

Segmentation d’image par optimisation proximale

Laurent CONDAT¹, Nelly PUSTELNIK²

¹Univ. Grenoble Alpes & CNRS, GIPSA-Lab, F-38000 Grenoble

²Laboratoire de Physique de l’École Normale Supérieure de Lyon & CNRS, F-69364 Lyon Cedex 07
Laurent.Condat@gipsa-lab.grenoble-inp.fr, nelly.pustelnik@ens-lyon.fr

Résumé – L’optimisation proximale par éclatement permet de minimiser une somme arbitraire de fonctions convexes, en appelant individuellement leur gradient ou opérateur proximal. Ce dernier doit donc avoir une forme calculable, ce qui oblige à éclater le problème pour faire apparaître des fonctions suffisamment simples, au prix d’un ralentissement de la convergence. Dans ce travail, nous étudions une formulation convexe robuste de la segmentation d’une image couleurs et montrons que l’opérateur proximal, en apparence compliqué, qui apparaît, est calculable exactement et rapidement. En limitant l’éclatement, on obtient ainsi une convergence bien plus rapide de la méthode.

Abstract – Proximal splitting methods for optimization allow to minimize an arbitrary sum of convex functions, by individual application of their gradient or proximity operator. So, the latter must be computable, which forces to split the problem in sufficiently simple functions, at the cost of slow convergence. In this work, we study a robust convex formulation of color image segmentation and we show that the proximity operator, which appears and seems complicated, can be computed exactly and quickly. So, by limiting splitting, we get a much faster convergence.

1 Contexte, motivation et contribution

Durant ces dix dernières années, l’essor des méthodes d’optimisation convexe non lisse a essentiellement été guidé par des problématiques de restauration d’image [1–8]. La difficulté majeure consiste alors à proposer des algorithmes permettant de converger vers le minimiseur d’une somme de fonctions ayant des propriétés distinctes, par exemple une fonction différentiable et une fonction non lisse. Ces méthodes procèdent par éclatement (*splitting*) [9, 10] : elles opèrent en appelant individuellement le gradient ou l’opérateur proximal des fonctions, qui doit donc être calculable. Mais si le nombre de fonctions dont on veut minimiser la somme peut être arbitraire, la convergence est généralement d’autant plus lente que ce nombre est important, autrement dit que l’algorithme est éclaté. On a donc intérêt à regrouper les fonctions. S’il existe de nombreuses formes explicites pour les opérateurs proximaux associées à des fonctions usuelles de la littérature des problèmes inverses [9–11], seuls des configurations bien particulières conduisent à des formes explicites pour une somme de fonctions [12]. Dans ce travail, nous nous intéressons à la fonction suivante. Soit un entier non nul K , un vecteur $c \in \mathbb{R}^K$ et une fonction h convexe, semi-continue inférieurement et propre de \mathbb{R}^N dans $]-\infty, +\infty]$. On définit alors la fonction convexe f qui, pour tout $p \in \mathbb{R}^K$ vaut

$$f(p) = c^\top p + \sum_{k=1}^K h(p_k) \quad \text{s.t.} \quad p_1 \geq \dots \geq p_K. \quad (1)$$

Dans le paragraphe suivant, nous motiverons l’intérêt de cette étude à travers un problème de segmentation, ou partitionnement (*clustering*), d’image. La Section 3 détaillera le résultat théorique associé au calcul de l’opérateur proximal de la fonc-

tion f . Le gain obtenu en utilisant cet opérateur, en termes de vitesse de convergence, sera illustré en Section 4.

2 Approche considérée

Pour résoudre le problème de segmentation, de nombreuses approches variationnelles ont été proposées dans la littérature [13–18]. Parmi toutes ces méthodes, la plupart sont basées sur des critères non convexes et, par conséquent, la solution dépend fortement de l’initialisation choisie.

Le problème original de Chan-Vese [14, 17, 19], auquel nous nous intéressons ici, consiste à segmenter une image y à C canaux ($C = 3$ pour une image couleurs) en Q zones $\Omega_1, \dots, \Omega_Q$ distinctes, caractérisées par leur vraisemblance $f_q : \mathbb{R}^C \rightarrow \mathbb{R}$. Le problème de minimisation consiste à résoudre

$$\begin{aligned} \min_{\Omega_1, \dots, \Omega_Q} \quad & \sum_{q=1}^Q \int_{\Omega_q} -\log f_q(y(x)) dx + \frac{1}{2} \sum_{q=1}^Q \text{Per}(\Omega_q) \\ \text{s.t.} \quad & \begin{cases} \bigcup_{q=1}^Q \Omega_q = \Omega, \\ (\forall q \neq p), \Omega_q \cap \Omega_p = \emptyset, \end{cases} \end{aligned} \quad (2)$$

où la pénalisation $\text{Per}(\Omega_q)$ impose à la solution d’avoir un périmètre minimal et où la contrainte assure une partition sans recouvrement du domaine Ω de l’image. Dans plusieurs travaux [19, 20], il apparaît qu’une relaxation convexe et discrète efficace de ce problème, pour une image $y \in (\mathbb{R}^C)^N$ sous forme vectorisée (avec N le nombre de pixels), est de la forme

$$\min_{\theta} \sum_{q=1}^Q (-\log f_q(y))^\top (\theta_q - \theta_{q+1}) + \lambda \text{MTV}(\theta) \quad \text{s.t.} \quad 1 \geq \theta_2 \geq \dots \geq \theta_Q \geq 0, \quad (3)$$

où les θ_q sont des vecteurs de \mathbb{R}^N , $\theta_1 \equiv 1$, $\theta_{Q+1} \equiv 0$, $\theta = [\theta_2^\top \dots \theta_Q^\top]^\top \in \mathbb{R}^{(Q-1)N}$, les inégalités sur les θ_Q s'entendent pixel à pixel, f_q est appliquée pixel à pixel, $\lambda > 0$ contrôle le niveau de régularisation spatiale et la variation totale multilabels $\text{MTV}(\theta)$ vaut, au choix,

$$\text{MTV}(\theta) = \sum_{q=2}^Q \text{TV}(\theta_q), \quad (4)$$

$$\text{MTV}(\theta) = \sum_{q=1}^Q \text{TV}(\theta_q - \theta_{q+1}), \quad (5)$$

voire une forme plus complexe basée sur des convolutions infimales [20], où $\text{TV}(u)$ est la fonctionnelle usuelle de variation totale isotrope appliquée à une image scalaire $u \in \mathbb{R}^N$:

$$\text{TV}(u) = \sum_{n=1}^N \|(Du)_n\|_2 \quad (6)$$

avec $D : \mathbb{R}^N \rightarrow (\mathbb{R}^2)^N$ l'opérateur de différentiation discrète horizontale/verticale.

Lorsque l'on veut segmenter l'image en zones de labels (couleurs) fixées $\delta_q \in \mathbb{R}^C$, $q = 1, \dots, Q$, un choix typique pour l'attache aux données du problème (3) est de prendre $-\log f_q(y_n) = \|y_n - \delta_q\|_2^2$, pour chaque valeur de pixel $y_n \in \mathbb{R}^C$. Pour chaque pixel d'indice n et chaque $q = 1, \dots, Q$, on peut interpréter $(\theta_q - \theta_{q+1})_n \in [0, 1]$ comme la proportion de label δ_n dans l'image segmentée. Ainsi, cette dernière s'écrit

$$v = \sum_{q=1}^Q (\theta_q - \theta_{q+1}) \delta_q. \quad (7)$$

La solution algorithmique proposée pour résoudre le problème (3) dans [21] est basée sur un algorithme primal-dual de type Arrow-Hurwicz mais il requiert des itérations internes. Dans un contexte d'estimation de carte de disparité [22], et parallèlement pour la détection de changement de régularité locale [23], un algorithme proximal a été proposé : (3) est reformulé comme une somme de quatre fonctions faisant intervenir des opérateurs linéaires, permettant d'effectuer un changement de variable de la forme $\alpha_q = \theta_q - \theta_{q+1}$. Cette reformulation conduit à des fonctions dont les opérateurs proximaux ont une forme explicite. Deux algorithmes sont implémentés : un algorithme proximal primal PPXA+ [24] et un algorithme proximal primal-dual [25]. Une autre alternative, proposée dans [23], vise à reformuler (3), également comme une somme de quatre fonctions, mais en considérant une autre décomposition basée sur l'éclatement de la contrainte de décroissance.

Dans ce travail nous proposons de réécrire (3) comme une somme de deux fonctions

$$\min_{\theta} g(\theta) + \lambda \text{MTV}(\theta) \quad (8)$$

Algorithme pour résoudre le Problème 3.1

1. Affecter $\chi^{\text{top}} := (1, 1, r_1)$ et former la pile $\mathcal{S} := \{\chi^{\text{top}}\}$.
 2. Pour $k = 2, \dots, K$, faire {
 3. Affecter $\chi := (k, k, r_k)$.
 4. Tant que \mathcal{S} n'est pas vide et $r_\chi \geq r_{\chi^{\text{top}}}$, faire {
 5. Mettre à jour $r_\chi := (r_\chi(k - \chi_1 + 1) + r_{\chi^{\text{top}}}(\chi_2^{\text{top}} - \chi_1^{\text{top}} + 1)) / (k - \chi_1^{\text{top}} + 1)$.
 6. Mettre à jour $\chi := (\chi_1^{\text{top}}, k, r_\chi)$.
 7. Enlever (POP) χ^{top} de \mathcal{S} .
 8. Ajouter (PUSH) χ en haut de \mathcal{S} .
 9. Pour chaque élément $\chi = (\chi_1, \chi_2, r_\chi)$ de \mathcal{S} et pour tout $k = \chi_1, \dots, \chi_2$, affecter $\hat{z}_k := r_\chi$.
-

FIGURE 1 – Algorithme prenant en entrée un vecteur $r \in \mathbb{R}^K$ et calculant le vecteur \hat{z} solution du Problème 3.1. Il consiste à maintenir une pile \mathcal{S} de segments constants de \hat{z} , qui sont fusionnés au fur et à mesure pour assurer la décroissance de \hat{z} . Chaque segment de \mathcal{S} est représenté par un triplet $\chi = (\chi_1, \chi_2, r_\chi) : \chi_1$ et χ_2 sont les indices de début et de fin du segment et r_χ est la valeur prise par \hat{z} sur le segment ; r_χ vaut toujours $(\sum_{k=\chi_1}^{\chi_2} r_k) / (\chi_2 - \chi_1 + 1)$, soit la moyenne de r sur le segment. χ_{top} indique toujours l'élément en haut de \mathcal{S} .

avec la fonction g égale à la somme sur les pixels de la fonction f définie en (1) :

$$g(\theta) = \sum_{n=1}^N \sum_{q=2}^Q c_{n,q} \theta_{q,n} + h(\theta_{q,n}) \quad \text{s.t.} \quad \theta_{2,n} \geq \dots \geq \theta_{Q,n}, \quad (9)$$

où $c_{n,q} = \log f_q(y_n) - \log f_{q-1}(y_n)$ et $h(a) = 0$ si $a \in [0, 1]$, $+\infty$ sinon (l'opérateur proximal de h est $\text{prox}_h(a) = \max(0, \min(a, 1))$). Ainsi, l'opérateur proximal de g est simplement l'opérateur proximal de f appliqué à chaque pixel. Nous montrons comment calculer ce dernier dans la section suivante.

Finalement, le problème (8) n'impliquant que deux fonctions, dont une composée avec un opérateur linéaire, l'algorithme primal-dual de Chambolle et Pock [6] est un choix judicieux ; nous l'appliquons en section 4, avec la régularisation donnée en (4).

3 Calcul de l'opérateur proximal de f

Le résultat principal et nouveau de l'article est le suivant : l'opérateur proximal de la fonction f , définie en (1) avec h quelconque, se ramène à l'opérateur proximal pour $h = 0$, suivi de l'opérateur proximal de h . Plus précisément, pour un entier K non nul :

Problème 3.1 Soit un vecteur $r \in \mathbb{R}^K$. On veut calculer

$$\hat{z} = \arg \min_{z \in \mathbb{R}^K} \frac{1}{2} \|z - r\|_2^2 \quad \text{s.t.} \quad z_1 \geq \dots \geq z_K.$$

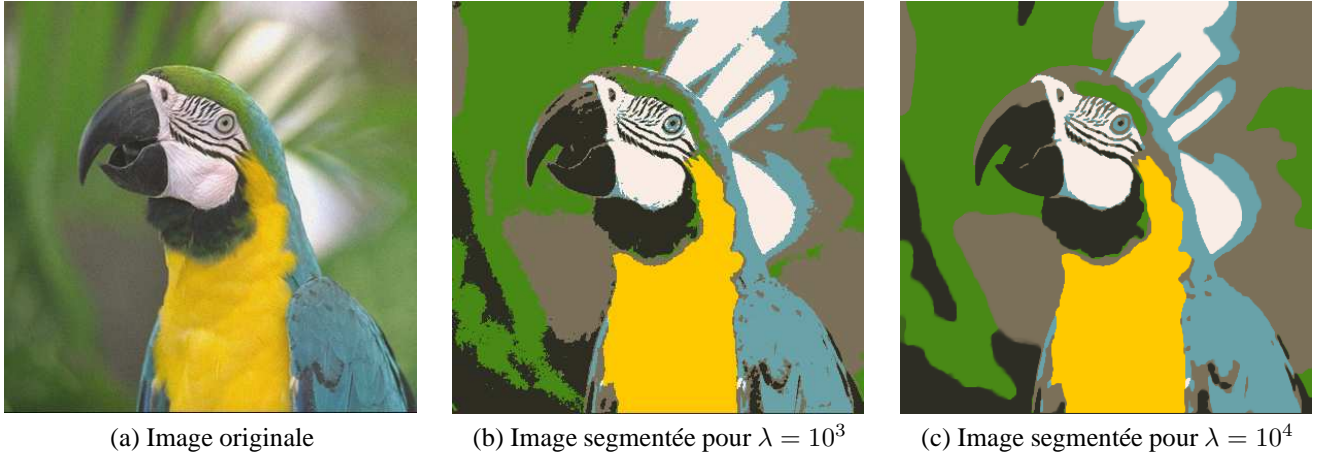


FIGURE 2 – Résultats de segmentation de l'image bruitée (a) pour $Q = 6$ classes avec l'algorithme proposé (1000 itérations).

Le vecteur \hat{z} ne possède pas de forme explicite en fonction de r mais un algorithme exact et très rapide, de complexité $O(K)$, existe pour le calculer [26, 27]. Il est rappelé en Fig. 1.

Problème 3.2 Soit h une fonction convexe semi-continue inférieurement et propre de \mathbb{R} dans $]-\infty, +\infty]$, deux vecteurs $c \in \mathbb{R}^K$ et $s \in \mathbb{R}^K$. On veut calculer

$$\hat{p} = \arg \min_{p \in \mathbb{R}^K} \frac{1}{2} \|p - s\|_2^2 + c^\top p + \sum_{k=1}^K h(p_k) \quad \text{s.t.} \quad p_1 \geq \dots \geq p_K.$$

Proposition 3.3 Les solutions \hat{p} et \hat{z} des Problèmes 3.1 et 3.2, avec $r = s - c$, vérifient :

$$\hat{p}_k = \text{prox}_h(\hat{z}), \quad \forall k = 1, \dots, K. \quad (10)$$

Preuve. \hat{z} est l'unique vecteur satisfaisant les conditions de Karush-Kuhn-Tucker d'optimalité du premier ordre, à savoir : il existe des multiplicateurs de Lagrange μ_1, \dots, μ_{K-1} tels que

1. (admissibilité primale) $\hat{z}_1 \geq \dots \geq \hat{z}_K$
2. (admissibilité duale) $\mu_k \geq 0$, pour tout $k = 1, \dots, K$,
3. (stationnarité) $\hat{z}_k - r_k + \mu_{k-1} - \mu_k = 0$, pour tout $k = 1, \dots, K$, avec $\mu_0 = \mu_K = 0$.
4. (relâchement complémentaire) $\mu_k(\hat{z}_k - \hat{z}_{k+1}) = 0$, pour tout $k = 1, \dots, K - 1$.

De manière similaire, \hat{p} est l'unique vecteur satisfaisant les conditions d'optimalité de Karush-Kuhn-Tucker, à savoir : il existe des multiplicateurs de Lagrange ν_1, \dots, ν_{K-1} et des variables duales t_1, \dots, t_K qui satisfont

1. (admissibilité primale) $\hat{p}_1 \geq \dots \geq \hat{p}_K$
2. (admissibilité duale 1) $\nu_k \geq 0$, pour tout $k = 1, \dots, K$,
3. (admissibilité duale 2) $t_k \in \partial h(\hat{p}_k)$, $\forall k = 1, \dots, K$,
4. (stationnarité) $\hat{p}_k - s_k + c_k + t_k + \nu_{k-1} - \nu_k = 0$, pour tout $k = 1, \dots, K$, avec $\nu_0 = \nu_K = 0$.
5. (relâchement complémentaire) $\nu_k(\hat{p}_k - \hat{p}_{k+1}) = 0$, pour tout $k = 1, \dots, K - 1$.

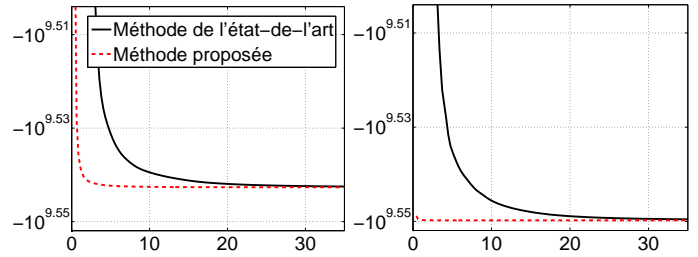


FIGURE 3 – Valeur de la fonction coût en (8) pour $\lambda = 10^3$ (à gauche) et pour $\lambda = 10^4$ (à droite) pour l'expérience illustrée en Fig. 2, en fonction du temps de calcul, en secondes.

Il est donc suffisant de montrer que \hat{p} définie par (10) satisfait les conditions ci-dessus. Tout d'abord, par définition de l'opérateur proximal, nous avons

$$0 \in \partial h(\hat{p}_k) + \hat{p}_k - \hat{z}_k, \quad \text{pour tout } k = 1, \dots, K.$$

Par conséquent, en posant $t_k = \hat{z}_k - \hat{p}_k$, la condition d'admissibilité duale 2 est satisfaite. D'autre part, en posant $\mu_k = \nu_k$, la condition d'admissibilité duale 1 et la condition de stationnarité sont également satisfaites. De plus, si $\hat{z}_k = \hat{z}_{k+1}$, alors $\hat{p}_k = \hat{p}_{k+1}$, ce qui conduit à la condition de relâchement complémentaire. Enfin, l'opérateur proximal est monotone : pour tous $a, a' \in \mathbb{R}$, $a \geq a'$ implique $\text{prox}_h(a) \geq \text{prox}_h(a')$; on a donc l'admissibilité primale, ce qui conclut la preuve. \square

4 Validation expérimentale

Nous proposons d'évaluer les performances en terme de temps de calcul entre 1) une méthode de l'état de l'art [23], qui consiste à séparer le critère en une somme de quatre fonctions, et 2) la méthode proposée, à savoir la minimisation de (8) par l'algorithme de Chambolle-Pock [6], avec l'opérateur proximal de g implémenté comme en Section 3.

L'image que nous analysons est présentée en Fig. 2 (a), il s'agit d'une version bruitée de l'image « parrot ». La segmentation est effectuée pour $Q = 6$ classes et l'image segmentée, définie en (7) est présentée en Fig. 2 (b) (resp. (c)) pour

$\lambda = 10^3$ (resp. $\lambda = 10^4$). Les temps de calcul correspondants sont présentés en Fig. 3 (implémentation Matlab avec un Mex-file en C pour l'opérateur prox_g , afin de passer outre la gestion lente des boucles en Matlab). Le gain en terme de vitesse de convergence apparaît nettement à travers ces résultats, quelle que soit la valeur de λ choisie.

Références

- [1] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Applied Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.
- [2] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [3] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, Jun. 2007.
- [4] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Visual Communication and Image Representation*, vol. 21, no. 3, pp. 193–199, Apr. 2010.
- [5] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [6] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.
- [7] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.
- [8] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *preprint arXiv:1406.5429*, to appear in *IEEE Signal Proc. Mag.*
- [9] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.
- [10] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.
- [11] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2014.
- [12] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 730–762, June 2009.
- [13] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comp. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
- [14] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Applied Math.*, vol. 42, pp. 577–685, 1989.
- [15] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Proc. IEEE Int. Conf. Comput. Vis.*, Boston, USA, 1995, pp. 694–699.
- [16] S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi, "Gradient flows and geometric active contour models," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1995, pp. 810–815.
- [17] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.
- [18] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed: A unifying graph-based optimization framework," *IEEE Trans. Pattern Anal. Match. Int.*, vol. 33, no. 7, pp. 1384–1399, Jul. 2011.
- [19] T. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM Journal of Applied Mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.
- [20] A. Chambolle, D. Cremers, and T. Pock, "A convex approach to minimal partitions," *SIAM J. Imaging Sci.*, vol. 5, no. 4, pp. 1113–1158, 2012.
- [21] T. Pock, A. Chambolle, D. Cremers, and B. Horst, "A convex relaxation approach for computing minimal partitions," in *Proc. Computer Vision and Pattern Recognition*, Miami Beach, Florida, USA, 20–25 June 2009.
- [22] S. Hiltunen, J.-C. Pesquet, and B. Pesquet-Popescu, "Comparison of two proximal splitting algorithms for solving multilabel disparity estimation problems," in *Proc. Eur. Sig. Proc. Conference*, Bucharest, Romania, Aug., 27–31 2012.
- [23] N. Pustelnik, H. Wendt, and A. Abry, "Local regularity for texture segmentation: combining wavelet leaders and proximal minimization," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 26–31 2013.
- [24] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pac. J. Optim.*, vol. 8, no. 2, pp. 273–305, Apr. 2012.
- [25] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, 2011.
- [26] L. W. Zhong and J. T. Kwok, "Efficient sparse modeling with automatic feature grouping," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 23, no. 9, pp. 1436–1447, Sep. 2012.
- [27] M. Bogdan, E. Van Den Berg, W. Su, and E. Candès, "Statistical estimation and testing via the sorted L1 norm," *Preprint, arXiv:1310.1969*, 2015.