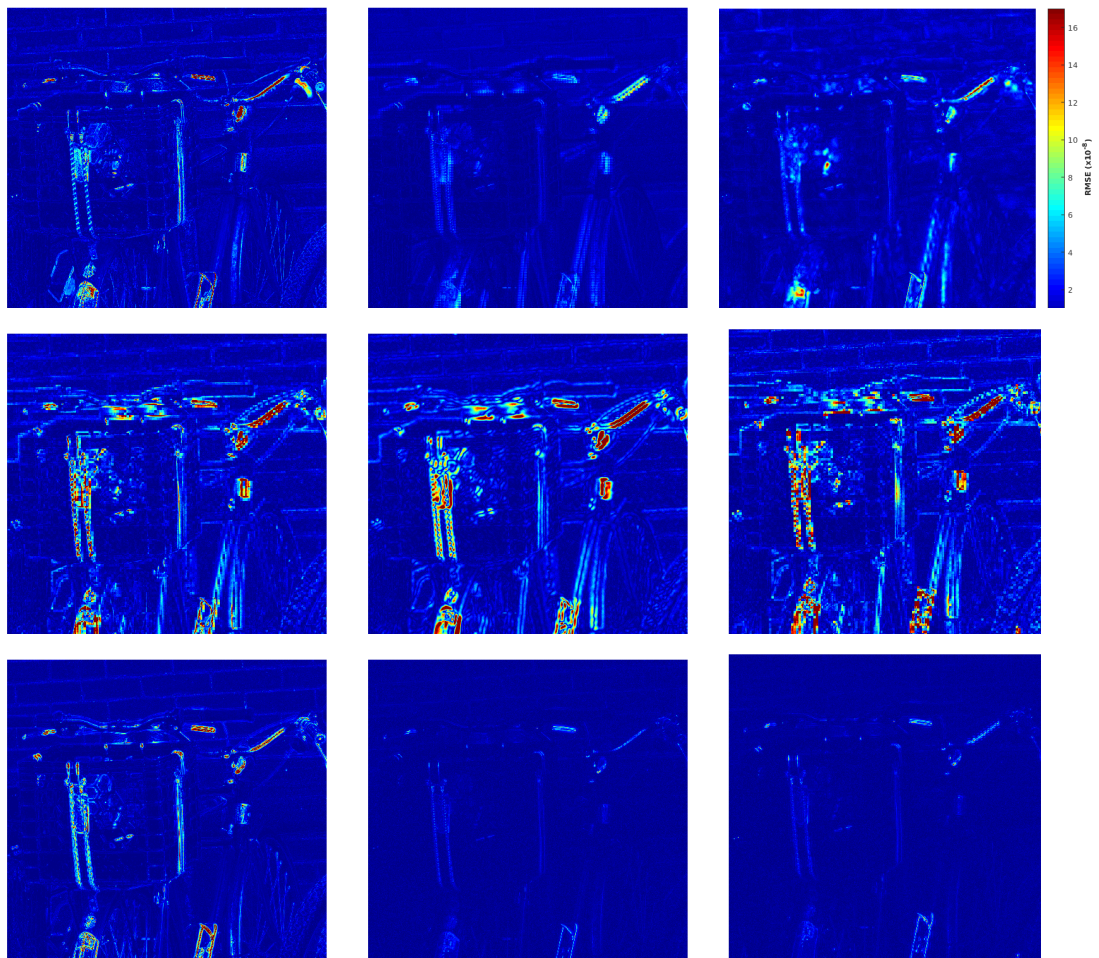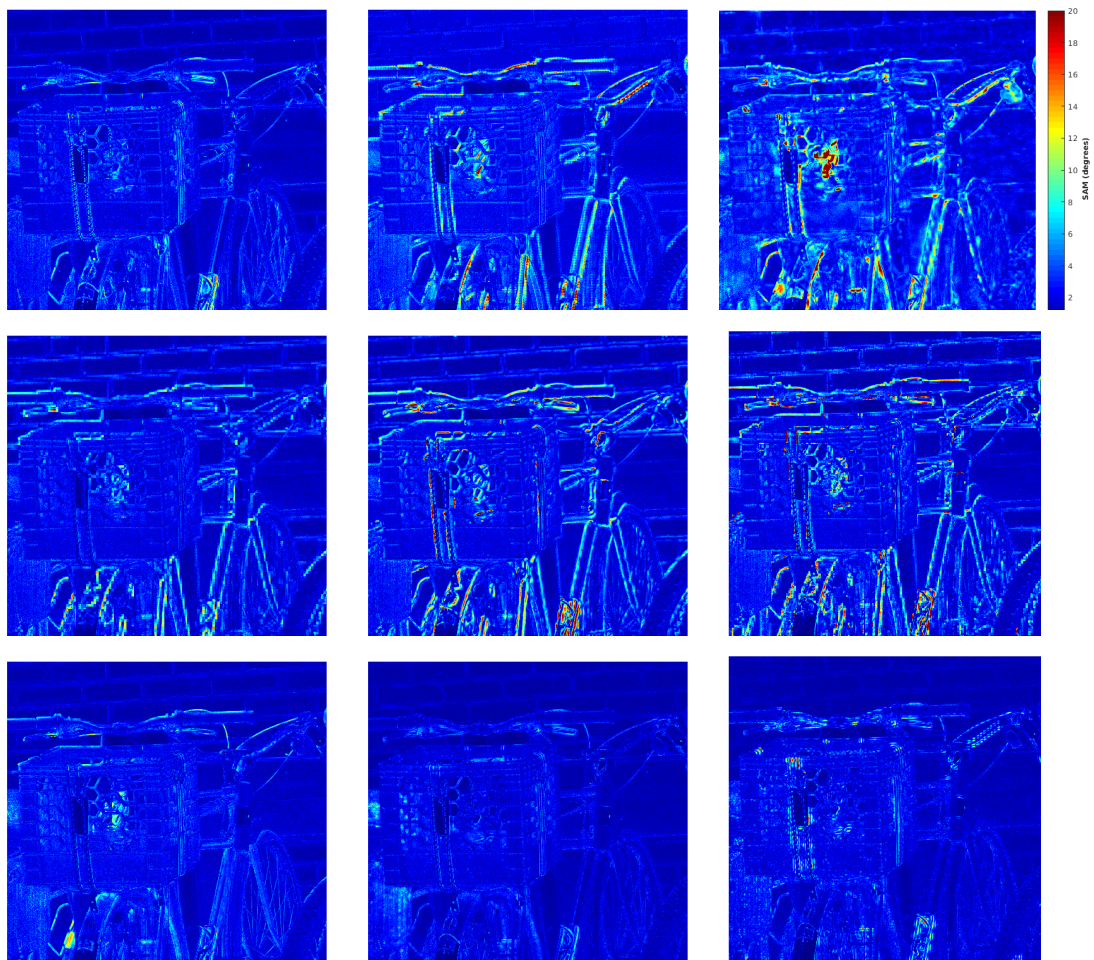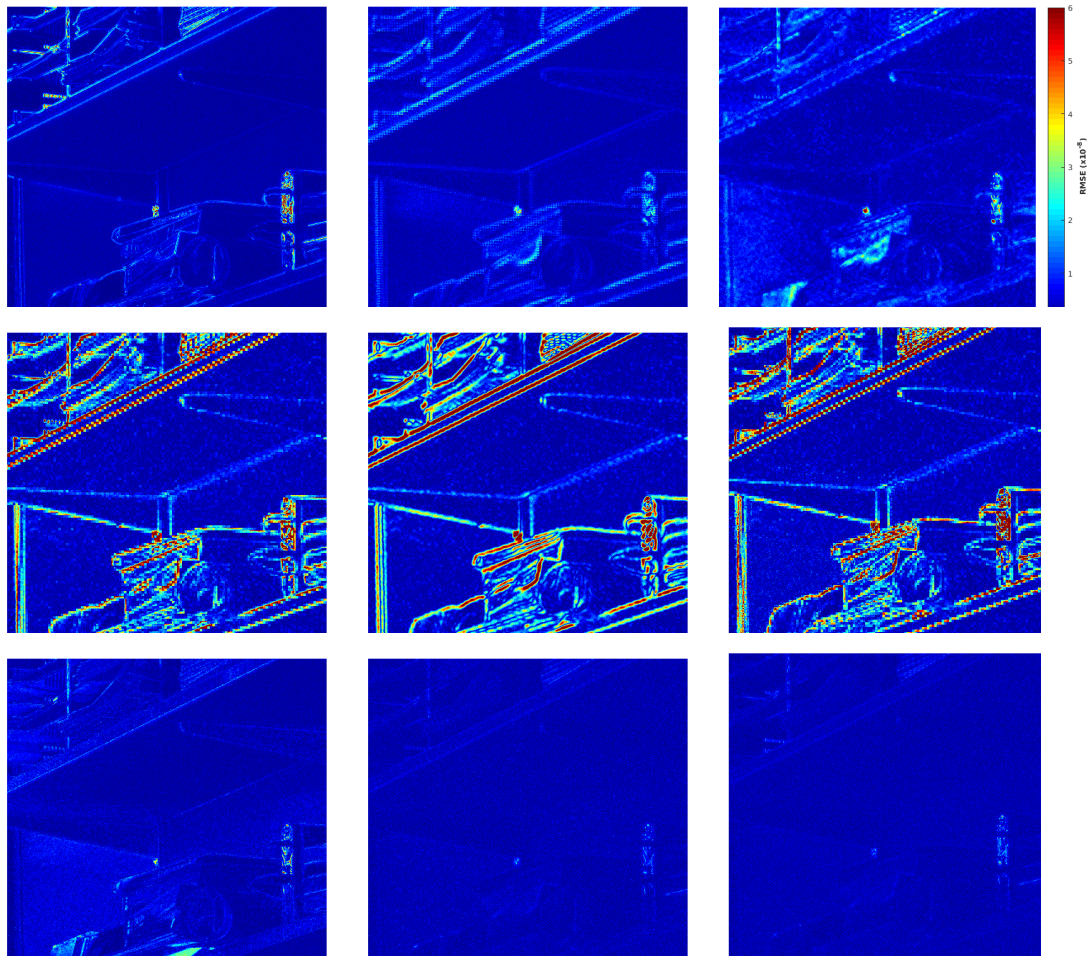
Figure 3.13: Mapping of the RMSE error computed for each pixel on Harvard dataset, image "Bookshelves" of size $512 \times 512 \times 31$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB, L^2 norm and L^1 norm.
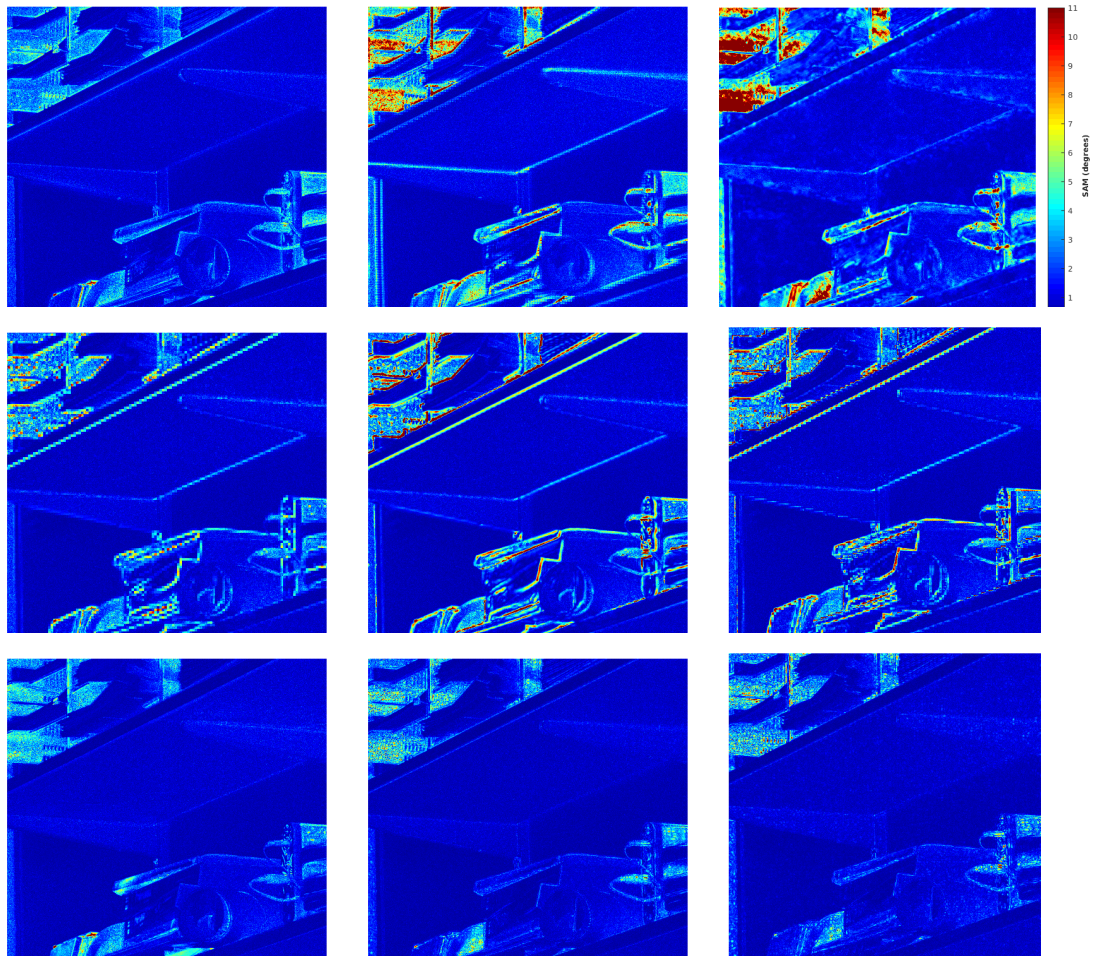
Figure 3.14: Mapping of the SAM error computed for each pixel on Harvard dataset, image "Bookshelves" of size $512 \times 512 \times 31$. From top left to bottom. CNMF, HySure, GSA, SFIMHS, GLPHS, MAPSMM, HMWB, L^2 norm and L^1 norm.

	RMSE	PSNR	ERGAS	SAM	$Q2^n$	CC	DD
Reference	0	∞	0	0	1	1	0
CNMF	5.0651	46.1166	1.2013	1.0964	<u>0.8995</u>	0.9972	3.4515
HySure	5.6362	45.1884	1.5163	1.7113	0.8916	0.9964	3.7267
GSA	6.4841	43.9712	1.3972	1.9213	0.8834	0.9958	4.3021
SFIMHS	18.7981	34.7261	3.9293	1.3359	0.8161	0.9681	9.0311
GLPHS	17.6361	35.2811	3.6893	1.6166	0.8269	0.9717	8.9936
MAPSMM	20.8921	33.8084	4.4446	1.7394	0.7927	0.9595	9.9244
HMWB	5.0875	46.1192	1.1788	1.1674	**0.9065**	0.9966	2.9568
L^2 norm	<u>3.5014</u>	<u>49.3234</u>	**0.8284**	**0.9747**	0.8966	**0.9985**	<u>2.5481</u>
L^1 norm	**3.4814**	**49.3731**	<u>0.8471</u>	<u>1.0474</u>	0.8959	**0.9985**	**2.5217**

Table 3.4: Quantitative quality evaluation of each fused product on Harvard dataset, image "Bookshelves" of size $512 \times 512 \times 31$ and SNR=35dB. RMSE and DD values are expressed in magnitude of 10^{-9}.

3.9 Concluding remarks

In this chapter, two HS-MS fusion techniques based on the nonlocal gradient were suggested. The first model consisted in minimizing a functional composed of four energy terms. The first term is a regularization one computed with the nonlocal gradient. The latter is expressed as the difference between two pixels times a similarity coefficient. This coefficient's role is to copy the geometry of the multispectral image and to inject it in the final fused image. The second and third terms are data fidelity ones, their goal is the penalize the deviation of the fused image from the input hyperspectral and multispectral image. The last term is a radiometric constraint that forces the fused and the multispectral images to share the same high spatial frequencies. The regularization term was penalized with the L^1 and the L^2 norm, the data fidelity terms and the radiometric constraint were penalized with the L^2 norm.

The experiments were carried out on two datasets: Chikusei and Urban. On Chikusei, the chosen SNR is 35dB. The visual result of the proposed method is very similar to the reference image, it is not corrupted by noise and does

not contains artifacts. The mappings of the RMSE and the SAM errors showed that the suggested fusion model provided a result with minimal errors with respect to the other methods. The quantitative results showed that the suggested method overtook all the methods in most of the quality measures. For Urban, the SNR is 30dB, which means that the noise is increased but is still moderate. The visual result shows that the fused image is not corrupted by noise. The RMSE and SAM error mappings showed that, compared to the rest of the methods, the errors of the suggested model are minimized. Also, the fusion result outperformed the other methods in all the quality indices. These results show that the suggested method is quite robust to noise.

In the second part of the chapter, a modification of the first model was suggested: the radiometric constraint was penalized with the L^1 norm in order to account for possible impulse noise, also, a new way of computing the similarity weights was suggested. The experiments were carried out on Harvard dataset which is not a remote sensing one and contains images of indoor and outdoor scenes taken by a hyperspectral camera. The experiment on the first image "Bicycles" showed that, the fused image is very similar to the reference one, it is denoised and does not contain color artifacts unlike the methods of the state-of-the-art. The RMSE mapping of the errors showed that the suggested method compares favorably to the other methods. However, in terms of SAM error mapping, the L^2 norm provided better visual results. The quantitative measures showed than the L^1 norm is very competitive compared to the other methods and overcame the L^2 in some quality indices.

The experiments on the second image "Bookshelves" showed that the fused image is denoised, is artifacts free and contains all the geometrical information present in the reference image. The RMSE mapping showed that the error present in the result with the L^1 norm is visually weaker than the ones present

in all the other methods. The SAM mapping of the L^1 norm compares favorably to the other methods, however, it seems to be stronger than the one provided by the result with the L^2 norm. The quantitative results showed that L^1 norm compares favorably to the other methods and performs better than the L^2 norm in some quality indices. From those results we can also see that, on one hand, the L^2 norm gives better results when it comes to the SAM measure. On the other hand, the measures show that the L^1 norm performs better in denoising and preserving the spatial quality of the fused image, which confirms the visual observations.

Chapter 4

Conclusions

This thesis explored the fusion of hyperspectral and multispectral images within two frameworks: optimal transport and nonlocal methods. Due to various sensor-specifics, as discussed in Section 1.3.2: on-board storage capacity, transmission bandwidth, financial constraints, etc, imaging satellites capture images with either a good spatial resolution and a low spectral one (multispectral) or images with a low spatial resolution and a high spectral one (hyperspectral). Thus, image fusion aims at combining the rich spatial resolution of the multispectral image and the high spectral information of the hyperspectral one into one high-resolution hyperspectral image. The fused image has superior quality and is visually better, in terms of spatial and spectral resolution, than any of the input images. High-resolution images are very useful in many missions such as classification, facial recognition, forests mapping, etc. In this thesis, image fusion was carried out with techniques from the optimal transport theory and from nonlocal-based methods. The present work focused on satellite images. In the last chapter, the nonlocal fusion technique was also tested on images taken by hyperspectral cameras of indoor and outdoor scenes.

In Chapter 2, methods from the optimal transport theory were employed to model and solve the image fusion problem. The latter was formulated as the

minimization of the sum of two regularized Wasserstein distances. Each one of these distances describes either the spectral or the spatial information to be injected in the final fused image. The injected information is controlled with coefficients associated with the regularized Wasserstein distances. The solution of the minimization problem is the barycenter in the sense of the Wasserstein metric solved with Sinkhorn-Knopp algorithm. In the optimal transport model, the noise is not taken into account, thus, a denoising preprocessing step was added in order to downplay the effect of noise. The fusion model contains five parameters in total. These parameters were optimized with objective measures on three datasets: Pavia, Chikusei and Urban, the optimal parameters were fixed for all the experiments. The fusion result was evaluated visually and with quality indices on two datasets: Pavia and Urban. Also, the suggested method was compared to six state-of-the-art fusion models. For both datasets, the visual results showed that the fused image is close to the reference one and artifacts free. Also, RMSE and SAM error mappings were visualized in order to study the distribution of the errors across the images. The results showed that the errors of the proposed method are visually less important than those of the other methods. The fused image was compared to six state-of-the-art techniques with various quality measures. The proposed method outperformed the other ones in all the measures. In future work, it would be interesting to upgrade the fusion model so that it can take into account the minimization of noise, without having to apply a preprocessing denoising step.

In the second part of Chapter 2, a modification of the first model was suggested. The modification consisted in introducing nonlocal means (NLM) weights in the cost matrix. The goal behind these weights was to take into account the values of pixels during the transport in order to improve the latter and to account for the additive noise that affects the data. With the adding of

the NLM weights, the total number of parameters was eight. Thus, a study of the sensitivity of parameters was carried out to have a first idea about the behavior of the new fusion technique. Various experiments were conducted on three datasets: Chikusei, Botswana and Pavia. The results of these experiments showed that the NLM weights did not contribute to improving the previous fusion results. In fact, the different figures showed that the best results the new technique can reach are the ones of the model in the absence of the NLM weights. Nevertheless, the NLM weights might be useful in downplaying misregistration problems in the case where the hyperspectral and the multispectral data are not aligned. The study of the contribution of the NLM weights under misalignment assumptions will be carried out in future work.

In Chapter 3, two variational techniques, for hyperspectral and multispectral image fusion, based on the nonlocal gradient were suggested. The first fusion problem was modeled as the minimization of an energy functional composed of four energy terms. The first term is a regularization one, computed with the nonlocal gradient. The latter is computed as the difference between two pixels times a similarity coefficient computed from the multispectral image. The goal of the nonlocal term is to copy the geometric information of the multispectral image into the fused one. The second and third terms are data fidelity ones, they penalize the deviation of the fused image from the input hyperspectral and multispectral images. The last term is called radiometric constraint, its goal is to force the fused and the multispectral image to share the same high spatial frequencies. The nonlocal term was penalized with the L^1 and the L^2 norms, the data fidelity terms as well as the radiometric constraint were penalized with the L^2 norm. The optimal parameters of this model were found by exhaustive search and they were fixed for all the experiments. The first experiment was carried out on Chikusei dataset with a SNR of 35dB, the results showed that, visually, the

proposed method is similar to the reference one and is not corrupted by noise or artifacts. The mappings of RMSE and SAM errors showed that the suggested method has fewer errors with respect to the other techniques. The quantitative results showed that the suggested model outperformed all the methods in most of the quality indices. In order to test the robustness to noise of our model, the experiments were run on Urban dataset with a SNR of 30dB. The visual results showed that the fused image is not corrupted by noise. The RMSE and the SAM error mappings showed that the errors of the proposed method are minimal compared to the other methods. Also, the fused result outperformed the other methods in all the quality indices. These results highlight the robustness of the suggested model to noise.

The second fusion technique developed in Chapter 3 was a modification of the first one: the radiometric constraint was penalized with the L^1 norm in order to account for possible impulse noise, moreover, a new way of computing the similarity weights, that mixes information from the hyperspectral and the multispectral image, was suggested. The experiments in this part were run on two images, "Bicycles" and "Bookshelves", from outdoor and indoor scenes with a SNR of 35dB. The results on "Bicycles" showed that, visually, the fused image is similar to the reference one, is denoised and artifacts free, unlike some state-of-the-art methods. The RMSE and SAM mappings showed that the L^1 norm model compares favorably to the other methods, even though, the result of the L^2 norm model had less errors in terms of SAM than the rest of the methods. In terms of quantitative results, the L^1 norm is very competitive with respect to the other methods and outperformed the L^2 norm in many quality indices. Similarly to "Bicycles", the experiments on "Bookshelves" showed that, visually, the fused image contains the color information and all the geometric details present in the reference image. Moreover, the fusion result is denoised and

artifacts free. On one hand, the RMSE mapping showed that the error present in the result with the L^1 norm is weaker than the results of the other methods. On the other hand, the SAM mapping of the result with the L^1 norm compares favorably to the other methods but it stays stronger than the one of the result with the L^2 norm. The quality measures showed that the result with the L^1 norm compares favorably to the other methods and outperforms the result with the L^2 norm in many quality indices. To sum up, the first model seems to give good results when it comes to recovering the spectral information, whereas the model with the L^1 norm performs better in denoising and recovering the spatial details. Future work will be focused on studying further the difference between both models and their contributions to fusing non-remote sensing and also remote sensing data.

The present thesis explored various techniques for fusing hyperspectral and multispectral images. The experiments focused mainly on remote sensing data, but in the last chapter, the fusion was also tested on non-remote sensing data and it proved efficient. Thus, an interesting line of research might be to test the studied methods on fusing different types of data such as SAR (Synthetic-Aperture Radar) and optical images.

Appendix A

Appendix of chapter 2

A.1 Computations in the case of a triply block-Toepltiz matrix

The link between the multiplication matrix-vector and the convolution was made in the case of a Toeplitz matrix, except that in our case, the matrix ξ is block-Toeplitz and the multiplication matrix-vector is handled differently.

Let us go back to the matrix ξ corresponding to the $2 \times 2 \times 2$ grid as highlighted in figure (2.6). As illustrated in figure (A.1), ξ is a block-Toeplitz and symmetric matrix that can be seen as a 4×4 matrix with elements being themselves matrices. The matrix ξ can be written under the following form

$$\xi = \begin{pmatrix} \xi_0 & \xi_1 \\ \xi_1 & \xi_0 \end{pmatrix} \tag{A.1}$$

Where ξ_0 and ξ_1 are matrices.

As shown in figure (A.2), each element of ξ is itself a 2×2 block-Toeplitz and symmetric matrix that can be written as follows

$$\xi_0 = \begin{pmatrix} \xi_{0,0} & \xi_{0,1} \\ \xi_{0,1} & \xi_{0,0} \end{pmatrix} \qquad \xi_1 = \begin{pmatrix} \xi_{1,0} & \xi_{1,1} \\ \xi_{1,1} & \xi_{1,0} \end{pmatrix} \tag{A.2}$$

1	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{2}{\gamma}}$	$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(2+\alpha)}$
$e^{-\frac{1}{\gamma}}$	1	$e^{-\frac{2}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{1}{\gamma}(2+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$
$e^{-\frac{1}{\gamma}}$	$e^{-\frac{2}{\gamma}}$	1	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(2+\alpha)}$	$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$
$e^{-\frac{2}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	1	$e^{-\frac{1}{\gamma}(2+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{\alpha}{\gamma}}$
$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(2+\alpha)}$	1	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{2}{\gamma}}$
$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{1}{\gamma}(2+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}}$	1	$e^{-\frac{2}{\gamma}}$	$e^{-\frac{1}{\gamma}}$
$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(2+\alpha)}$	$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{2}{\gamma}}$	1	$e^{-\frac{1}{\gamma}}$
$e^{-\frac{1}{\gamma}(2+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{1}{\gamma}(1+\alpha)}$	$e^{-\frac{\alpha}{\gamma}}$	$e^{-\frac{2}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	$e^{-\frac{1}{\gamma}}$	1

Figure A.1: The matrix ξ is block symmetric

So in the end, the matrix ξ can be written under the form

$$\xi = \begin{pmatrix} \xi_{0,0} & \xi_{0,1} & \xi_{1,0} & \xi_{1,1} \\ \xi_{0,1} & \xi_{0,0} & \xi_{1,1} & \xi_{1,0} \\ \xi_{1,0} & \xi_{1,1} & \xi_{0,0} & \xi_{0,1} \\ \xi_{1,1} & \xi_{1,0} & \xi_{0,1} & \xi_{0,0} \end{pmatrix} \tag{A.3}$$

where $\xi_{0,i}$ and $\xi_{1,i}$ are 2×2 Toeplitz matrices with $i \in \{0,1\}$

Let us consider in a general setting an image of size $n_l \times n_c \times b$ where n_l, n_c and b are the number of rows, columns and spectral bands respectively. The cost matrix ξ of this image is computed in a row-major order, it is block-Toeplitz, symmetric and it is defined as follows

$$\xi = \begin{pmatrix} \xi_0 & \xi_1 & \xi_2 & \dots & \xi_{b-1} \\ \xi_1 & \xi_0 & \xi_1 & \dots & \xi_{b-2} \\ \xi_2 & \xi_1 & \xi_0 & \dots & \xi_{b-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \xi_{b-1} & \dots & \xi_2 & \xi_1 & \xi_0 \end{pmatrix} \tag{A.4}$$
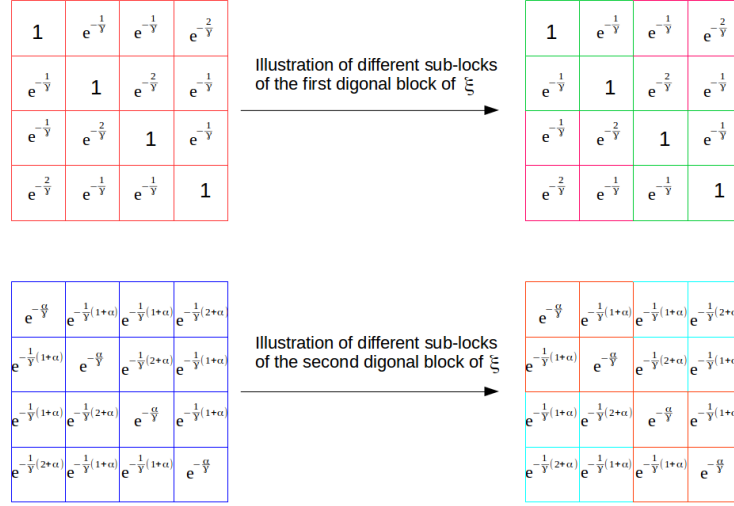
Illustration of different sub-locks of the first digonal block of ξ →

Illustration of different sub-locks of the second digonal block of ξ →

Figure A.2: Sub-matrices of ξ

Where ξ_i are themselves block-Toeplitz and symmetric matrices of size $n_l \times n_l$ with $i \in [\![0, b-1]\!]$

Let F be the following block-vector

$$
F = \begin{pmatrix} F_0 \\ F_1 \\ \vdots \\ F_{b-1} \end{pmatrix}
\tag{A.5}
$$

where F_i are block-vectors of size $n_l \times 1$ and $i \in [\![1, b-1]\!]$.

Let X be a vector defined as follows

$$
X = \begin{bmatrix}
X(2b-2) & X(2b-3) & & X(b) & X(b-1) & X(b-2) & & X(1) & X(0) \\
\xi_{b-1} & \xi_{b-2} & \cdots & \xi_1 & \xi_0 & \xi_1 & \cdots & \xi_{b-2} & \xi_{b-1}
\end{bmatrix}
\tag{A.6}
$$

The block-convolution between a block-vector X that contains block-matrices

and a block-vector F that contains block-vectors at the instant n is defined as follows

Definition 1.

$$Y(n) = (X * F)(n) = \sum_{i=0}^{b-1} X(i)F(n-i) \tag{A.7}$$

Remark. The block convolution (A.7) implicates multiplication operations between matrices and vectors, thus there is a certain way for the multiplication to be respected which makes the convolution not commutative.

The block-matrix-block-vector product $\xi \times F$ is as follows

$$\xi \times F = \begin{pmatrix} \xi_0 F_0 + \xi_1 F_1 + \ldots + \xi_{b-1}F_{b-1} \\ \xi_1 F_0 + B_0 F_1 + \ldots + \xi_{b-2}F_{b-1} \\ \vdots \\ \xi_{b-1}F_0 + \xi_{b-2}F_1 + \ldots + \xi_0 F_{b-1} \end{pmatrix} = \begin{pmatrix} Y(b-1) \\ Y(b) \\ \vdots \\ Y(2b-2) \end{pmatrix} \tag{A.8}$$

From the structure of the ξ matrix we noticed that the following relationship can be made between the matrix ξ_0 and the matrices $(\xi_i)_{i \in [\![1,b-1]\!]}$

$$\xi_i = e^{-i\frac{\alpha}{\gamma}} \xi_0$$

Thus the vector X defined in (A.6) becomes

$$X = [\xi_{b-1} \ \xi_{b-2} \ \ldots \ \xi_1 \ \xi_0 \ \xi_1 \ \ldots \ \xi_{b-2} \ \xi_{b-1}]$$
$$= [e^{-(b-1)\frac{\alpha}{\gamma}} \ e^{-(b-2)\frac{\alpha}{\gamma}} \ \ldots \ 1 \ \ldots e^{-(b-2)\frac{\alpha}{\gamma}} \ e^{-(b-1)\frac{\alpha}{\gamma}}] \otimes \xi_0$$

where $\otimes$ is the kronecker product.
we define

$$V = [\overset{V(2b-2)}{e^{-(b-1)\frac{\alpha}{\gamma}}} \ \overset{V(2b-3)}{e^{-(b-2)\frac{\alpha}{\gamma}}} \ \ldots \ \overset{V(b)}{e^{-\frac{\alpha}{\gamma}}} \ \overset{V(b-1)}{1} \ \overset{V(b-2)}{e^{-\frac{\alpha}{\gamma}}} \ \ldots \ \overset{V(1)}{e^{-(b-2)\frac{\alpha}{\gamma}}} \ \overset{V(0)}{e^{-(b-1)\frac{\alpha}{\gamma}}}]$$

Proposition 3. *The multiplication* (A.8) *becomes*

$$\xi \times F = (X * F)((b-1) \to (2b-2))$$

$$= \left(\left(V \otimes \xi_0 \right) * F \right)((b-1) \to (2b-2))$$

$$= \left(V * \left(\xi_0 * F \right)(0 \to b-1) \right)((b-1) \to (2b-2))$$

Proof. We define $R = \xi_0 * F$, following the definition of the block-convolution defined in (A.7), the computation of R is as follows

$$R(n) = (\xi_0 * F)(n) = \xi_0 F(n) \tag{A.9}$$

Thus, the result of the convolution (A.9) is a block vector defined as

$$R = \begin{bmatrix} \overset{R(0)}{\xi_0 F_0} & \overset{R(1)}{\xi_0 F_1} & \ldots & \overset{R(b-2)}{\xi_0 F_{b-2}} & \overset{R(b-1)}{\xi_0 F_{b-1}} \end{bmatrix}$$

We define $Q = V * R$, the computation of Q is as follows

$$Q(n) = (V * R)(n) = \sum_{i=0}^{b-1} R(i)V(n-i)$$

$$Q(0) = R(0)V(0) + R(1)V(-1)$$
$$Q(1) = R(0)V(1) + R(1)V(0)$$
$$\vdots$$
$$\left. \begin{aligned} Q(b-1) &= R(0)V(b-1) + R(1)V(b-2) + \ldots + R(b-1)V(0) \\ Q(b) &= R(0)V(b) + R(1)V(b-1) + \ldots + R(b-1)V(1) \\ &\vdots \\ Q(2b-2) &= R(0)V(2b-2) + R(1)V(2b-1) + \ldots + R(b-1)V(b-1) \end{aligned} \right\} = (\xi \times F)$$
$$Q(2b-1) = R(1)V(2b-2) + \ldots + R(b-1)V(b)$$
$$\vdots$$

$\blacksquare$

From the structure of the matrices $\xi_{i,j}$ where $i \in [\![0, b-1]\!]$ and $j \in [\![0, n_l - 1]\!]$, the following relationship between $\xi_{i,j}$ and $\xi_{0,0}$ is established

$$\xi_{i,j} = e^{-\frac{(i-j)^2}{\gamma}}\xi_{0,0}$$

Thus, the multiplication (A.8) can be further simplified by studying the operation $R = \xi_0 \times F$ defined in (A.9). By repeating the same reasoning as for $\xi \times F$ for each of the components of the vector R, the operation $\xi_0 \times F_i$, where $i \in [\![0, b-1]\!]$, can be written as follows

$$\xi_0 \times F_i = (X' * F_i)((n_l - 1) \to (2n_l - 2))$$
$$= \left(V' * (\xi_{0,0} * F_i)(0 \to n_l - 1) \right)((n_l - 1) \to (2n_l - 2))$$

where

$$X' = \begin{bmatrix} \overset{X'(2n_l-2)}{\xi_{0,n_l-1}} & \overset{X'(2n_l-3)}{\xi_{0,n_l-2}} & \cdots & \overset{X'(n_l)}{\xi_{0,1}} & \overset{X'(n_l-1)}{\xi_{0,0}} & \overset{X'(n_l-2)}{\xi_{0,1}} & \cdots & \overset{X'(1)}{\xi_{0,n_l-2}} & \overset{X'(0)}{\xi_{n_l-1}} \end{bmatrix}$$

$$= [e^{\frac{-(n_l-1)^2}{\gamma}} \quad e^{\frac{-(n_l-2)^2}{\gamma}} \quad \cdots \quad 1 \quad \cdots e^{\frac{-(n_l-2)^2}{\gamma}} \quad e^{\frac{-(n_l-1)^2}{\gamma}}] \otimes \xi_{0,0}$$

and

$$V' = \begin{bmatrix} \overset{V'(2n_l-2)}{e^{\frac{-(n_l-1)^2}{\gamma}}} & \overset{V'(2n_l-3)}{e^{\frac{-(n_l-2)^2}{\gamma}}} & \cdots & \overset{V'(n_l)}{e^{-\frac{1}{\gamma}}} & \overset{V'(n_l-1)}{1} & \overset{V'(n_l-2)}{e^{-\frac{1}{\gamma}}} & \cdots & \overset{V'(1)}{e^{\frac{-(n_l-2)^2}{\gamma}}} & \overset{V'(0)}{e^{\frac{-(n_l-1)^2}{\gamma}}} \end{bmatrix}$$

and

$$F_i = \begin{pmatrix} F_{0,i} \\ F_{1,i} \\ \vdots \\ F_{n_l-1,i} \end{pmatrix}$$

where $F_{0,i}$ are vectors of size $n_c \times 1$ and $i \in [\![0, b-1]\!]$.

We define $R'_i = \xi_{0,0} * F_i$, R'_i is computed as follows

$$R'_i(n) = (\xi_{0,0} * Fi)(n) = \xi_{0,0}F_i(n)$$

With $n \in [\![0, n_l - 1]\!]$ and R'_i is defined as

$$R'_i = \begin{bmatrix} \overset{R'_i(0)}{\xi_{0,0}F_{0,i}} & \overset{R'_i(1)}{\xi_{0,0}F_{1,i}} & \cdots & \overset{R'_i(n_l-2)}{\xi_{0,0}F_{n_l-2,i}} & \overset{R'_i(n_l-1)}{\xi_{0,0}F_{n_l-1,i}} \end{bmatrix}$$

Let us recall that the matrix $\xi_{0,0}$ is Toeplitz, the multiplication $\xi_{0,0}F_i(n)$ can be then computed using the convolution as shown in (2.5.1) in the case of non block-Toeplitz matrices. In this case, the storing of the matrix $\xi_{0,0}$ is not needed but only a vector as the one defined in (2.5.1).

Thus, we can conclude that all the components $(\xi_{0,0}F_{j,i})_{j\in[\![0,n_l-1]\!]}$ of R'_i can be computed independently and in parallel, after that, the convolution with the vector V' can be then carried out. Therefore, all the components $(R(i))_{i\in[\![0,b-1]\!]}$ can also be computed independently and in parallel, the application of the vector at the end allows the computation of the final multiplication $\xi \times F$.

As one can notice, the decomposition of the matrix-vector multiplication into convolutions thanks to the interesting properties of the matrix (block-Toeplitz and symmetric), allows the gain of an important execution time thanks to the possibility of parallelization of the computations. It also allows the saving of an important memory space because, according to the previous computations, only three vectors are needed to carry out the convolution: V, V' and the vector x drawn from the sub-sub matrix $\xi_{0,0}$ that allows the computation of the last convolutions $(\xi_{0,0}F_{j,i})_{j\in[\![0,n_l-1]\!]}$. Moreover, these last three vectors are symmetric, which means that only half of them can be stored and the other half can be concluded by doing the symmetry.

A.2 Algorithm for the search of optimal parameters when using NLM weights

The algorithm presented below is used for studying the sensitivity of parameters when using the NLM weights in the cost matrix. The study of the NLM weights in the cost matrix as well as the experiments where this algorithm

is used are described in sections (2.9) and (2.10).

Algorithm 5: Optimisation of parameters with the NLM weights

  Input :$\widetilde{\mathbf{S}}(\mathbf{f}),\widetilde{\mathbf{T}}(\mathbf{g})$, γ_M, γ_H, α_M, α_H, s

1 $\xi_M^{dist} \leftarrow$ compute_kernel_dist (multispectral component),

2 $\xi_H^{dist} \leftarrow$ compute_kernel_dist (hyperspectral component),

3 array_sizes_patches={(1×1),(3×3),(5×5), (3×3×3),(5×5×5)},

4 array_filtering_parameters=logspaced[-12,4]

5 **for** *i← 0* **to** *length (array_sizes_patches)* **do**

6   **for** *j← 0* **to** *length (array_filtering_parameters)* **do**

7         $\xi_M^{dist} \leftarrow$ compute_kernel_NLM $\Big(\widetilde{\mathbf{S}}(\mathbf{f})$, array_sizes_patches(i),

          array_filtering_parameters(j), γ_M, α_M, s $\Big)$

8         $\xi_H^{dist} \leftarrow$ compute_kernel_NLM $\Big(\widetilde{\mathbf{T}}(\mathbf{g})$, array_sizes_patches(i),

          array_filtering_parameters(j), γ_H, α_H, s $\Big)$

9         $\xi_M \leftarrow \xi_M^{dist} \odot \xi_M^{NLM}$

10        $\xi_H \leftarrow \xi_H^{dist} \odot \xi_H^{NLM}$

11        $\mathbf{u} \leftarrow$ compute_barycenter $\Big(\widetilde{\mathbf{S}}(\mathbf{f}),\widetilde{\mathbf{T}}(\mathbf{g}),\gamma_M, \gamma_H, \alpha_M, \alpha_H \Big)$

12   res(i,j)← $\mathbf{u}$

13   **end**

14 **end**

  Output:res

References

[1] J. Mifdal, B. Coll, N. Courty, J. Froment, and B. Vedel, "Hyperspectral and multispectral Wasserstein barycenter for image fusion," in *Proc. IEEE Geosci. Remote Sens. Symp. (IGARSS)*, (Texas, TX, USA), 2017.

[2] J. Mifdal, B. Coll, and J. Duran, "A variational formulation for hyperspectral and multispectral image fusion," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3328–3332, IEEE, 2018.

[3] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B*, vol. 207, no. 1167, pp. 187–217, 1980. 4

[4] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, pp. 106–154, 1962. 4

[5] C. H. Grohnfeldt, *Multi-sensor data fusion for multi-and hyperspectral resolution enhancement based on sparse representations*. PhD thesis, Technische Universität München, 2017. xxvii, 7

[6] I. Kant, *Critique of pure reason*. Cambridge University Press, 1998. 9

[7] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New Haven Yale University Press, 1923. 11

[8] A. Ben-Israel and T. N. Greville, *Generalized inverses: theory and applications*, vol. 15. Springer Science & Business Media, 2003. 13

[9] A. N. Tihonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math.*, vol. 4, pp. 1035–1038, 1963. 13

[10] A. N. Tikhonov and V. I. Arsenin, *Solutions of ill-posed problems*, vol. 14. Vh Winston, 1977. 13

[11] H. Van Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 44–51, IEEE, 2010. 15

[12] D. Wu and D.-W. Sun, "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—part i: Fundamentals," *Innovative Food Science & Emerging Technologies*, vol. 19, pp. 1–14, 2013. 15

[13] W. Turner, S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger, "Remote sensing for biodiversity science and conservation," *Trends in ecology & evolution*, vol. 18, no. 6, pp. 306–314, 2003. 15

[14] S. T. Thurman and J. R. Fienup, "Analysis of the general image quality equation," in *Visual Information Processing XVII*, vol. 6978, p. 69780F, International Society for Optics and Photonics, 2008. 17

[15] J. C. Leachtenauer and R. G. Driggers, *Surveillance and reconnaissance imaging systems: modeling and performance prediction*. Artech House, 2001. 17

[16] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal processing magazine*, vol. 19, no. 1, pp. 17–28, 2002. 18

[17] K. Navulur, *Multispectral image analysis using the object-oriented paradigm*. CRC press, 2006. 18

[18] W. Dou, Y. Chen, X. Li, and D. Z. Sui, "A general framework for component substitution image fusion: An implementation using the fast image fusion method," *Computers & Geosciences*, vol. 33, no. 2, pp. 219–228, 2007. 19

[19] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 79, 2011. 19, 33

[20] L. Loncan, L. Almeida, J. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J.-Y. Tourneret, M. Veganzoes, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, 2015. 19, 34

[21] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for ikonos imagery," *IEEE Geoscience and Remote sensing letters*, vol. 1, no. 4, pp. 309–312, 2004. 19

[22] M. González-Audícana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved ihs and pca mergers based on wavelet decomposition," *IEEE Transactions on Geoscience and Remote sensing*, vol. 42, no. 6, pp. 1291–1299, 2004. 19

[23] M. Joshi and A. Jalobeanu, "Map estimation for multiresolution fusion in remotely sensed images using an igmrf prior model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1245–1255, 2010. 19

[24] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 738–746, 2011. 19

[25] M. Cetin and N. Musaoglu, "Merging hyperspectral and panchromatic image data: qualitative and quantitative analysis," *International Journal of Remote Sensing*, vol. 30, no. 7, pp. 1779–1804, 2009. 19

[26] G. A. Licciardi, M. M. Khan, J. Chanussot, A. Montanvert, L. Condat, and C. Jutten, "Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear pca band reduction," *EURASIP Journal on Advances in Signal processing*, vol. 2012, no. 1, p. 207, 2012. 19

[27] R. Molina, A. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, 1999. 21

[28] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 29–56, 2017. 29

[29] J. M. Facil, A. Concha, L. Montesano, and J. Civera, "Deep single and direct multi-view depth fusion," *CoRR, abs/1611.07245*, 2016. 30

[30] M. Tao, L. Yang, Y. Gu, and S. Cheng, "Object-oriented change detection based on change magnitude fusion in multitemporal very high resolution images," in *2017 9th International Conference on Modelling, Identification and Control (ICMIC)*, pp. 418–423, IEEE, 2017. 30

[31] M. S. Farid, A. Mahmood, and S. A. Al-Maadeed, "Multi-focus image fusion using content adaptive blurring," *Information Fusion*, vol. 45, pp. 96–112, 2019. 31

[32] L. Guanter, H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner, C. Chlebek, *et al.*, "The enmap spaceborne imaging spectroscopy mission for earth observation," *Remote Sensing*, vol. 7, no. 7, pp. 8830–8857, 2015. 31

[33] A. Iwasaki, N. Ohgi, J. Tanii, T. Kawashima, and H. Inada, "Hyperspectral imager suite (hisui)-japanese hyper-multi spectral radiometer," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pp. 1025–1028, IEEE, 2011. 31

[34] P. Stefano, P. Angelo, P. Simone, R. Filomena, S. Federico, S. Tiziana, A. Umberto, C. Vincenzo, N. Acito, D. Marco, *et al.*, "The prisma hyperspectral mission: Science activities and opportunities for agriculture and land monitoring," in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pp. 4558–4561, IEEE, 2013. 31

[35] R. O. Green, G. Asner, S. Ungar, and R. Knox, "Nasa mission to measure global plant physiology and functional types," in *Aerospace Conference, 2008 IEEE*, pp. 1–7, IEEE, 2008. 31

[36] S. Michel, M.-J. LEFEVRE-FONOLLOSA, and S. HOSFORD, "Hypxim–a hyperspectral satellite defined for science, security and defence users," *PAN*, vol. 400, no. 800, p. 400, 2011. 31

[37] A. Eckardt, J. Horack, F. Lehmann, D. Krutz, J. Drescher, M. Whorton, and M. Soutullo, "Desis (dlr earth sensing imaging spectrometer for the iss-muses platform)," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, pp. 1457–1459, IEEE, 2015. 31

[38] R. Haydn, G. W. Dalke, J. Henkel, and J. Bare, "Application of the ihs color transform to the processing of multisensor data and image enhancement," in *Proceedings of the International Symposium on Remote Sensing of Environment, First Thematic Conference: Remote sensing of arid and semi-arid lands, 19-25 January, 1982, Cairo, Egypt*, Ann Arbor, Mich.: Center Remote Sens. Information & Analysis, Environ. Res. Inst., Mich., 1982., 1982. 32

[39] V. PohlC, "Multisensorimagefusioninremotesensing: concepts, methodsandapplications," *InternationalJournalofRemoteSensing*, vol. 19, no. 5, p. 823r854, 1998. 32

[40] W. CARPER, T. LILLESAND, and R. KIEFER, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogrammetric Engineering and remote sensing*, vol. 56, no. 4, pp. 459–467, 1990. 32

[41] P. Chavez, S. C. Sides, J. A. Anderson, *et al.*, "Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic," *Photogrammetric Engineering and remote sensing*, vol. 57, no. 3, pp. 295–303, 1991. 32

[42] M. Ehlers, "Multisensor image fusion techniques in remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 46, no. 1, pp. 19–30, 1991. 32

[43] R. A. Mandhare, P. Upadhyay, and S. Gupta, "Pixel-level image fusion using brovey transforme and wavelet transform," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 6, pp. 2690–2695, 2013. 32

[44] W. A. Hallada and S. Cox, "Image sharpening for mixed spatial and spectral resolution satellite systems," 1983. 32

[45] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," Jan. 4 2000. US Patent 6,011,875. 32

[46] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, 2007. 32, 33

[47] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295–309, 2011. 32

[48] D. A. Yocky, "Image merging and data fusion by means of the discrete two-dimensional wavelet transform," *JOSA A*, vol. 12, no. 9, pp. 1834–1841, 1995. 32

[49] B. Garguet-Duport, J. Girel, J.-M. Chassery, and G. Patou, "The use of multiresolution analysis and wavelets transform for merging spot panchromatic and multispectral image data," *Photogrammetric Engineering and remote sensing*, vol. 62, no. 9, pp. 1057–1066, 1996. 32

[50] J. Zhou, D. Civco, and J. Silander, "A wavelet transform method to merge landsat tm and spot panchromatic data," *International journal of remote sensing*, vol. 19, no. 4, pp. 743–757, 1998. 32

[51] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989. 32

[52] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and statistics*, pp. 281–299, Springer, 1995. 32

[53] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *IEEE Transactions on signal processing*, vol. 40, no. 10, pp. 2464–2482, 1992. 32

[54] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2091–2106, 2005. 32

[55] J.-L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 297–309, 2007. 32

[56] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," in *Readings in Computer Vision*, pp. 671–679, Elsevier, 1987. 32

[57] T. A. Wilson, S. K. Rogers, and M. Kabrisky, "Perceptual-based image fusion for hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, 1997. 32

[58] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Transactions on geoscience and remote sensing*, vol. 40, no. 10, pp. 2300–2312, 2002. 32

[59] Q. Wei, N. Dobigeon, and J. Tourneret, "Bayesian fusion of multiband images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–11227, 2015. 33

[60] Q. Wei, N. Dobigeon, and J. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, (Florence, Italy), pp. 3176–3180, 2014. 33

[61] M. Simoes, J. Bioucas-Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2015. xxix, 33, 35, 37, 79

[62] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015. xxix, 33, 35, 37, 75

[63] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, 2006. 33

[64] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1301–1312, 2008. 33

[65] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, R. Restaino, G. Licciardi, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2015. 33, 34

[66] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 446–453, 2011. 33

[67] J. Duran, A. Buades, B. Coll, C. Sbert, and G. Blanchet, "A survey of pansharpening methods with a new band-decoupled variational model," *ISPRS J. Photogramm. Remote Sens.*, vol. 125, pp. 78–105, 2017. 33, 34, 113, 115, 121

[68] R. B. Gomez, A. Jazaeri, and M. Kafatos, "Wavelet-based hyperspectral and multispectral image fusion," in *Geo-Spatial Image and Data Exploitation II*, vol. 4383, pp. 36–43, International Society for Optics and Photonics, 2001. 34

[69] Y. Zhang and M. He, "Multi-spectral and hyperspectral image fusion using 3-d wavelet transform," *Journal of Electronics (China)*, vol. 24, no. 2, pp. 218–224, 2007. 34

[70] Z. Chen, H. Pu, B. Wang, and G.-M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1418–1422, 2014. 34

[71] J. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000. xxix, 34, 79

[72] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of ms + pan data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, 2007. xxix, 34, 79

[73] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006. xxix, 35, 79

[74] M. T. Eismann and R. C. Hardie, "Application of the stochastic mixing model to hyperspectral resolution enhancement," *IEEE transactions on geoscience and remote sensing*, vol. 42, no. 9, pp. 1924–1933, 2004. 35

[75] M. T. Eismann, *Resolution enhancement of hyperspectral imagery using maximum a posteriori estimation with a stochastic mixing model*. 2004. xxix, 35, 79

[76] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, 2009. 35

[77] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1212–1226, 1999. 36

[78] H. N. Gross and J. R. Schott, "Application of spectral mixture analysis and image fusion techniques for image sharpening," *Remote Sensing of Environment*, vol. 63, no. 2, pp. 85–94, 1998. 36

[79] O. Berné, A. Helens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pp. 1–4, IEEE, 2010. 36

[80] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2012. xxix, 36, 79

[81] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 2329–2336, IEEE, 2011. 36

[82] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 1409–1413, IEEE, 2013. 36

[83] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3586–3594, 2015. 37

[84] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *European Conference on Computer Vision*, pp. 63–78, Springer, 2014. 37

[85] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1117–1127, 2015. 37

[86] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a sylvester equation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4109–4121, 2015. 37

[87] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, 2017. 38

[88] J. Yang, Y.-Q. Zhao, and J. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sensing*, vol. 10, no. 5, p. 800, 2018. 38

[89] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016. 38

[90] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*, pp. 391–407, Springer, 2016. 38

[91] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016. 38

[92] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," *arXiv preprint arXiv:1607.07680*, 2016. 38

[93] Y. Zhao, G. Li, W. Xie, W. Jia, H. Min, and X. Liu, "Gun: Gradual upsampling network for single image super-resolution," *IEEE Access*, vol. 6, pp. 39363–39374, 2018. 38

[94] N. Bonneel, G. Peyré, and M. Cuturi, "Wasserstein barycentric coordinates: histogram regression using optimal transport.," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 71–1, 2016. 41

[95] J. Rabin and N. Papadakis, "Convex color image segmentation with optimal transport distances," in *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 256–269, Springer, 2015. 41, 46

[96] J. Delon, "Movie and video scale-time equalization application to flicker reduction," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 241–248, 2006. 41

[97] F. Pitié, A. C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," *Computer Vision and Image Understanding*, vol. 107, no. 1-2, pp. 123–137, 2007. 41

[98] N. Bonneel, M. Van De Panne, S. Paris, and W. Heidrich, "Displacement interpolation using lagrangian mass transport," in *ACM Transactions on Graphics (TOG)*, vol. 30, p. 158, ACM, 2011. 41, 46

[99] J. Lellmann, D. A. Lorenz, C. Schonlieb, and T. Valkonen, "Imaging with kantorovich–rubinstein discrepancy," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2833–2859, 2014. 41

[100] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in neural information processing systems*, pp. 2292–2300, 2013. 41, 48

[101] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000. 41, 46, 48

[102] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, "Wasserstein propagation for semi-supervised learning," in *International Conference on Machine Learning*, pp. 306–314, 2014. 41, 46

[103] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 9, pp. 1853–1865, 2017. 41

[104] J. Rabin, J. Delon, and Y. Gousseau, "Regularization of transportation maps for color and contrast transfer," *2010 IEEE International Conference on Image Processing*, pp. 1933–1936, 2010. xxviii, 41

[105] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas, "Convolutional wasserstein distances: Efficient optimal transportation on geometric domains," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 66, 2015. xxviii, 41, 42, 65

[106] C. Villani, *Topics in optimal transportation*. No. 58, American Mathematical Soc., 2003. 42, 46, 48

[107] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Histoire de l'Académie Royale des Sciences de Paris*, 1781. 42

[108] F. Santambrogio, "Optimal transport for applied mathematicians," *Birkäuser, NY*, pp. 99–102, 2015. 45

[109] G.-S. Xia, S. Ferradans, G. Peyré, and J.-F. Aujol, "Synthesizing and mixing stationary gaussian texture models," *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 476–508, 2014. 46

[110] N. Courty, R. Flamary, D. Tuia, and T. Corpetti, "Optimal transport for data fusion in remote sensing," in *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pp. 3571–3574, IEEE, 2016. 46

[111] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, "A reconstruction of the initial conditions of the universe by optimal mass transportation," *Nature*, vol. 417, no. 6886, p. 260, 2002. 46

[112] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment problems, revised reprint*, vol. 106. Siam, 2012. 47

[113] G. Loeper and F. Rapetti, "Numerical solution of the monge–ampère equation by a newton's algorithm," *Mij*, vol. 1, p. 1, 2005. 47

[114] N. Papadakis, G. Peyré, and E. Oudet, "Optimal transport with proximal splitting," *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 212–238, 2014. 47

[115] E. Schrödinger, *Über die Ümkehrung der Naturgesetze*. Verlag der Akademie der Wissenschaften, in Kommission bei Walter de Gruyter, 1931. 47

[116] A. G. Wilson, "The use of entropy maximising models, in the theory of trip distribution, mode split and route split," *Journal of Transport Economics and Policy*, pp. 108–126, 1969. 47

[117] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices," *Pacific Journal of Mathematics*, vol. 21, no. 2, pp. 343–348, 1967. 48

[118] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *The annals of mathematical statistics*, vol. 35, no. 2, pp. 876–879, 1964. 48

[119] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, "Iterative bregman projections for regularized transportation problems," *SIAM Journal on Scientific Computing*, vol. 37, no. 2, pp. A1111–A1138, 2015. 48, 49, 52, 56

[120] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR computational mathematics and mathematical physics*, vol. 7, no. 3, pp. 200–217, 1967. 48

[121] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 435–446, Springer, 2011. 48

[122] M. Agueh and G. Carlier, "Barycenters in the wasserstein space," *SIAM Journal on Mathematical Analysis*, vol. 43, no. 2, pp. 904–924, 2011. 49

[123] J. Bigot, T. Klein, *et al.*, "Consistent estimation of a population barycenter in the wasserstein space," *ArXiv e-prints*, 2012. 49

[124] M. Cuturi and A. Doucet, "Fast computation of wasserstein barycenters," in *International Conference on Machine Learning*, pp. 685–693, 2014. 49

[125] R. Sinkhorn, "Diagonal equivalence to matrices with prescribed row and column sums," *The American Mathematical Monthly*, vol. 74, no. 4, pp. 402–405, 1967. 53

[126] F. A. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 145–163, 1993. 72

[127] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?," in *Third conference" Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, pp. 99–103, SEE/URISCA, 2000. 73

[128] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002. 74

[129] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 442–446, 2016. 76

[130] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005. 85, 95, 112

[131] J. Froment, "Parameter-free fast pixelwise non-local means denoising," *Image Processing On Line*, vol. 4, pp. 300–326, 2014. 86

[132] V. Aurich and J. Weule, "Non-linear gaussian filters performing edge preserving diffusion," in *Mustererkennung 1995*, pp. 538–545, Springer, 1995. 88

[133] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005. 88

[134] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Tech. Rep. SAL-2016-05-27, Space Application Laboratory, University of Tokyo, Japan, May 2016. 92

[135] T. F. Chan and J. J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*, vol. 94. Siam, 2005. 108

[136] K. Chen, *Matrix preconditioning techniques and applications*, vol. 19. Cambridge University Press, 2005. 108

[137] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring images: matrices, spectra, and filtering*, vol. 3. Siam, 2006. 108

[138] J. Modersitzki, *FAIR: flexible algorithms for image registration*, vol. 6. SIAM, 2009. 108

[139] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*, vol. 153. Springer Science & Business Media, 2006. 108

[140] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on pure and applied mathematics*, vol. 42, no. 5, pp. 577–685, 1989. 108

[141] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992. 108

[142] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for p+ xs image fusion," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 43–58, 2006. 108

[143] J. Duran, A. Buades, B. Coll, and C. Sbert, "A nonlocal variational model for pansharpening image fusion," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 761–796, 2014. 112

[144] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011. 122, 125, 126

[145] R. Rockafellar, *Convex Analysis*. Princeton Landmarks in Mathematics and Physics, Princeton University Press, 1997. 124

[146] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996. 137

[147] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE transactions on image processing*, vol. 13, no. 10, pp. 1327–1344, 2004. 137

[148] M. Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 99–120, 2004. 137, 138

[149] K. Özkan and E. Seke, "Combination of l 1 and l 2 norms for image deconvolution problems," in *Electrical and Electronics Engineering (ELECO), 2011 7th International Conference on*, pp. II–148, IEEE, 2011. 137

[150] C. A. Micchelli, L. Shen, Y. Xu, and X. Zeng, "Proximity algorithms for image models ii: L1/tv denosing," *Advances in Computational Mathematics, online version available*, 2011. 138

[151] S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of the absolute norm," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 913–917, 1997. 138

[152] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *CVPR 2011*, pp. 193–200, IEEE, 2011. 142