Exact discrete minimization for TV+L0 image decomposition models

Minimisation discrète exacte pour les modèles de décomposition d'images TV+L0

Loïc Denis
Florence Tupin
Xavier Rondot

2010D007

février 2010

Département Traitement du Signal et des Images
Groupe TII : Traitement et Interprétation des Images

EXACT DISCRETE MINIMIZATION FOR TV+L0 IMAGE DECOMPOSITION MODELS

————————

MINIMISATION DISCRETE EXACTE POUR LES MODELES DE DECOMPOSITION D'IMAGES TV+L0

Loïc Denis[†], Florence Tupin[‡], Xavier Rondot[‡][1]

[†]CRAL UMR 5574 CNRS/ENS Lyon/UCBL, Lyon, France

[‡]Institut Telecom, Telecom ParisTech, CNRS LTCI, Paris, France

Abstract

Penalized maximum likelihood denoising approaches seek a solution that fulfills a compromise between data fidelity and agreement with a prior model. Penalization terms are generally chosen to enforce smoothness of the solution and to reject noise. The design of a proper penalization term is a difficult task as it has to capture image variability. Image decomposition into two components of different nature, each given a different penalty, is a way to enrich the modeling. We consider the decomposition of an image into a component with bounded variations and a sparse component. The corresponding penalization is the sum of the total variation of the first component and the L0 pseudo-norm of the second component. The minimization problem is highly non-convex, but can still be globally minimized by a minimum s-t-cut computation on a graph. The decomposition model is applied to synthetic aperture radar image denoising.

Résumé

La variation totale a été très utilisée ces dernières années dans les applications de débruitage. Elle permet de décomposer une image bruitée en une composante à variations bornées et une composante de bruit. La composante à variations bornées contient des zones homogènes séparées par des contours francs.

Il est en pratique trop restrictif de considérer les images comme étant constantes par morceaux. Les images SAR, par exemple, sont certes formées de régions homogènes, mais également de cibles ponctuelles très brillantes. La décomposition de la scène en deux composantes (une constante par morceaux, l'autre parcimonieuse) permet d'enrichir la modélisation. Cette décomposition est réalisée en calculant un estimateur MAP nécessitant la minimisation d'une énergie formée d'un terme d'attache aux données, de la variation totale et de la pseudo-norme L0. Ce problème d'optimisation est très difficile car l'énergie est non continue, non convexe et non séparable. Au lieu de considérer une minimisation alternée d'une relaxation convexe du problème initial, nous montrons qu'une minimisation discrète exacte est possible.

La méthode est illustrée sur une image SAR réelle, montrant l'apport de la décomposition par rapport à un débruitage basé uniquement sur la minimisation de la variation totale.

1 Introduction

Image denoising has been considered in the seminal work of Rudin, Osher and Fatemi (ROF) [1] as the decomposition of an image into a component with bounded variations (BV), and a noise component. The noise is removed by minimizing the total variation (TV) of the BV component, subject to a data fidelity constraint. The popularity of TV-based denoising comes from its ability to preserve sharp edges in the denoised image (i.e., in the BV component).

It has long been noted that ROF model leads to "cartoon-like" images. This can be better understood by considering ROF in a Bayesian framework. Data fidelity expresses the likelihood while total variation models the prior on the denoised image. Piecewise constant images have low total variation, and are therefore given a high prior probability. TV prior therefore biases the solution towards piecewise constant (i.e., "cartoon") images. This is often considered a problem, sometimes referred to as a staircasing effect [2].

Meyer proposed to take advantage of this effect to perform a decomposition of an image into its geometrical (so-called "structure") and textural components. To model the prior of textures as highly oscillating, he suggested the use of the G-norm [3]. Since that work, several image decomposition models have been studied in the literature [4, 5, 6, 7]. In some approaches, the focus is put on the decomposition of a noiseless image into two parts. In others, the image is decomposed into two parts plus a noise component. We follow the latter approach in this paper.

We consider the restoration of an image formed by the sum of a piecewise constant component and by some isolated pixels. Such a model can for example be applied to synthetic aperture radar (SAR) images in which the scene can often be considered as a mix of homogeneous regions and point-like strong scatterers. The restoration of the two components can be obtained by minimizing the sum of TV and L0, as discussed in section 2. The minimization problem to handle is challenging: it is combinatorial due to the L0 term. The specific form of the energy (namely, the separability of the non-convex

3

part) can be exploited to design an algorithm for exact discrete minimization. The algorithm, described in section 3, is based on a minimum s-t-cut computation on a graph. We illustrate the method on numerical experiments and SAR data in section 4.

2 The TV+L0 decomposition model

We describe the problem of the recovery a scene $\boldsymbol{u}$ from an observed image $\boldsymbol{v}$ corrupted by noise. Images are considered sampled and quantized. The scene decomposition model is introduced in the following paragraph. Then, noise modeling is described. Finally, the TV+L0 minimization problem is stated.

2.1 Scene decomposition model

We consider scenes $\boldsymbol{u}$ that can be decomposed as the sum of two terms: $\boldsymbol{u} = \boldsymbol{u}_{\mathrm{BV}} + \boldsymbol{u}_{\mathrm{S}}$, where $\boldsymbol{u}_{\mathrm{BV}}$ is a component with low total variation (i.e., close to piecewise constant), and $\boldsymbol{u}_{\mathrm{S}}$ is a sparse component (i.e., with all pixels but a few equal to zero).

The total variation of component $\boldsymbol{u}_{\mathrm{BV}}$ can be approximated[2] by $\mathrm{TV}(\boldsymbol{u}_{\mathrm{BV}}) = \sum_{(k,l)} w_{k,l} |u_{\mathrm{BV}k} - u_{\mathrm{BV}l}|$, where (k,l) denotes a clique (i.e., pair of neighboring pixels), and $w_{k,l}$ is a weight. The sparsity of component $\boldsymbol{u}_{\mathrm{S}}$ is defined by the value of the L0 pseudo-norm, i.e., the number of non-zero pixels of $\boldsymbol{u}_{\mathrm{S}}$: $\mathrm{L0}(\boldsymbol{u}_{\mathrm{S}}) = \sum_k 1 - \delta(u_{\mathrm{S}k})$.

By considering that the two components $\boldsymbol{u}_{\mathrm{BV}}$ and $\boldsymbol{u}_{\mathrm{S}}$ are statistically independent, the joint prior distribution $p(\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}})$ can be modeled in the framework of Markov random fields by:

$$
\begin{aligned}
-\log p(\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}}) &= -\log p(\boldsymbol{u}_{\mathrm{BV}}) - \log p(\boldsymbol{u}_{\mathrm{S}}) \\
&= \beta_{\mathrm{BV}} \, \mathrm{TV}(\boldsymbol{u}_{\mathrm{BV}}) + \beta_{\mathrm{S}} \, \mathrm{L0}(\boldsymbol{u}_{\mathrm{S}}).
\end{aligned}
\tag{1}
$$

---

[2]our choice of anistropic TV rather than isotropic TV will find its justification in the use of the discrete minimization technique described in section 3

2.2 Image formation model

We assume that the likelihood distribution that relates the noisy image $\boldsymbol{v}$ to the scene $\boldsymbol{u}$ is *separable*, i.e., can be written as a product of the likelihood at each pixel k:

$$p(\boldsymbol{v}|\boldsymbol{u}) = \prod_k p(v_k|u_k). \tag{2}$$

We therefore consider no mixing (such as that caused by blurring) or noise correlation in our image formation model.

If the noise is additive, the image $\boldsymbol{v}$ can be written as $\boldsymbol{v} = \boldsymbol{u}_{\mathrm{BV}} + \boldsymbol{u}_{\mathrm{S}} + \boldsymbol{n}$, with $\boldsymbol{n}$ a noise component. In the case of multiplicative noise, like speckle noise occurring in coherent imaging technique such as SAR, the image writes $\boldsymbol{v} = (\boldsymbol{u}_{\mathrm{BV}} + \boldsymbol{u}_{\mathrm{S}}) \times \boldsymbol{n}$. Note that for some noise distributions (e.g., Poisson noise, or Rice distribution), no such relation exist.

The likelihood distribution may depend on each term of the scene decomposition $p(v_k|u_{\mathrm{BV}k}, u_{\mathrm{S}k})$, as will be discussed in section 4.

2.3 Energy minimization problem

Starting from an observed noisy image $\boldsymbol{v}$, the components $\boldsymbol{u}_{\mathrm{BV}}$ and $\boldsymbol{u}_{\mathrm{S}}$ can be estimated in the maximum a posteriori sense by solving the following minimization problem:

$$\widehat{(\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}})} = \underset{(\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}})}{\arg\min} \quad -\log p(\boldsymbol{v}|\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}}) - \log p(\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}})$$

$$= \underset{(\boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}})}{\arg\min} \quad \mathrm{D}(\boldsymbol{v}, \boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}}) + \beta_{\mathrm{BV}} \, \mathrm{TV}(\boldsymbol{u}_{\mathrm{BV}}) + \beta_{\mathrm{S}} \, \mathrm{L0}(\boldsymbol{u}_{\mathrm{S}}) \tag{3}$$

Given an input image $\boldsymbol{v}$ with N pixels, minimization problem (3) has $2N$ unknowns. The objective function is non-continuous and highly non-convex in variable $\boldsymbol{u}_{\mathrm{S}}$ (possibly also in variable $\boldsymbol{u}_{\mathrm{BV}}$ depending on $\mathrm{D}(\cdot)$) and is not separable in variable $\boldsymbol{u}_{\mathrm{BV}}$ (the TV term induces a coupling between values at different sites). It is thus very challenging to solve.

5

We describe next section how the global minimum can still be found by exact discrete minimization via a combinatorial optimization technique based on a minimum s-t-cut computation.

3 Exact discrete minimization by graph-cuts

We show in the following paragraphs how to solve minimization problem (3) exactly among discrete images $\boldsymbol{u}_{\mathrm{BV}}$ and $\boldsymbol{u}_{\mathrm{S}}$. At each of their N pixels, these images take values respectively in the discrete ordered sets $\{\alpha_1, \alpha_2, \cdots, \alpha_{m_{\mathrm{BV}}}\}$ and $\{\gamma_1, \gamma_2, \cdots, \gamma_{m_{\mathrm{S}}}\}$.

3.1 Problem reformulation

We begin by considering minimization problem (3) for $\boldsymbol{u}_{\mathrm{BV}}$ fixed. We write $\boldsymbol{u}_{\mathrm{S}}^{\star}(\boldsymbol{u}_{\mathrm{BV}})$ the minimizer of one such restricted problem:

$$\boldsymbol{u}_{\mathrm{S}}^{\star}(\boldsymbol{u}_{\mathrm{BV}}) = \arg\min_{\boldsymbol{u}_{\mathrm{S}}} \quad \mathrm{D}(\boldsymbol{v}, \boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}}) + \beta_{\mathrm{S}} \, \mathrm{L0}(\boldsymbol{u}_{\mathrm{S}}) \tag{4}$$

It comes from the separability of the data fidelity and L0 terms (see our hypotheses discussed in paragraph 2.2) that the computation of $\boldsymbol{u}_{\mathrm{S}}^{\star}(\boldsymbol{u}_{\mathrm{BV}})$ requires to solve N mono-dimensional minimization problems:

$$\begin{cases} u_{\mathrm{S}k}^{\star}(u_{\mathrm{BV}k}) = u_{\mathrm{S}k}^{\dagger} & \text{if } -\log p(v_k|u_{\mathrm{BV}k}, u_{\mathrm{S}k}^{\dagger}) + \beta_{\mathrm{S}} \\ & \qquad\qquad < -\log p(v_k|u_{\mathrm{BV}k}, u_{\mathrm{S}k} = 0), \\ u_{\mathrm{S}k}^{\star}(u_{\mathrm{BV}k}) = 0 & \text{otherwise;} \end{cases} \tag{5}$$

with $u_{\mathrm{S}k}^{\dagger} = \arg\min_{u_{\mathrm{S}k}} -\log p(v_k|u_{\mathrm{BV}k}, u_{\mathrm{S}k})$ either known in closed form, or numerically computed in $O(m_{\mathrm{S}})$ operations in the worst case.

We can now reformulate minimization problem (3) into an equivalent

problem involving only $\boldsymbol{u}_{\mathrm{BV}}$:

$$\arg\min_{\boldsymbol{u}_{\mathrm{BV}}} \quad \underbrace{\mathrm{D}(\boldsymbol{v}, \boldsymbol{u}_{\mathrm{BV}}, \boldsymbol{u}_{\mathrm{S}}^{\star}(\boldsymbol{u}_{\mathrm{BV}})) + \beta_{\mathrm{S}} \, \mathrm{L0}(\boldsymbol{u}_{\mathrm{S}}^{\star}(\boldsymbol{u}_{\mathrm{BV}}))}_{(i)}$$

$$\underbrace{+\beta_{\mathrm{BV}} \, \mathrm{TV}(\boldsymbol{u}_{\mathrm{BV}})}_{(ii)} \quad (6)$$

The term (i) is of the form $\sum_k f_k(u_{\mathrm{BV}k})$ (i.e., it is separable), and (ii) is a convex term of the form $\sum_{(k,l)} g_{kl}(u_{\mathrm{BV}k}, u_{\mathrm{BV}l})$ (i.e., involving only first order cliques).

Solving problem (6) gives $\boldsymbol{u}_{\mathrm{BV}}^{\star}$. The pair $(\boldsymbol{u}_{\mathrm{BV}}^{\star}, \boldsymbol{u}_{\mathrm{S}}^{\star}(\boldsymbol{u}_{\mathrm{BV}}^{\star}))$ obtained is then a global minimizer of problem (3). We describe in next paragraph how problem (6) can be stated equivalently as a minimum s-t-cut computation on a graph.

3.2 Graph representation

Minimization problems of the form of problem (6), with a non-convex but separable discrete potential f_k and a convex discrete potential over first-order cliques g_{kl} can be exactly solved by computing a minimum s-t-cut on a graph with $N \times m_{\mathrm{BV}}$ nodes [8, 9]. We give here a quick overview of how to restate such a minimization problem as a min-cut computation.

We begin by giving a brief description of Ishikawa's graph construction [8]. As depicted in figure 1, the graph is a superimposition of layers of nodes. Each layer has N nodes, one for each image pixel, and pairs of arcs connect neighboring nodes (the cliques of the underlying pixel grid shown in Fig.1(a)). The m_{BV} different layers represent each possible level of the discrete set $\{\alpha_1, \alpha_2, \cdots, \alpha_{m_{\mathrm{BV}}}\}$. Two special nodes (terminals) are added: a source s and a sink t. Each node is also connected to its two counterparts in the next layer below and above. The first and last layers are connected respectively to the source and sink. Finally, a capacity is set to each arc, as depicted in Fig.1(c). The node corresponding to pixel k, located on the i-th layer, represents the value $u_{\mathrm{BV}k} = \alpha_i$. It is connected to its neighbors from the same layer with

7

horizontal arcs with capacity $\boxed{\text{a}} = \beta_{\text{BV}} \cdot (\alpha_{i+1} - \alpha_i)$. The vertical arc going downstream (from source to sink) is given the weight $\boxed{\text{b}} = f_k(\alpha_i)$. The vertical arc going upstream is given infinite cost $\boxed{\text{c}} = \infty$.

With this construction, cuts with finite cost are in bijection with the set $\{\alpha_1, \alpha_2, \cdots, \alpha_{m_{\text{BV}}}\}^N$ of all possible discrete images $\boldsymbol{u}_{\text{BV}}$ (see [8]). The cost of each cut equals the value of the energy minimized in (6) for the corresponding image $\boldsymbol{u}_{\text{BV}}$. Thus, the minimum cut gives the *global* minimizer $\boldsymbol{u}_{\text{BV}}^{\star}$ of problem (6), from which we deduce $\boldsymbol{u}_{\text{S}}^{\star}(\boldsymbol{u}_{\text{BV}}^{\star})$ to solve the initial image decomposition problem (3).

4   Results

We illustrate in the following paragraphs the application of the discrete minimization algorithm to decompose a SAR image into homogeneous regions and strong scatterers with the TV+L0 model.

4.1   Model for SAR images

High and very high resolution SAR scenes are generally a mix of rather homogenous regions and point-like strong scatterers. These scatterers have strong contrast (amplitudes several orders of magnitude higher than that of the homogeneous background). Man-made structures (buildings, bridges, pylons, ...) produce such strong echoes.

The observed SAR amplitude image $\boldsymbol{v}$ is corrupted by speckle noise generated by the coherent summation of echoes coming from different scatterers. In a homogeneous region, noise follows Rayleigh's distribution. When a strong scatterer is present in a resolution cell, noise distribution is better

(a) Pixel grid
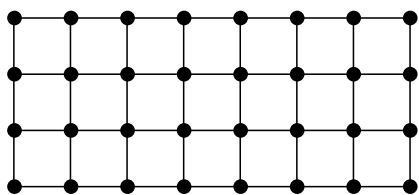
(b) Ishikawa's graph construction
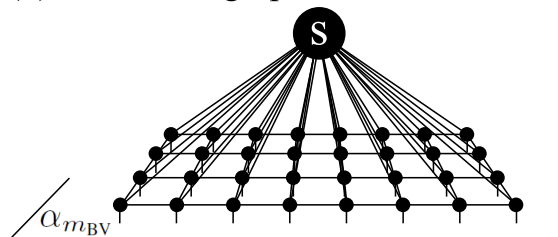
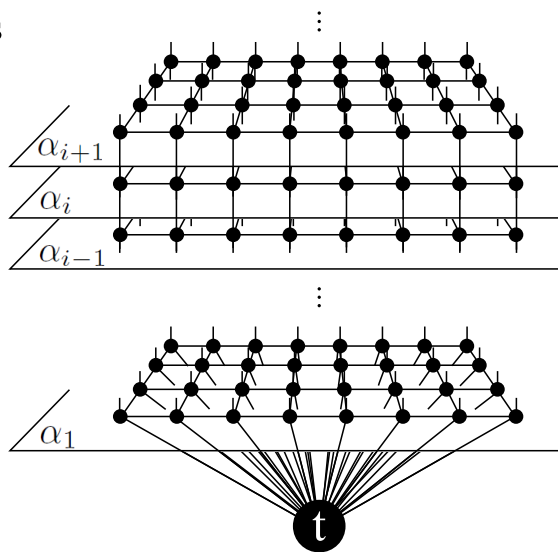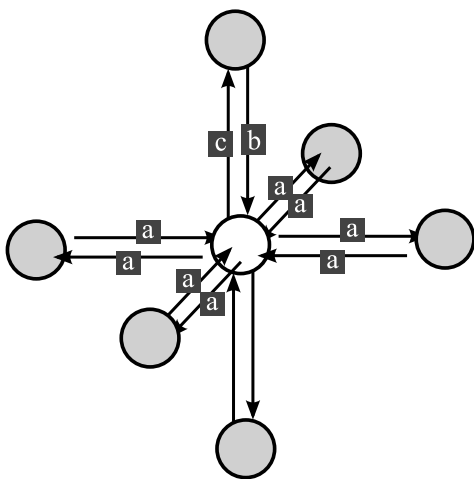(c) Arc orientations and weights

Figure 1: Graph construction for discrete minimization.

described by a Rice distribution[10] (see Fig. 2):

$$-\log p(v_k|u_{\mathrm{BV}k}, u_{\mathrm{S}k}) = \frac{v_k{}^2 + u_{\mathrm{S}k}{}^2}{2u_{\mathrm{BV}k}} + 2\log u_{\mathrm{BV}k}$$

$$-\log \mathrm{I}_0\left(\frac{v_k \cdot u_{\mathrm{S}k}}{u_{\mathrm{BV}k}{}^2}\right), \quad (7)$$

with I_0 modified Bessel's function of the first kind. Note that the amplitude $u_{\mathrm{BV}k}$ must be strictly positive and that $u_{\mathrm{S}k}$ is either equal to zero (absence of strong scatterer) or strictly positive. In the absence of a strong scatterer, equation (7) simplifies to Rayleigh's model.

4.2 Energy minimization

Rice's neg-log-likelihood defined in equation (7) is quasi-convex (but non-convex) in each variable. For a given value of $\boldsymbol{u}_{\mathrm{BV}}$, it reaches its unique minimum at $\boldsymbol{u}_{\mathrm{S}}^*(\boldsymbol{u}_{\mathrm{BV}})$ (red curve on Fig. 2). At each pixel k, and for each value $\{\alpha_1, \alpha_2, \cdots, \alpha_{m_{\mathrm{BV}}}\}$, the corresponding value of $u_{\mathrm{S}k}{}^*(u_{\mathrm{BV}k})$ can be computed efficiently by bisection. The number of quantization levels m_{S} of the sparse component can be chosen large to account for the high dynamic range of strong scatterers with very little overhead. The bounded variations component has very low dynamic range and choosing $m_{\mathrm{BV}} \approx 100$ is often enough in practice. The positivity constraints ($\forall k$, $u_{\mathrm{BV}k} > 0$ and $u_{\mathrm{S}k} \geq 0$) are straightforwardly enforced by considering discrete levels that satisfy the constraints. The tuning of parameters β_{BV} and β_{S} is done by hand by first setting a suitable β_{BV} value with $\beta_{\mathrm{S}} = 0$, and then choosing a value of β_{S} that gives a satisfying number of strong scatterers.

The main drawback of the graph-based minimization algorithm is the cost of the memory representation of the graph: $N \times m_{\mathrm{BV}}$ nodes and about $3N \times m_{\mathrm{BV}}$ pairs of arcs.
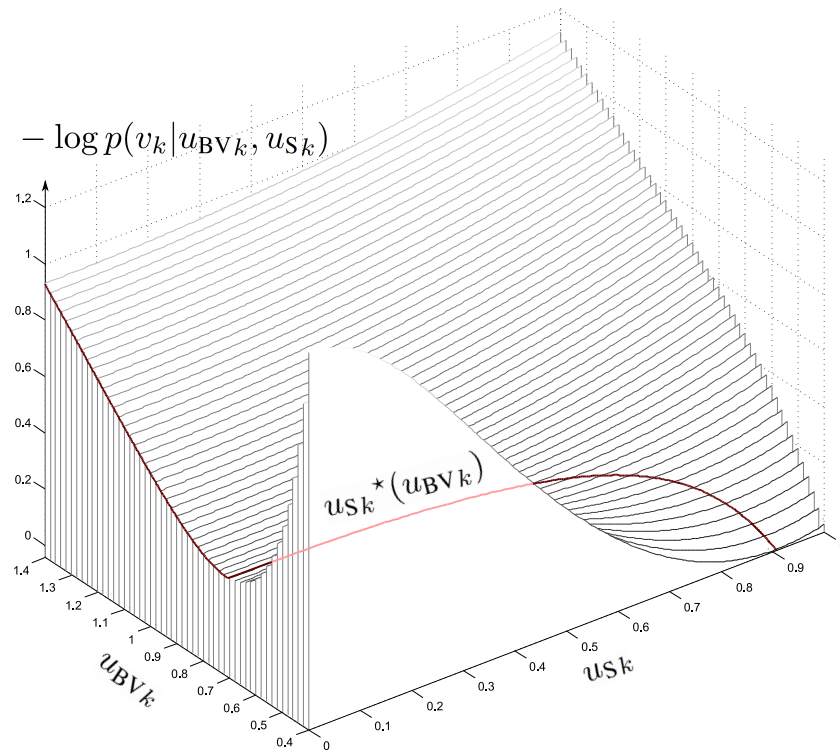
Figure 2: Surface representation of Rice's neg-log-likelihood (surface drawn for $v_k = 1$).

4.3 Decomposition results

We illustrate the method on a real SAR image (Fig. 3.a: south of Toulouse city, France, ©CNES/DGA). Fig. 3 compares the results obtained when denoising with a total variation minimization approach and with the TV+L0 decomposition approach.

The TV approach corresponds to setting β_S to infinity. Strong scatterers are either suppressed or distorted (reduced amplitude, spreading) as they do not fit to the prior model (result of TV denoising shown Fig. 3.b).

TV+L0 decomposition produces much smoother regions in the BV component (Fig. 3.c left) and sharp targets in the sparse component (Fig. 3.c right). It models better the scene and makes it possible to use a more precise noise model (namely Rice distribution). Note that the alignments of strong scatterers found in $\boldsymbol{u}_S$ component are likely to be meaningful as no spatial prior has been enforced on $\boldsymbol{u}_S$.

(a) noisy SAR amplitude image $\boldsymbol{v}$ (b) TV denoising
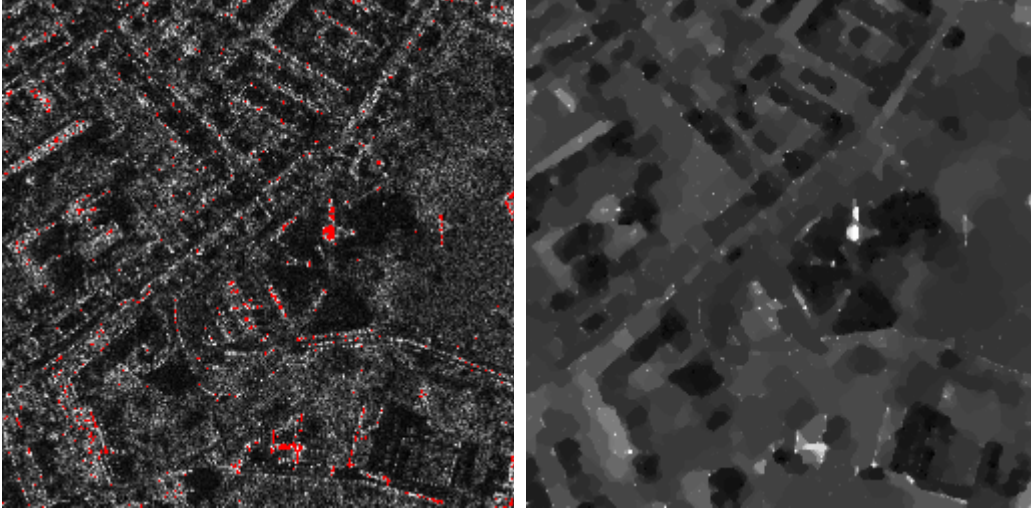
(c) image decomposition: bounded variations $\boldsymbol{u}_{\mathrm{BV}}$ and sparse $\boldsymbol{u}_{\mathrm{S}}$ parts
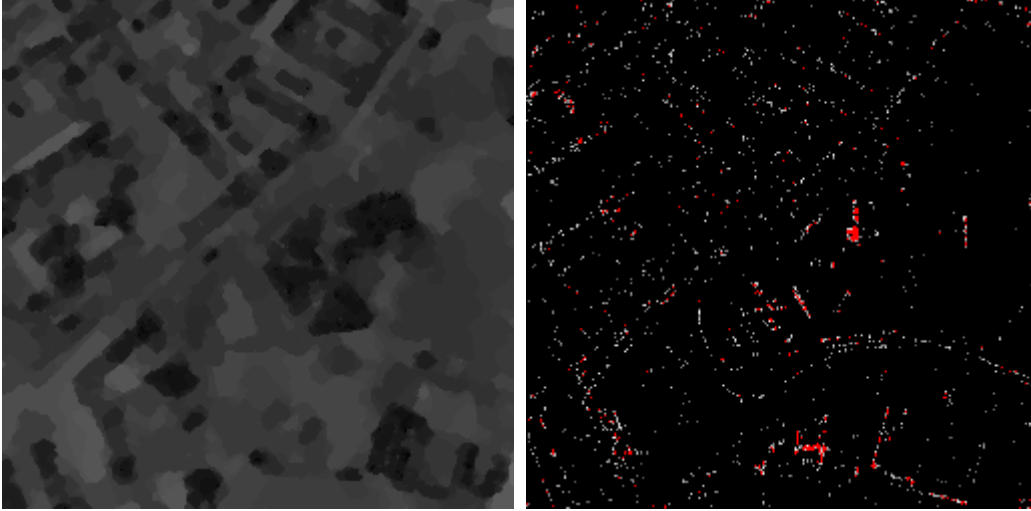
Figure 3: SAR image denoising: comparison of TV denoising and TV+L0 decomposition (SAR image ©CNES/DGA).

References

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259–268, 1992.

[2] A. Buades, B. Coll, and J. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE trans. on Image Processing*, vol. 15, no. 6, pp. 1499–1505, 2006.

[3] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, AMS, 2001.

[4] J.F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition: modeling, algorithms, and parameter selection," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 111–136, 2006.

[5] W. Yin, D. Goldfarb, and S. Osher, "Image Cartoon-Texture Decomposition and Feature Selection Using the Total Variation Regularized L^1 Functional," *Lecture Notes in Computer Science*, vol. 3752, pp. 73, 2005.

[6] M. Elad, J.L. Starck, P. Querre, and DL Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[7] J. Gilles, "Image Decomposition: Theory, Numerical Schemes, and Performance Evaluation," *Advances in Imaging and Electron Physics*, vol. 158, pp. 89–137, 2009.

[8] H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1333–1336, 2003.

[9] J. Darbon and M. Sigelle, "Image restoration with discrete constrained Total Variation part II: Levelable functions, convex priors and non-convex cases," *J. of Mathematical Imaging and Vision*, vol. 26, no. 3, pp. 277–291, 2006.

[10] J.W. Goodman, *Speckle phenomena in optics: theory and applications*, Roberts& Co Publishers, 2006.