



# Proximal Based Strategies for Solving Discrete Mumford-Shah With Ambrosio-Tortorelli Penalization on Edges

Hoang T V Le, Marion Foare, Nelly Pustelnik

## ► To cite this version:

Hoang T V Le, Marion Foare, Nelly Pustelnik. Proximal Based Strategies for Solving Discrete Mumford-Shah With Ambrosio-Tortorelli Penalization on Edges. IEEE Signal Processing Letters, 2022, 10.1109/lsp.2022.3155307 . hal-03381093

**HAL Id: hal-03381093**

**<https://hal.science/hal-03381093>**

Submitted on 15 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A PROXIMAL BASED STRATEGY FOR SOLVING DISCRETE MUMFORD-SHAH AND AMBROSIO-TORTORELLI MODELS

Hoang T.V. Le<sup>1</sup>, Marion Foare<sup>2,3</sup>, Nelly Pustelnik<sup>1\*</sup>

<sup>1</sup>Univ Lyon, Ens de Lyon, Univ Lyon 1, CNRS, Laboratoire de Physique, Lyon, 69342, France.

<sup>2</sup>Univ Lyon, Ens de Lyon, Univ Lyon 1, CNRS, INRIA, LIP, Lyon, 69342, France.

<sup>3</sup>CPE Lyon, Villeurbanne, 69100, France.

## ABSTRACT

This work is dedicated to joint image restoration and contour detection considering the Ambrosio-Tortorelli functional. Two proximal alternating minimization schemes with convergence guarantees are provided, PALM-AT and SL-PAM-AT, as well as closed-form expressions of the involved proximity operators. A thorough numerical study is conducted in order to evaluate the performance of both numerical schemes as well as comparisons to a more standard  $\ell_1$ -based discrete Mumford-Shah functional.

**Index Terms**— Image restoration, edge detection, Mumford-Shah & Ambrosio-Tortorelli functional, proximal algorithms

## 1. INTRODUCTION

The field of image reconstruction is at the core of image analysis, often being a preprocessing task before other crucial tasks such as segmentation or contour detection. Hence, image restoration is rarely the end of the processing chain, since the user usually needs to extract descriptors to analyze the resulting images, e.g. the disease course when tackle with medical image analysis and computer-aided diagnosis [1, 2, 3] or the estimation of the phase contact area to identify hydrodynamic regimes [4].

Formally, an image restoration problem relies first on a well-defined degradation model, formulated here as:

$$\mathbf{z} = A\bar{\mathbf{u}} + \mathbf{n}, \quad (1)$$

where  $\bar{\mathbf{u}} \in \mathbb{R}^N$  is the original image,  $A \in \mathbb{R}^{M \times N}$  models a linear degradation (e.g. a blur) and  $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_M)$  denotes a realization of a white Gaussian noise with standard deviation  $\sigma > 0$ .

Traditionally, the purpose of image restoration framework is to only estimate a restored image  $\hat{\mathbf{u}}$  from the degraded data  $\mathbf{z}$ . In this work, we improve it by performing a joint estimation of both the restored image  $\hat{\mathbf{u}}$  and its contours, which will be denoted by  $\hat{\mathbf{e}}$  in the following.

The first class of variational formulations aiming to estimate jointly the image and its contours traces back to the Mumford-Shah (MS) functional [5], in the continuous setting, and the Geman-Geman model [6], as a discrete counterpart. Following [7, 8], a discrete counterpart of the MS (D-MS) functional can be written as:

$$\underset{\mathbf{u} \in \mathbb{R}^N, \mathbf{e} \in \mathbb{R}^{|\mathbb{E}|}}{\text{minimize}} \quad \frac{1}{2} \|A\mathbf{u} - \mathbf{z}\|_2^2 + \beta \|(1 - \mathbf{e}) \odot D_0 \mathbf{u}\|^2 + \lambda \mathcal{R}(\mathbf{e}), \quad (2)$$

where  $D_0 \in \mathbb{R}^{|\mathbb{E}| \times N}$  models a finite difference operator and  $\mathbf{e} \in \mathbb{R}^{|\mathbb{E}|}$  denotes the edges between nodes<sup>1</sup>, and whose value is 1 when a contour change is detected, and 0 otherwise. The first term is the data fidelity term, the second one aims to enforce smoothness everywhere in the image except at the location on the contours (i.e. when  $e_i \equiv 1$  for every  $i \in \{1, \dots, |\mathbb{E}|\}$ ), while the last term  $\mathcal{R}$  penalizes the length of the contours, being the discrete counterpart of the 1D Hausdorff measure. Finally,  $\beta > 0$  and  $\lambda > 0$  denote the regularization parameters controlling the smoothness and the length of the contour respectively.

**State-of-the-art** – The MS problem being non-convex and difficult to solve, several studies provide approximations of this model. Most of these approaches are formulated in the following general two-terms minimization problem:

$$\hat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathbb{R}^N}{\text{argmin}} \quad \frac{1}{2} \|A\mathbf{u} - \mathbf{z}\|_2^2 + \mu \mathcal{P}(\mathbf{u}), \quad (3)$$

where  $\mu > 0$ . The contours  $\hat{\mathbf{e}}$  are then deduced from the reconstructed image  $\hat{\mathbf{u}}$ . For instance, the restriction to piecewise constant images leads to the ROF model in the denoising framework [9] (corresponding to  $\mathcal{P}(\mathbf{u}) = \|D_0 \mathbf{u}\|_{1,2}$ ), and to TV restoration in a more general framework [10]. Some works improve over the standard TV model, e.g. the TGV model [11], or the hybrid regularization proposed in [12], which combines the TV penalization and a smooth penalization. Such a class of convex models can efficiently be solved with convergence guarantees when using proximal algorithms framework [13, 14]. Regarding contour extraction, Cai et al. [15, 16] provide a post-processing step that consists in thresholding the ROF solution, making the bridge between ROF minimizers and the piecewise constant MS solutions.

The MS energy is also closely related to the weak membrane energy suggested by Blake and Zisserman [17], and to the Potts model [18], that are both non-convex. Storath, Weimann and coauthors [19, 20] provided several schemes based on dynamic programming to solve them. They proposed a fast and convergent algorithmic solution in 1D while the extension to 2D, based on Gauss-Seidl iterations has weaker convergence guarantees [21].

In a preliminary work [8], we explored the three-terms minimization problem (2) with various penalization terms

\*This work is supported by the ANR (Agence Nationale de la Recherche) from France ANR-19-CE48-0009 Multisc’In and Labex MILYON / ANR-10-LABX-0070.

<sup>1</sup>For instance if the set of edges are limited to the horizontal and vertical edges between two pixels, then  $|\mathbb{E}| = 2N - N_r - N_c$ , where  $N = N_r \times N_c$  is the size of the grid.

$\mathcal{R}(\mathbf{e})$  designed to be minimum when  $e_i \equiv 0$ , for every  $i \in \{1, \dots, |\mathbb{E}|\}$ . In this case, there is no need to add additional constraints on  $\mathbf{e}$ , since both the second and third terms in (2) force the solution  $\hat{\mathbf{e}}$  to be between 0 and 1. In particular, we dealt with the  $\ell_1$ -norm and the BerHu penalization, whose main interest in terms of optimization is to be separable, facilitating the design of the algorithm schemes. However, the literature dedicated to edge detection mentions that other penalization could be more appropriated than the  $\ell_1$ -norm to penalize contour length, such as the Ambrosio-Tortorelli (AT) penalization [22, 7]. It is of particular interest since it  $\Gamma$ -converges (in the continuous setting) to the (continuous) MS model [5], and then accurately approximates the Hausdorff measure originally involved in the MS model.

**Contributions** – This work focuses on a special instance of (2), when  $\mathcal{R}$  models the AT penalization over  $\mathbf{e}$  [22]. We derive two proximal alternating schemes PALM and SL-PAM in this context, leading to algorithmic schemes with convergence guarantees to a critical point of (2). A particular attention is paid to the derivation of the involved proximity operators. Numerous experiments are run in order to evaluate the performance of the proposed PALM and SL-PAM for minimizing D-MS with AT penalization over  $\mathbf{e}$ . A multiresolution Golden-grid search strategy is proposed to efficiently extract the optimal set of hyperparameters  $(\beta, \lambda)$  and to provide fair comparisons with state-of-the-art methods for different types of degradation.

**Outline** – In Section 2, we provide a discrete version of the AT penalization using Discrete Exterior Calculus operators. Section 3 is dedicated to the proposed algorithmic schemes PALM and SL-PAM and Section 4 provides numerical experiments.

## 2. AMBROSIO-TORTORELLI

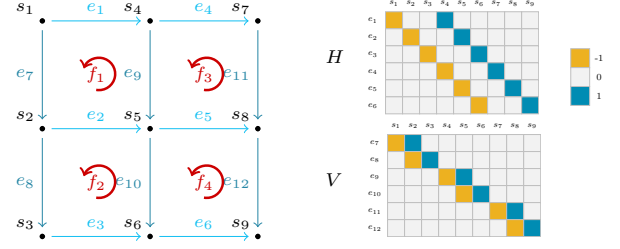
**Ambrosio-Tortorelli regularization** – In the discrete setting, the AT penalization over  $\mathbf{e}$  [22, 7] writes:

$$\mathcal{R}_{\text{AT}}(\mathbf{e}) = \frac{1}{4\varepsilon} \|\mathbf{e}\|_2^2 + \varepsilon \|D_1 \mathbf{e}\|_2^2 \quad (4)$$

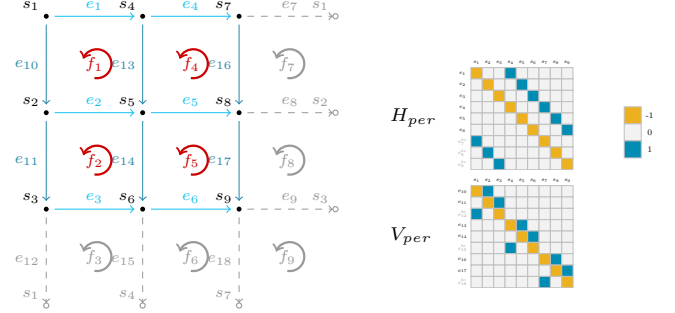
with  $\varepsilon > 0$  the  $\Gamma$ -convergence parameter. Large values of  $\varepsilon$  lead to thick contours but help to detect the set of discontinuities. Then, as  $\varepsilon$  tends to 0, the penalization of  $\|\mathbf{e}\|_2^2$  increases and enforces  $\mathbf{e}$  to become sparser, and thus contours become thinner.

This penalization term relies on the derivative operator  $D_1 \in \mathbb{R}^{|\mathbb{F}| \times |\mathbb{E}|}$ , stemming from the Discrete Exterior Calculus (DEC) framework [23]. Following [7] it allows us to better control the topology of the problem by decomposing the image domain into a 2D cell complex, composed of vertices, edges, and faces (living in space  $\mathbb{F}$ ), as illustrated on the left-hand side of Fig. 1. Following the original idea of Mumford and Shah in their seminal paper [5], the key point of this work is to set the contours  $\mathbf{e}$  to live in-between the vertices encoding the pixels value, that is, on the edges. As a consequence,  $\mathbf{e}$  can truly tend to a 1D contour.

**Design of  $D_1$**  – DEC provides differential operators  $D_k$  between  $k$ -forms. Basically, a discrete differential  $k$ -forms is defined as a column vector living on  $k$ -dimensional cells. In particular, the discrete gradient operator  $D_0 : \mathbb{R}^N \rightarrow$



**Fig. 1:** Left: example of a discrete image complex for illustrating  $D_k$  without boundary conditions. Pixel values live on the vertices  $(s_i)_i$ , while each arrow  $(e_j)_j$  encodes the contour between two pixels. Right: horizontal and vertical pixelwise discrete gradient operators  $H$  and  $V$ .



**Fig. 2:** Left: example of a discrete image complex for illustrating  $D_k$  with periodic boundary conditions. Right: corresponding horizontal and vertical pixelwise discrete gradient operators  $H_{\text{per}}$  and  $V_{\text{per}}$ .

$\mathbb{R}^{|\mathbb{E}|}$  is the oriented vertex-to-edge incidence matrix. As illustrated in Fig. 1 in a simple case where  $N_r = N_c = 3$ , it can be expressed in terms of the horizontal and vertical pixel-wise discrete gradient operators  $H$  and  $V$  as  $D_0 = [H^\top \ V^\top]^\top$ . In other words, it maps our image  $\mathbf{u}$  into a pair of  $\mathbf{e} = [\mathbf{e}_h^\top, \mathbf{e}_v^\top]^\top$ , where  $\mathbf{e}_h$  (resp.  $\mathbf{e}_v$ ) denotes horizontal (resp. vertical) edges.

Similarly,  $D_1 \in \mathbb{R}^{|\mathbb{F}| \times |\mathbb{E}|}$  is defined as the edge-to-face incidence matrix. By computing a kind of *curl* on the faces, it acts as a differential operator on edges. We can observe that it results in a combination of the vertical derivatives of horizontal edges, and conversely. As a consequence, we can simply rewrite  $D_1$  as

$$D_1 = [V \ -H] \quad (5)$$

even if [23] does not provide such a trivial form.

In order to design efficient algorithmic schemes, discrete gradient operators with boundary effects are used numerically (see Fig. 2 for the periodic counterpart of Fig. 1).

## 3. ALGORITHMIC SCHEME

The minimization of (2) when  $\mathcal{R}$  is defined by (4) appears to be a bi-convex minimization problem. In [7], the authors propose to alternate the resolution of two linear systems, derived from the optimality conditions. Mimicking the  $\Gamma$ -convergence process, they decrease  $\varepsilon$  during the optimization process. It allows them to better capture thin structures. This numerical scheme converges to a stationary point, but at the price of a huge computational time.

In this work, we propose to derive two proximal alternating schemes (following [24, 25, 8]) relying on implicit gradient (proximal) steps rather than using a standard Gauss-Seidel method, which consists in alternately minimizing the objective function w.r.t each variable.

**Proximal alternating minimization** – The first scheme, referred as PALM-AT, is presented in Algorithm 1, while

the second one, called SL-PAM-AT, is presented in Algorithm 2. The benefit of considering a full proximal step in the update of  $\mathbf{e}^{[k+1]}$  in Algorithm 2 is to relax the bound associated with the step-size parameter, computed from the Lipschitz constant of the gradient of the linearized coupling term. The convergence of the sequence  $\{(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})\}_{k \in \mathbb{N}}$  generated in Algorithm 1 (resp. Algorithm 2) to a critical point of (2) when  $\mathcal{R}$  is defined by (4) follows similar arguments than those in [25, Ass. A-B, Theorem 3.1] (resp. [8]). Among the different requested technical assumptions, one deserves a specific attention in our context. We need to have a closed form expression of the involved proximity operators, whose definition for a proper convex lower semicontinuous function  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$ , when  $\mathcal{H}$  models a real Hilbert space, is

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \text{prox}_f(\mathbf{x}) = \underset{\mathbf{y} \in \mathcal{H}}{\text{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + f(\mathbf{y}).$$

**Proximity operators closed forms** – The computation of  $\text{prox}_{\frac{1}{2c_k} \|A \cdot - \mathbf{z}\|_2^2}$  is standard [13, Table 10.1 (xx.i)]. It can be efficiently computed when  $A$  models a shift-invariant blur with periodic effects. The key issue in Algorithms 1 and 2 is the computation of  $\text{prox}_{\frac{1}{d_k} \lambda \mathcal{R}_{\text{AT}}}(\mathbf{e})$  and  $\text{prox}_{\frac{\beta}{d_k} \mathcal{S}(\cdot, \mathbf{u}^{[k+1]}) + \frac{\lambda}{d_k} \mathcal{R}_{\text{AT}}}(\mathbf{e})$  whose closed form expressions are given by the following propositions.

**Proposition 1.** *The update of the edge variable in Algorithm 1 takes a closed form that is:*

$$\mathbf{e}^{[k+1]} = \left[ \frac{2\lambda\varepsilon}{d_k} D_1^* D_1 + \left( 1 + \frac{2\lambda}{4\varepsilon d_k} \right) \text{Id} \right]^{-1} \mathbf{e}^{[k]}.$$

**Proposition 2.** *The update of the edge variable in Algorithm 2 takes a closed form that is:*

$$\mathbf{e}^{[k+1]} = \left[ \frac{2\lambda\varepsilon}{d_k} D_1^* D_1 + \frac{2\beta}{d_k} \text{diag} \left( (D_0 \mathbf{u}^{[k+1]})^2 \right) + \dots \right. \\ \left. \left( \frac{\lambda}{2\varepsilon d_k} + 1 \right) \text{Id} \right]^{-1} \left( \frac{2\beta}{d_k} (D_0 \mathbf{u}^{[k+1]})^2 + \mathbf{e}^{[k+1]} \right).$$

These results are obtained from straightforward computations relying on [13, Table 10.1 (xx.i)].

In terms of implementation, the major limitation relies on the inversions involved in Proposition 1 and in Proposition 2. When  $D_1$  is defined by (5), the update of  $\mathbf{e}^{[k+1]}$  for PALM-AT can be efficiently obtained by using inversion lemma leading to

$$\mathbf{e}^{[k+1]} = \begin{bmatrix} F + GQM & GQ \\ QM & Q \end{bmatrix} \mathbf{e}^{[k]}$$

where  $\eta_1 = (1 + \frac{\lambda}{2\varepsilon d_k})$ ,  $\eta_2 = \frac{2\lambda\varepsilon}{d_k}$ ,  $F = (\eta_2 V^* V + \eta_1 \text{Id})^{-1}$ ,  $G = F(\eta_2 V^* H)$ ,  $M = (\eta_2 H^* V)F$ , and  $Q = (\eta_2 H^* H + \eta_1 \text{Id} - (\eta_2 H^* V)F(\eta_2 V^* H))^{-1}$  where inversion is a cheap operation in the Fourier domain. However, similar inversion properties cannot be used for SL-PAM, requiring the inversion of  $\frac{2\lambda\varepsilon}{d_k} D_1^* D_1 + \frac{2\beta}{d_k} \text{diag}[(D_0 \mathbf{u}^{[k+1]})^2] + (\frac{\lambda}{2\varepsilon d_k} + 1) \text{Id}$  which starts to become intractable even for medium size restoration problems as it is performed at each iteration.

---

#### Algorithm 1: PALM-AT

---

**Input:** Noisy image  $\mathbf{z}$

**Output:** Denoised image  $\mathbf{u}^*$ , contour  $\mathbf{e}^*$

**Set:**  $\mathcal{S}(\mathbf{u}, \mathbf{e}) = \|(1 - \mathbf{e}) \odot D_0 \mathbf{u}\|^2$ ,  $L_{\nabla \mathbf{u} \mathcal{S}}^{[k]}$  and  $L_{\nabla \mathbf{e} \mathcal{S}}^{[k]}$  the Lipschitz constants of  $\nabla_{\mathbf{u}} \mathcal{S}(\cdot, \mathbf{e}^{[k]})$  and  $\nabla_{\mathbf{e}} \mathcal{S}(\mathbf{u}^{[k+1]}, \cdot)$

**Initialization:**  $\mathbf{u}^{[0]} \in \mathbb{R}^N$ ,  $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$

**while**  $\frac{\Phi^{[k+1]} - \Phi^{[k]}}{\Phi^{[k]}} < \zeta$  **and**  $k \in \mathbb{N}$  **do**  
    Choose  $c_k > \beta L_{\nabla \mathbf{u} \mathcal{S}}^{[k]}$   
     $\mathbf{u}^{[k+1]} = \text{prox}_{\frac{1}{2c_k} \|A \cdot - \mathbf{z}\|_2^2} \left( \mathbf{u}^{[k]} - \frac{\beta}{c_k} \nabla_{\mathbf{u}} \mathcal{S}(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}) \right)$   
    Choose  $d_k > \beta L_{\nabla \mathbf{e} \mathcal{S}}^{[k]}$   
     $\mathbf{e}^{[k+1]} = \text{prox}_{\frac{\lambda}{d_k} \mathcal{R}_{\text{AT}}} \left( \mathbf{e}^{[k]} - \frac{\beta}{d_k} \nabla_{\mathbf{e}} \mathcal{S}(\mathbf{e}^{[k]}, \mathbf{u}^{[k+1]}) \right)$

---



---

#### Algorithm 2: SL-PAM-AT

---

**Input:** Noisy image  $\mathbf{z}$

**Output:** Denoised image  $\mathbf{u}^*$ , contour  $\mathbf{e}^*$

**Set:**  $\mathcal{S}(\mathbf{u}, \mathbf{e}) = \|(1 - \mathbf{e}) \odot D_0 \mathbf{u}\|^2$  and  $L_{\nabla \mathbf{u} \mathcal{S}}^{[k]}$  is the Lipschitz constant of  $\nabla_{\mathbf{u}} \mathcal{S}(\cdot, \mathbf{e}^{[k]})$

**Initialization:**  $\mathbf{u}^{[0]} \in \mathbb{R}^N$ ,  $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$

**while**  $\frac{\Phi^{[k+1]} - \Phi^{[k]}}{\Phi^{[k]}} < \zeta$  **and**  $k \in \mathbb{N}$  **do**  
    Choose  $c_k > \beta L_{\nabla \mathbf{u} \mathcal{S}}^{[k]}$   
     $\mathbf{u}^{[k+1]} = \text{prox}_{\frac{1}{2c_k} \|A \cdot - \mathbf{z}\|_2^2} \left( \mathbf{u}^{[k]} - \frac{\beta}{c_k} \nabla_{\mathbf{u}} \mathcal{S}(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}) \right)$   
    Choose  $d_k > 0$   
     $\mathbf{e}^{[k+1]} = \text{prox}_{\frac{1}{d_k} (\beta \mathcal{S}(\cdot, \mathbf{u}^{[k+1]}) + \lambda \mathcal{R}_{\text{AT}})} (\mathbf{e}^{[k]})$

---

## 4. EXPERIMENTS

**Multiresolution golden search** – The selection of the hyperparameters  $(\beta, \lambda)$  impacts significantly the estimation so it is important to select them in an optimal way if we want to compare several types of regularization terms.

---


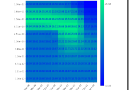
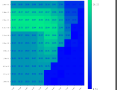
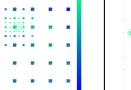
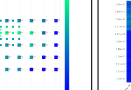
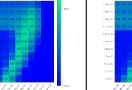
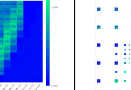
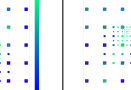
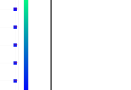


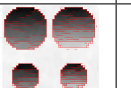

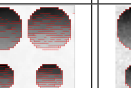




#### Algorithm 3: Multiresolution Golden-grid search

---

**for**  $\ell = 0, 1, \dots$  **do**

- 1- Run the minimizers of (2) on a  $5 \times 5$ -equally spaced grid with  $(\beta, \lambda) \in [\beta_{L,\ell}, \beta_{R,\ell}] \times [\lambda_{L,\ell}, \lambda_{R,\ell}]$ .
  - 2- Identify the pair  $(\beta_\ell^*, \lambda_\ell^*)$  maximizing the score (e.g. PSNR or Jaccard index).
  - 3- The grid bounds  $\beta_{L,\ell+1}$ ,  $\beta_{R,\ell+1}$ ,  $\lambda_{L,\ell+1}$ , and  $\lambda_{R,\ell+1}$  are updated in order to be centered in  $(\beta_\ell^*, \lambda_\ell^*)$  and with a twice smaller width.
- 

**Simulation settings** – Our experiments are performed on a toy example displayed in Table. 1 (top-left), which allow us to have access to the ground truth both in terms of image to restore and contour to extract and to study carefully the impact of the different algorithmic strategies and penalization. We consider two degradation models (1): (i) a Gaussian noise of variance  $\sigma^2$  without linear degradation (i.e.  $A = \mathbb{I}$ ) and (ii) a degradation combining a linear Gaussian blur with standard deviation of 1.1 and Gaussian noise of variance  $\sigma^2$  provided in Fig. 3. We evaluate the results both in terms of PSNR and Jaccard index, to evaluate respectively the image restoration and contour detection performances. The stopping criterion for all the algorithms is set to  $\zeta = 10^{-4}$ ,  $c_k = 1.01 * \beta * 2$ ,  $d_k = 10^{-4}$  (for SL-PAM) and  $d_k = 1.01 * \beta L_{\nabla \mathbf{e} \mathcal{S}}^{[k]}$  (for PALM) where the Lipschitz constant is computed with the power method. The  $\varepsilon$  parameter can be either selected fixed or decreasing  $\varepsilon_{\max} \searrow \varepsilon_{\min}$ .

		Grid search to maximize PSNR		Golden-grid search to maximize PSNR		Grid search to maximize Jaccard		Golden-grid search to maximize Jaccard	
Original ( $\mathbf{u}, \mathbf{e}$ )	Algorithm	PALM-AT							
		$\varepsilon = 0.2$	$\varepsilon = 2 \searrow 0.02$	$\varepsilon = 0.2$	$\varepsilon = 2 \searrow 0.02$	$\varepsilon = 0.2$	$\varepsilon = 2 \searrow 0.02$	$\varepsilon = 0.2$	$\varepsilon = 2 \searrow 0.02$
	Map								
	$\begin{bmatrix} \mathbf{u}^* \\ \mathbf{e}^* \end{bmatrix}$								
PSNR		26.83	26.31	<b>28.53</b>	<b>28.53</b>	21.20	24.93	22.62	26.13
Jaccard		0.518	0.527	0.556	0.556	0.857	0.869	0.869	<b>0.886</b>
CT per ( $\beta, \lambda$ )		1-10s per $\varepsilon$							

**Table 1:** Comparison between grid search and the proposed multiresolution Golden-grid search for  $(\beta, \lambda)$ -hyperparameters selection in the context of Gaussian noise degradation with variance  $\sigma^2 = 0.05$ . The original and degraded data are displayed on the left. The column 2-5 (resp. 6-9) present the results obtained when maximizing the PSNR (resp. Jaccard index). The map row shows the SNR or Jaccard index map for several values of  $(\beta, \lambda)$ . The grid search map is composed with 121 values while the multiresolution Golden-grid is displayed for 125 in order to provide fair comparisons. The images displayed correspond to the optimal solution obtained with different grid search approaches and algorithms. The associated PSNR and Jaccard index are also provided.

	ROF[9]	T-ROF	PALM-11	SL-PAM-11	PALM-AT $\varepsilon = 0.2$	SL-PAM-AT $\varepsilon = 0.2$	PALM-AT $\varepsilon = 2 \searrow 0.02$	SL-PAM-AT $\varepsilon = 2 \searrow 0.02$
$\begin{bmatrix} \mathbf{u}^* \\ \mathbf{e}^* \end{bmatrix}$								
Jaccard		0.530	0.869	0.876	0.869	0.883	0.886	<b>0.899</b>
PSNR	19.17	22.62	22.62	23.73	22.62	22.52	<b>26.13</b>	26.10
Min-Max CT per $(\beta, \lambda)$		1s	1-10s	1s	1-10s	60-840s	1-10s per $\varepsilon$	50-840s per $\varepsilon$

**Table 2:** Comparison between different schemes using observation  $\mathbf{z}$  provided in Table. 1 (top-left). All the results are obtained with a multiresolution golden-grid search strategy (even for T-ROF) to maximize the Jaccard index. The second row displays the optimal solution for each methods. The optimum Jaccard index, the associated PSNR and the computational time is also provided.

### Grid search versus Multiresolution Golden-grid search

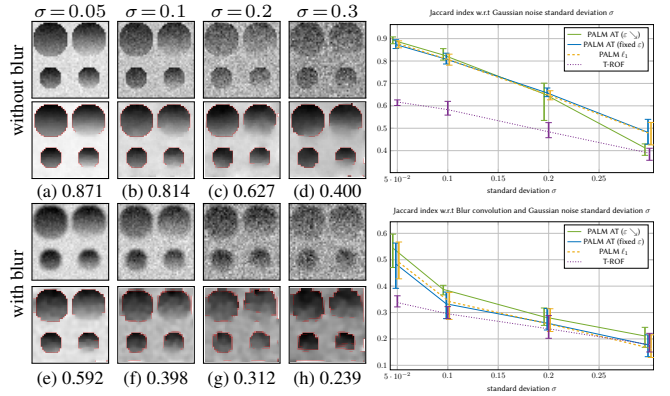
– The performances of PALM-AT for two different parameter selection strategies: grid search and the proposed multiresolution Golden-grid search (Algorithm 3) are provided in Table 1. We can observe that the proposed hyperparameter selection procedure always leads to better performances whatever the considered algorithm (more than 1.5dB in PSNR). Since the latter leads to the best parameter estimation both in terms of PSNR and Jaccard, next, we only use the Golden-grid search strategy.

**PSNR versus Jaccard** – In Table 1, we observe that maximizing the PSNR leads to a poor contour estimation while maximizing the Jaccard index performs much better. We also notice that maximizing Jaccard index, AT penalization with decreasing  $\varepsilon$  does not only give the best Jaccard index but also a good restoration related with high PSNR score compared to the others. In the following, we then focus on Jaccard index.

**PALM versus SL-PAM** – In Tab. 2, we display the optimal results in terms of Jaccard index using a multiresolution Golden-grid search strategy for PALM and SL-PAM with different choices of  $\mathcal{R}$ . The choice  $\mathcal{R} = \|\cdot\|_1$  refers to the method presented in [8] while AT penalization refers to the strategies presented in this work. We can observe that SL-PAM and PALM strategies for a similar penalization lead to close performance both quantitatively and visually, even if SL-PAM is slightly better and more precisely, among all strategies. PALM-AT and SL-PAM-AT with  $\varepsilon = 2 \searrow 0.02$  leads to the best results. However, we can notice that SL-PAM-AT requires huge computational cost due to the inversion so for more extensive results and comparisons we focus on PALM-AT.

**Impact of the choice of  $\mathcal{R}$  in PALM** – In Fig. 3, we perform comparisons between PALM-11, PALM-AT with  $\varepsilon = 0.2$ , and PALM-AT with  $\varepsilon = 2 \searrow 0.02$ . We also provide a comparison with the state-of-the-art method T-ROF. We

evaluate the performance for different degradation models, different levels of noise variance, and several realizations of them. We can observe that for the configuration without blur, all penalization lead to similar performance, while with Gaussian blur PALM-AT with  $\varepsilon = 2 \searrow 0.02$  appears to be the best choice.



**Fig. 3:** (Left) 1st & 3rd rows: observation degraded by an additive white Gaussian noise and (a)-(d) without blur, (e)-(h) with a Gaussian blur. 2nd & 4th rows: Best estimated image  $\mathbf{u}^*$  and contours  $\mathbf{e}^*$  (delineated in red) using PALM AT with  $\varepsilon = 2 \searrow 0.02$  maximizing the Jaccard index, provided below. (Right) Comparisons between T-ROF, PALM-11, PALM-AT with  $\varepsilon = 0.2$ , and PALM-AT with  $\varepsilon = 2 \searrow 0.02$  obtained with multiresolution Golden-grid search strategy maximizing the Jaccard index. For each method, we display the mean (min and max) of Jaccard index for several realizations of Gaussian noise.

## 5. CONCLUSION

In this work, we derive two iterative schemes to solve the AT functional, relying on PALM and SL-PAM. We also introduce a multiresolution Golden-grid search hyperparameters selection to compare their performances on contour detection and restoration. From numerical experiments, SL-PAM gives better results than PALM for AT regularization but at the price of larger computational time. Extensive experiments illustrate the benefit of PALM-AT with  $\varepsilon = 2 \searrow 0.02$  compared to PALM-11 and T-ROF.

## 6. REFERENCES

- [1] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of Positron Emission Tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76–96, 2014.
- [2] M. Storath, A. Weinmann, J. Friel, and M. Unser, "Joint image reconstruction and segmentation using the Potts model," *Inverse Problems*, vol. 31, no. 2, 2015.
- [3] H. Lim, Y. K. Dewaraja, and J. A. Fessler, "Joint low-count PET/CT segmentation and reconstruction with paired variational neural networks," *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312, pp. 113120, 2020.
- [4] B. Pascal, N. Pustelnik, P. Abry, J.-C. G  minard, and V. Vidal, "Parameter-free and fast nonlinear piecewise filtering: application to experimental physics," *Annals of Telecommunications*, vol. 75, pp. 655–671, 2020.
- [5] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Applied Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Match. Int.*, vol. 6, no. 6, pp. 721 – 741, Nov. 1984.
- [7] M. Foare, J.-O. Lachaud, and H. Talbot, "Image restoration and segmentation using the Ambrosio-Tortorelli functional and discrete calculus," in *Proc. 23th Int. Conf. Patt. Rec. (ICPR2016)*, Cancun, Mexico, December 4-8, 2016.
- [8] M. Foare, N. Pustelnik, and L. Condat, "Semi-linearized proximal alternating minimization for a discrete Mumford–Shah model," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 2176–2189, 2019.
- [9] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys.D.*, vol. 60, no. 1, pp. 259–268, 1992.
- [10] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restoration: Total variation, wavelet frames, and beyond," *Journal of the American Mathematical Society*, vol. 25, pp. 1033–1089, 2012.
- [11] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, 2010.
- [12] X. Cai, R. Chan, M. Nikolova, and T. Zeng, "A three-stage approach for segmenting degraded color images: Smoothing, lifting and thresholding (SLaT)," *J. Sci. Comput.*, vol. 72, no. 3, pp. 1313–1332, 2017.
- [13] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, 2011.
- [14] H. H. Bauschke, P. L. Combettes, et al., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408, Springer, 2011.
- [15] X. Cai and G. Steidl, "Multiclass segmentation by iterated ROF thresholding," in *Proc. Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lund, Sweden, 2013, Springer, pp. 237–250.
- [16] X. Cai, R. Chan, C.-B. Sch  nlieb, G. Steidl, and T. Zeng, "Linkage between piecewise constant Mumford–Shah model and Rudin–Osher–Fatemi model and its virtue in image segmentation," *SIAM J. Sci. Comput.*, vol. 41, no. 6, pp. B1310–B1340, 2019.
- [17] A. Blake and A. Zisserman, *Visual reconstruction*, MIT press, 1987.
- [18] R. Potts, "Some generalized order-disorder transformations," *Mathematical Proc. of the Cambridge Philosophical Society*, vol. 48, pp. 106–109, 1952.
- [19] A. Weinmann and M. Storath, "Iterative Potts and Blake–Zisserman minimization for the recovery of functions with discontinuities from indirect measurements," *Proc. R. Soc. A*, vol. 471, no. 2176, 2015.
- [20] L. Kiefer, M. Storath, and A. Weinmann, "Iterative Potts minimization for the recovery of signals with discontinuities from indirect measurements: the multivariate case," *Foundations of Computational Mathematics*, vol. 21, pp. 1–46, 2020.
- [21] K. Hohm, M. Storath, and A. Weinmann, "An algorithmic framework for Mumford–Shah regularization of inverse problems in imaging," *Inverse Problems*, vol. 31, no. 11, 2015.
- [22] L. Ambrosio and V. M. Tortorelli, "Approximation of functional depending on jumps by elliptic functional via t-convergence," *Comm. Pure Applied Math.*, vol. 43, no. 8, pp. 999–1036, 1990.
- [23] L. J. Grady and J. Polimeni, *Discrete Calculus: Applied analysis on graphs for computational science*, Springer, 2010.
- [24] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–  jasiewicz inequality," *Mathematics of operations research*, vol. 35, no. 2, pp. 438–457, 2010.
- [25] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.