# A new proximal method for joint image restoration and edge detection with the Mumford-Shah model

Marion Foare, Nelly Pustelnik, Laurent Condat

# A NEW PROXIMAL METHOD FOR JOINT IMAGE RESTORATION AND EDGE DETECTION WITH THE MUMFORD-SHAH MODEL

*Marion Foare[1], Nelly Pustelnik[1], Laurent Condat[2]*

[1] Univ. Lyon, ENS de Lyon, Univ. Lyon 1, CNRS,
Laboratoire de Physique, F-69342 Lyon, France
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

## ABSTRACT

In this paper, we propose an adaptation of the PAM algorithm to the minimization of a nonconvex functional designed for joint image denoising and contour detection. This new functional is based on the Ambrosio–Tortorelli approximation of the well-known Mumford–Shah functional. We motivate the proposed approximation, offering flexibility in the choice of the possibly non-smooth penalization, and we derive closed form expression for the proximal steps involved in the algorithm. We focus our attention on two types of penalization: $\ell_1$-norm and a proposed quadratic-$\ell_1$ function. Numerical experiments show that the proposed method is able to detect sharp contours and to reconstruct piecewise smooth approximations with low computational cost and convergence guarantees. We also compare the results with state-of-the-art relaxations of the Mumford–Shah functional and a recent discrete formulation of the Ambrosio–Tortorelli functional.

***Index Terms***— Segmentation, restoration, Ambrosio–Tortorelli, non-smooth optimization, proximal algorithm

## 1. INTRODUCTION

In the last decades, the Mumford–Shah (MS) model introduced in [1] was intensively studied since it allows us to perform joint image denoising and contour detection, and yields piecewise smooth results. Difficult to minimize, many relaxations of the MS functional have been proposed. The nonconvex approximation of Ambrosio and Tortorelli [2, 3] is of particular interest since it can perform both denoising and contour detection simultaneously, contrary to some convex relaxation, such as Total Variation (TV) minimization.

**Mumford–Shah and Ambrosio–Tortorelli functionals** –
Let $g \in \mathrm{L}^\infty(\Omega)$ be the grayscale, possibly degraded, input image, defined on an open bounded domain $\Omega \in \mathbb{R}^2$. In the MS model, one wants to find $u \in \mathrm{W}^{1,2}(\Omega)^1$, a piecewise smooth reconstruction of $g$, and $K \subset \Omega$ the set of discontinuities. The pair $(K, u)$ is an optimal solution of the following

---

This work is supported by Defi Imag'In SIROCCO.
[1]$\mathrm{W}^{1,2}(\Omega) = \left\{ u \in \mathrm{L}^2(\Omega) \ : \ \partial u \in \mathrm{L}^2(\Omega) \right\}$ where $\partial$ denotes the weak derivative operator.

---

functional:

$$\mathcal{MS}(K, u) = \frac{1}{2} \int_{\Omega \backslash K} |u - g|^2 \, \mathrm{d}x + \beta \int_{\Omega \backslash K} |\nabla u|^2 \, \mathrm{d}x + \lambda \mathcal{H}^1(K \cap \Omega),$$
(1)

where $\beta, \lambda > 0$, and $\mathcal{H}^1$ denotes the 1-dimensional Hausdorff measure. The image $u$ should be close to $g$ and at the same time smooth, except at the locations $K$ of the strong edges, whose total length is penalized. We can rewrite Eq. (1) as a function of $u$ [4]:

$$\mathrm{MS}(u) = \frac{1}{2} \int_{\Omega} |u - g|^2 \, \mathrm{d}x + \beta \int_{\Omega} |\nabla u|^2 \, \mathrm{d}x + \lambda \mathcal{H}^1(\mathcal{J}_u) \quad (2)$$

where $\mathcal{J}_u$ denotes the jump set of $u$ (see [5] for more details). However, the problem is not simplified because the computation of $\mathcal{H}^1(\mathcal{J}_u)$ is still required. Ambrosio and Tortorelli proposed in [2, 3] to use the notion of $\Gamma$-convergence to approach the last term by smooth functions. In the discrete setting, the Ambrosio–Tortorelli (AT) functional is defined for $(u, v) \in \mathbb{R}^N \times \mathbb{R}^M$ by

$$\mathrm{AT}_\varepsilon(u, v) = \frac{1}{2} \|u - g\|_2^2 + \beta \|\mathrm{diag}\,(1 - v)\, Du\|_2^2 + \lambda \mathcal{R}(v),$$
(3)

where

$$\mathcal{R}(v) = \left( \varepsilon \|Dv\|_2^2 + \frac{1}{4\varepsilon} \|v\|_2^2 \right),$$
(4)

$D \in \mathbb{R}^{M \times N}$ represents a discrete difference operator and $g \in \mathbb{R}^N$ is the vectorized representation of the noisy image. In this context, the contour is represented by a smooth component $v$, equals to 1 on the contour and 0 otherwise. The AT functional $\Gamma$-converges towards the MS functional (1) as $\varepsilon$ tends to 0 [2, 3]. Since the MS or AT functionals are nonconvex and difficult to optimize, because of either the length term or the presence of $\varepsilon$, many works have focused on finding suitable approximations of them.

**Related works** – One of the most popular but coarse relaxation of the MS functional is the Total Variation (TV) minimization [6, 7]. This approximation is convex, and allows us to denoise while preserving the discontinuities. However, TV minimization leads to piecewise constant results, i.e., reduces the contrast and produces staircasing effects, and it is not able to perform joint denoising and contour detection. In

[8] a finite element implementation of the AT functional was proposed, refining and realigning the meshing locally around the discontinuities but extremely costly for large images. Li et al. proposed in [9] a functional similar to AT functional, where (4) is approximated with a quadratic penalization. The convergence is proved but it does not perform contour detection. For all these methods, contour detection can be obtained by thresholding $\|Du\|^2$. Recently, [10] proposed a new formulation of AT functional in the discrete calculus framework, resulting with true 1-dimensional contours with convergence guarantees but at the price of high computational cost.

**Contributions and outline** – In this paper, we propose a new variational model and a proximal algorithm with convergence guarantees for joint image denoising and contour detection based on the Ambrosio–Tortorelli approximation of the Mumford–Shah functional. First, in Section 2 we propose a variant of the Ambrosio–Tortorelli functional independent of parameter $\varepsilon$. Then in Section 3, we describe the algorithm based on an alternating proximal minimization and give closed form expressions for the steps involving proximity operators, allowing to obtain convergence guarantees. In section 4, we display some contour detection and denoising results, compared to some state-of-the-art methods. Finally, we conclude in Section 5.

## 2. PROPOSED MODEL

In this work we focus on the following approximation of the AT problem.

**Problem 1.** *Let $\beta, \lambda > 0$, let $g \in \mathbb{R}^N$ denote the noisy image with $N$ pixels, and let $u \in \mathbb{R}^N$. We aim to solve*
$$\underset{(u,v)\in\mathbb{R}^N\times\mathbb{R}^M}{minimize} \ \mathcal{J}(u,v),$$
*with*
$$\mathcal{J}(u,v) = \frac{1}{2}\|u-g\|_2^2 + \beta\|\operatorname{diag}(1-v)\,Du\|_2^2 + \lambda\mathcal{S}(v)$$
$$(5)$$
*where $D \in \mathbb{R}^{M\times N}$ and the regularization term $\mathcal{S}$ is a, possibly non-smooth, separable function; i.e., $(\forall v = (v_i)_{1\leq i\leq M})$, $\mathcal{S}(v)=\sum_{i=1}^M \sigma_i(v_i)$, where $\sigma_i:\mathbb{R}^M\to(-\infty;+\infty]$, and whose proximity operator[2] has a closed form expression.*

The AT functional reduces to $\mathcal{S}(v) = \mathcal{R}(v)$ so that for large values of $\varepsilon$, the regularization term (4) mainly penalizes the gradient of $v$. Hence, the support of $v$ is very large, i.e., $K$ is very thick. As $\varepsilon$ tends to 0, the penalization of $\|v\|_2^2$ increases and forces $v$ to be close to 0 almost everywhere. The set of discontinuities is then progressively narrowed to the 1-dimensional contour when $\varepsilon \to 0$. However, numerically, it is not possible for $\varepsilon$ to be arbitrarily small, because $\varepsilon$ controls the thickness of the contours, and it needs to be greater than the gridstep. The main limitation of this approach is to require an $\varepsilon$ varying along the iterations.

---
[2]For every $x \in \mathbb{R}^N$, $\operatorname{prox}_{\tau f}(x) = \operatorname{argmin}_{y\in\mathbb{R}^N} \frac{1}{2}\|y-x\|_2^2 + \tau f(y)$
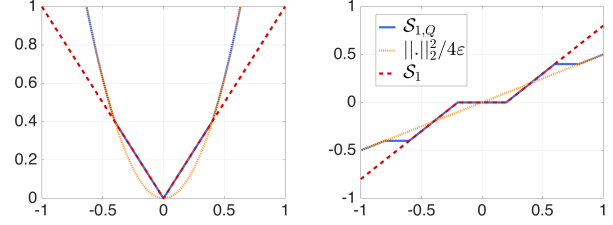


**Fig. 1**: Graphs and proximity operators of $\ell_1$ (in red), $\mathcal{S}_{1,Q}$ (in blue) and $\frac{1}{4\varepsilon}\|.\|_2^2$ (in orange). In this example, we choose $\varepsilon = 0.1$ and $\tau = 0.2$ for more readability. The choice of these parameters is discussed in Section 4. Since $v$ should take values between 0 and 1 by construction of the functional $\mathcal{J}$, we focus on the behavior in $[0, 1]$.

The model we focus on in this work is based on the observation that for small values of $\varepsilon$, the penalization term (4) on $v$ in the AT functional (3) is reduced to $\|v\|_2^2$, and forces $v$ to be close to 0 almost everywhere, except exactly on $K$. Thus, instead of minimizing the AT functional (3) depending on the parameter $\varepsilon$, one may directly consider the penalization $\frac{1}{4\varepsilon}\|v\|_2^2$. However, a simple $\ell_2$-penalization may not enforce sparsity as well as the functional (3) does. Hence, another possibility we explore in the work is to consider a sparse penalization over $v$, i.e. $\ell_0$ or $\ell_1$-penalization to approach $\mathcal{R}$ for small $\varepsilon$, and to obtain a similar (limit) behavior. Finally, in order to both reproduce the quadratic behavior of $\frac{1}{4\varepsilon}\|.\|_2^2$ for small $\varepsilon$ and enforce sparsity, one may also use a quadratic $\ell_1$-penalization for $v$ defined as follows.

**Definition 1.** *We call the quadratic $\ell_1$-penalization the convex function defined as*
$$(\forall v = (v_i)_{1\leq i\leq M} \in \mathbb{R}^M) \quad \mathcal{S}_{1,Q}(v) = \sum_{i=1}^M \max\left\{|v_i|, \frac{v_i^2}{4\varepsilon}\right\}.$$

The behavior of the proposed function is compared to the other penalization terms in Fig. 1. Note that this regularization term is different from the Huber function aiming at smoothing the $\ell_1$-norm. Here we want to keep the non-differentiability but imposing the quadratic behavior for large value of $v_i$.

In this paper, we focus on convex $\ell_1$ and quadratic-$\ell_1$ penalizations to limit the non-convexity of the problem. However, the algorithm derived in next section is adapted to separable non-convex penalization, e.g. $\ell_0$-penalization.

## 3. ALGORITHM

PAM algorithm [11] has been designed for nonsmooth and non-convex problems. Applied to Problem 1, it consists in updating alternately $u^k$ and $v^k$ using the following iterations:

$$u^{k+1} = \underset{\tilde{u}}{\operatorname{argmin}} \ \mathcal{J}(\tilde{u}, v^k) + \frac{1}{2\gamma^k}\|\tilde{u}-u^k\|_2^2 \qquad (6)$$

$$v^{k+1} = \underset{\tilde{v}}{\operatorname{argmin}} \ \mathcal{J}(u^{k+1}, \tilde{v}) + \frac{1}{2\delta^k}\|\tilde{v}-v^k\|_2^2 \qquad (7)$$

where $\gamma^k, \delta^k > 0$. The Proximal Alternating Linearized Minimization (PALM) algorithm [12] solves a linearization of iterations (6)-(7). In this work, we consider a hybrid algorithm, whose formulation for Problem 1 reads:

$$u^{k+1} = \text{prox}_{\gamma^k \|.-g\|_2^2}\left(u^k - \gamma^k \nabla_u \|\text{diag}(1-v^k)Du^k\|_2^2\right) \quad (8)$$

$$v^{k+1} = \underset{\tilde{v}}{\text{argmin}} \; \mathcal{J}(u^{k+1}, \tilde{v}) + \frac{1}{2\delta^k}\|\tilde{v} - v^k\|_2^2 \quad (9)$$

We can explicitly write these iterations as follows.

**Proposition 1.** *Let $\gamma^k > 0$. Iteration (8) is equivalent to*

$$u^{k+1} = \frac{u^k - \gamma^k \beta D^* \text{diag}\left((1-v^k)^2\right)Du^k + \gamma^k g}{\gamma^k + 1}, \quad (10)$$

*where $D^*$ denotes the adjoint of $D$ and $(1 - v^k)^2$ stands for the component-wise product.*

*Proof.* Let $\mathcal{P}(u,v) = \|\text{diag}\,(1-v)\,Du\|_2^2$ and its gradient $\nabla_u \mathcal{P}(u,v) = 2\beta D^* \text{diag}\left((1-v)^2\right)Du$. The result in (10) follows from the proximity operator of the $\ell_2$-norm. $\square$

**Proposition 2.** *Let $\delta^k > 0$. Iteration (9) is equivalent to, for all $i \in \{1, ..., M\}$,*

$$v_i^{k+1} = \text{prox}_{\frac{\lambda\sigma_i(.)}{2\beta(Du^{k+1})_i^2 + 1/\delta^k}}\left(\frac{\beta(Du^{k+1})_i^2 + \frac{v_i^k}{2\delta^k}}{\beta(Du^{k+1})_i^2 + \frac{1}{2\delta^k}}\right). \quad (11)$$

*Proof.* Let $i \in \{1, \ldots, M\}$ and $d_i = (Du^{k+1})_i^2$. We observe that

$$\underset{\tilde{v}}{\text{argmin}} \; \beta(1-\tilde{v}_i)^2 d_i + \lambda\sigma_i(\tilde{v}_i) + \frac{1}{2\delta^k}(\tilde{v}_i - v_i^k)^2$$

$$= \underset{\tilde{v}}{\text{argmin}} \; \frac{\lambda}{2\beta d_i + \frac{1}{\delta^k}}\sigma_i(\tilde{v}_i) + \frac{1}{2}\left(\tilde{v}_i - \frac{\beta d_i + \frac{v_i^k}{2\delta^k}}{\beta d_i + \frac{1}{2\delta^k}}\right)^2. \;\square$$

The resulting algorithm is displayed in Algorithm 1. Under similar assumptions than for PALM algorithm proposed by Bolte et al. [12] for $0 < \gamma^k < 1/\beta\|D\|^2$ and $0 < \delta^- \leq \delta^k \leq \delta^+ < +\infty$, the sequences $(u^k, v^k)$ converge to a (local) minimum of Problem 1.

When $\sigma_i = |\cdot|$, the associated proximity operator is the soft-thresholding, while for the quadratic $\ell_1$-penalization, the associated proximity operator is:

**Proposition 3.** *For every $\eta \in \mathbb{R}$,*

$$\text{prox}_{\tau\max\left\{|.|, \frac{|.|^2}{4\varepsilon}\right\}}(\eta) =$$
$$\text{sign}(\eta)\,\max\left\{0, \min\left[|\eta| - \tau, \max\left(4\varepsilon, \frac{|\eta|}{\frac{\tau}{2\varepsilon} + 1}\right)\right]\right\}. \quad (12)$$

*Proof.* One must remark that the maximum function in $\mathcal{S}_{1,Q}$ changes at value $x = 4\varepsilon$ and split cases. The result in (12) follows from differentiation:

$$\text{prox}_{\tau\max\left\{|.|, \frac{2}{4\varepsilon}\right\}}(\eta) =$$
$$\begin{cases} \text{sign}(\eta)\,\max(0, |\eta| - \tau) & \text{if } |\eta| < 4\varepsilon + \tau, \\ 4\varepsilon & \text{if } 4\varepsilon + \tau \leq |\eta| \leq 4\varepsilon + 2\tau, \\ \text{sign}(\eta)\,\frac{|\eta|}{\frac{\tau}{2\varepsilon} + 1} & \text{if } |\eta| > 4\varepsilon + 2\tau. \quad\square \end{cases}$$

---

**Algorithm 1:** Overview of the algorithm

**Input** $g$: vector of size $N$, $(\beta, \lambda)$: reals ;
**Var** $u^k$: vector of size $N$, $v^k$: vector of size $M$ ;
**begin**
$\quad u^k \leftarrow g, v^k \leftarrow 0;$
$\quad$**repeat**
$\qquad u^{k+1} = \dfrac{u^k - 2\gamma^k\beta D^*\text{diag}\left((1-v^k)^2\right)Du^k + 2\gamma^k g}{2\gamma^k + 1}$ ;
$\qquad v^{k+1} =$
$\qquad\qquad \left(\text{prox}_{\frac{\lambda\sigma_i(.)}{2\beta(Du^{k+1})_i^2 + 1/\delta^k}}\left(\dfrac{\beta(Du^{k+1})_i^2 + \frac{v_i^k}{2\delta^k}}{\beta(Du^{k+1})_i^2 + \frac{1}{2\delta^k}}\right)\right)_{1 \leq i \leq M}$ ;
$\quad$**until** *convergence*;
$\quad$**return** $(u, v)$ ;

---

## 4. EXPERIMENTS AND COMPARISONS

In this section, we evaluate the performance of the proposed algorithmic scheme where the regularization $\mathcal{S}$ models either a $\ell_1$-norm and a quadratic $\ell_1$-norm. We perform experiments on noisy data, where the noise is an additive white Gaussian noise with variance 0.2 for two types of images and three different sizes, in order to evaluate both the denoising and contour detection but also the performance in terms of computational time.

As discussed in Section 2, the proposed model works only for small values of $\varepsilon$ in (4). Thus, in all the experiments we arbitrarily fix $\varepsilon = 10^{-5}$. Moreover, we choose $\gamma_k = \delta_k = 0.99/\beta\|D\|^2$, so that the convergence assumptions are satisfied. In the experiments, we choose $D = [H^\top, V^\top]^\top \in \mathbb{R}^{M \times N}$, with $M = 2N$, where $H$ and $V$ are forward finite differences in the horizontal and vertical directions, respectively.

The proposed method is compared with some state-of-the-art methods: TV minimization, MS relaxation [13], and the discrete formulation of the AT functional (Discrete AT) [10]. If the proposed method and Discrete AT allows us to directly observe the contour from the estimated $v$, the two other methods require to threshold the gradient of $u$. Nonetheless, note that all these methods depend on several parameters, which can be chosen to perform different type of segmentation results. Here, for each method, we select the result leading to the best, visual-Signal to Noise Ratio-Structural Similarity (SNR-SSIM) compromise.

We can observe on Fig. 2 that the proposed method (for any choice of $\mathcal{S}$) leads to piecewise smooth results, without the typical staircasing effect obtained using TV. Moreover, they are similar to those obtained with Discrete AT and close to MS relaxation approach, which leads to the thinner contours. However, this last method does not have convergence guarantees (even local) and does not estimate the contour directly compared to both Discrete AT and the proposed method.

We compare in Table 1 the convergence speeds. Since the implementation in [13] is on GPU, we do not include
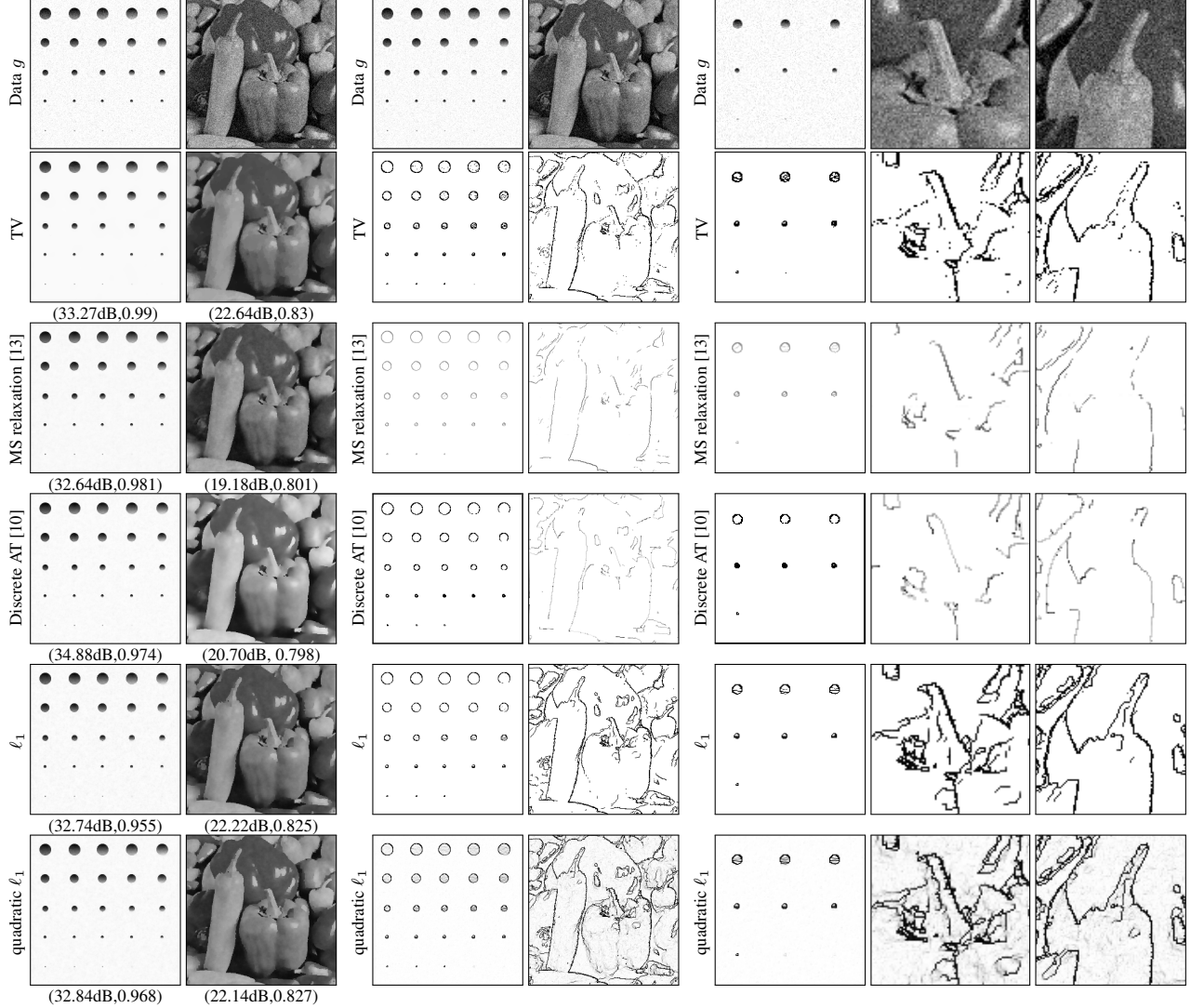
**Fig. 2**: Comparison of the proposed method with some state-of-the-art segmentation and denoising methods for two different images of sizes $128^2$ and $256^2$. Left: denoising results ($u$) with (SNR,SSIM) values, middle: contours (extracted from a threshold over $Du$ for TV, and from $v$ for the other methods), right: zoom on the contours.

this method in the comparison. The proposed model is much faster than [10] for both choices of $\mathcal{S}$, in particular with the quadratic $\ell_1$-penalization, for similar denoising and segmentation performances.

**Table 1**: Convergence speeds for images of Figure 2 with several resolutions (in sec.).

|                          | TV   | AT [10] | $\ell_1$ | quadratic $\ell_1$ |
|--------------------------|------|---------|----------|--------------------|
| dots ($N = 128^2$)       | 0.4  | 43.6    | 2.2      | 2.1                |
| dots ($N = 256^2$)       | 2.2  | 231.3   | 6.2      | 5.5                |
| dots ($N = 512^2$)       | 30.8 | 1446.5  | 116.3    | 90.3               |
| peppers ($N = 128^2$)    | 1.1  | 167.7   | 22.3     | 19.9               |
| peppers ($N = 256^2$)    | 8.8  | 1014.4  | 78.6     | 81.3               |
| peppers ($N = 512^2$)    | 61.8 | 10038.6 | 647.5    | 650.8              |

## 5. CONCLUSION AND DISCUSSION

We proposed a new functional to approximate the MS model, which gets rid of the main drawback of the Ambrosio–Tortorelli functional; namely, the presence of a theoretical parameter $\varepsilon$, for which no ideal value exists in practice. It allows to both perform image denoising and extract 1-dimensional contours while avoiding numerical difficulties due to $\varepsilon$ decreasing. Restoration and segmentation results are comparable with recent state-of-the-art methods, with a faster algorithm based on alternated proximal minimization, with proved convergence to a local minimum. For future work, it would be interesting to consider other regularizations $\mathcal{S}$, in particular nonconvex ones, and to work on real images, possibly textured images [14]. In particular, we are interested to extend this model for texture segmentation in order to apply it to multiphasic flow segmentation [15].

# 6. REFERENCES

[1] 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.

[2] 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.

[3] L. Ambrosio and V.M. Tortorelli, "On the approximation of free discontinuity problems," *Boll. Un. Mat. Ital.*, vol. 6, no. B, pp. 105–123, 1992.

[4] E. De Giorgi, M. Carriero, and A. Leaci, "Existence theorem for a minimum problem with free discontinuity set," *Archive for Rational Mechanics and Analysis*, vol. 108, no. 4, pp. 195–218, 1989.

[5] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford university press, 2000.

[6] 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.

[7] L. Condat, "Discrete total variation: New definition and minimization," *SIAM J. Imaging Sci.*, vol. 10, no. 3, pp. 1258–1290, 2017.

[8] B. Bourdin and A. Chambolle, "Implementation of an adaptive finite-element approximation of the Mumford–Shah functional," *Numerische Mathematik*, vol. 85, no. 4, pp. 609–646, 2000.

[9] Z. Li, F. Malgouyres, and T. Zeng, "Regularized nonlocal total variation and application in image restoration," *J. Math. Imag. Vis.*, vol. 59, no. 2, pp. 296–317, 2017.

[10] M. Foare, J.-O. Lachaud, and H. Talbot, "Image restoration and segmentation using the Ambrosio–Tortorelli functional and discrete calculus," in *Proc. Int. Conf. Patt. Rec.*, Cancun, Mexico, 2016.

[11] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.

[12] J. Bolte, S. Sabach, and M. Teboulle, "Proximal Alternating Linearized Minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, no. 1-2, pp. 459–494, 2014.

[13] E. Strekalovskiy and D. Cremers, "Real-time minimization of the piecewise smooth Mumford–Shah functional," in *Proc. European Conference on Computer Vision*, pp. 127–141. Springer, Zurich, Germany, 2014.

[14] N. Pustelnik, H. Wendt, P. Abry, and N. Dobigeon, "Combining local regularity estimation and total variation optimization for scale-free texture segmentation," *IEEE Trans. Computational Imaging*, vol. 2, no. 4, pp. 468–479, 2016.

[15] M. Serres, M.-L. Zanota, R. Philippe, and V. Vidal, "On the stability of taylor bubbles inside a confined highly porous medium," *International Journal of Multiphase Flow*, vol. 85, pp. 157–163, 2016.