The TV_p Regularized Mumford-Shah Model for Image Labeling and Segmentation

Yutong Li, Chunlin Wu, and Yuping Duan*

***Abstract*—The Mumford-Shah model [1] is an important tool for image labeling and segmentation, which pursues a piecewise smooth approximation of the original image and the boundaries with the shortest length. In contrast to previous efforts, which use the total variation regularization to measure the total length of the boundaries, we build up a novel piecewise smooth Mumford-Shah model by utilizing a non-convex ℓ^p regularity term for $p \in (0,1)$, which can well preserve sharp edges and eliminate geometric staircasing effects. We present optimization algorithms with convergence verification, where all subproblems can be solved by either the closed-form solution or fast Fourier transform (FFT). The method is compared to piecewise constant labeling algorithm and several state-of-the-art piecewise smooth Mumford-Shah models based on image decomposition approximations. Both labeling and segmentation results on synthetic and real images confirm the robustness and efficiency of the proposed method.**

***Index Terms*—Image labeling, segmentation, piecewise smooth, Mumford-Shah model, ℓ^p quasi-norm.**

I. Introduction

IMAGE labeling and segmentation are fundamental tasks in computer vision and pattern recognition, both of which aim to find a partition of an image into disjoint regions. Suppose that the image domain Ω is decomposed into L disjoint connected open subsets Ω_k each one with a piecewise smooth boundary and a union of boundaries of Ω_k inside Ω, such that

$$\Omega = \Omega_1 \cup \ldots \cup \Omega_L \cup \Gamma. \tag{1}$$

Mumford and Shah proposed the following minimization problem for image segmentation by penalizing the total length of the edges/boundaries of objects in an image

$$\min_{u,\Gamma} \ \frac{\lambda}{2}\int_\Omega (I-u)^2 dx + \int_{\Omega\backslash\Gamma} |\nabla u|^2 dx + \alpha|\Gamma|, \tag{2}$$

where $I : \Omega \to \mathbb{R}$ is the input image, $u \in H^1(\Omega\backslash\Gamma)$ (Sobolev space H^1 can be referred to [2]) is a nearly piecewise smooth approximation of I and $|\Gamma|$ stands for the length of 1-dimensional closed subsets $\Gamma \subset \Omega$. Since the Mumford-Shah functional is non-convex and the integral regions of the last two terms are discontinuous, finding the minimizers is not straightforward and may trap into local minima. Ambrosio and Tortorelli [3] proposed variational approximations of the Mumford-Shah functional by Γ-convergence and solved it by alternate minimizations.

Many variants have been proposed for efficient implementation of the MS model. Chan and Vese [4] sought for a binary approximation of the given image through a level set formulation [5], which has also been extended for multi-phase segmentation in [6]. Lie *et al.* [7] proposed the piecewise constant level set method to identify the boundaries and separate the image into different regions. Such kind of piecewise constant (PC) Mumford-Shah models can be formulated with respect to subregions Ω_k, $k = 1, \ldots, L$, as follows

$$\min_{\boldsymbol{c},\{\Omega_k\}_{k=1}^L} \ \frac{\lambda}{2}\sum_{k=1}^L \int_{\Omega_k} (I-c_k)^2 dx + \sum_{k=1}^L |\partial\Omega_k|. \tag{3}$$

In fact, when $\boldsymbol{c} = \{c_1, \ldots, c_L\}$ is fixed, the PC Mumford-Shah model becomes a special case of the Potts model [8] for image labeling

$$\min_{\{\Omega_k\}_{k=1}^L} \ \frac{\lambda}{2}\sum_{k=1}^L \int_{\Omega_k} f_k(x) dx + \sum_{k=1}^L |\partial\Omega_k|, \tag{4}$$

where f_k, $k = 1, \ldots, L$, are given to evaluate the performance of label assignment at each partition Ω_k. Due to the non-convex formulation of the level set methods, they may also get stuck into local minima. Chan *et al.* [9] reformulated the CV model into an equivalent convex minimization, which makes convex optimization techniques be applicable for image labeling problems. Lellmann and Schnörr [10] proposed a globally convergent Douglas-Rachford scheme for multiclass labeling problem. Bae *et al.* [11] developed a dual model for labeling problem, which proved that the global optimal of the Potts model can be achieved. Chan *et al.* [12] considered a specially designed non-convex regularization term which adapts spatially the image structures for better control of segmentation boundary. However, because such PC models approximate the image domain by a set of homogeneous regions, they fail in segmenting images with intensity inhomogeneity, that intensity varies for the same tissue within an image.

Piecewise smooth (PS) models are capable to deal with images with intensity inhomogeneous, the study of which can be traced back to almost twenty years ago [13], [14]. According to (1), the Mumford-Shah model (2) can be reformulated as

$$\min_{\{u_k,\Omega_k\}_{k=1}^L} \sum_{k=1}^L \Big(\frac{\lambda}{2}\int_{\Omega_k} (I-u_k)^2 dx + \int_{\Omega_k} |\nabla u_k|^2 dx + \alpha|\partial\Omega_k|\Big), \tag{5}$$

where $\boldsymbol{u} = \{u_1, \ldots, u_L\} : \Omega \to \mathbb{R}^L$ is the labeling function to classify each pixel $x \in \Omega$ into one out of L classes, subject to $u_k \in \{0,1\}$ and $\sum_{k=1}^L u_k = 1$. Based on the alternative

Asterisk indicates corresponding author.

Y. Li is with Center for Applied Mathematics, Tianjin University, Tianjin 300072, China.

C. Wu is with the School of Mathematical Sciences, Nankai University, Tianjin 300071, China.

*Y. Duan is with Center for Applied Mathematics, Tianjin University, Tianjin, China 300072. E-mail: yuping.duan@tju.edu.cn.

TABLE I
COMPARISON OF PIECEWISE SMOOTH MUMFORD-SHAH MODELS.

Method	Year	Decomposition	Multi-phase	Local estimation	Constant regularizer	Smooth regularizer
Tsai *et al.* [13]	2001	no	yes	no	N.A.	gradient
Vese and Chan [14]	2002	no	yes	no	TV	gradient
Le and Vese [15]	2007	additive	no	no	TV	Hessian
Li *et al.* [16]	2007	no	no	yes	TV	N.A.
Bresson *et al.* [17]	2007	no	no	no	TV	gradient
Li *et al.* [18]	2008	multiplicative	yes	yes	TV	no
Li *et al.* [19]	2010	multiplicative	yes	no	TV	gradient
Zhang *et al.* [20]	2010	no	no	yes	TV	N.A.
Li *et al.* [21]	2011	multiplicative	yes	yes	TV	no
Gu *et al.* [22]	2013	additive	yes	no	TV	gradient
Duan *et al.* [23]	2015	multiplicative	yes	yes	TV_0	gradient
Zhang *et al.* [24]	2016	multiplicative	yes	yes	TV	N.A.
Chang *et al.* [25]	2017	multiplicative	yes	no	TV_0	Hessian
Gu *et al.* [26]	2017	no	yes	yes	TV	N.A.
Jung *et al.* [27]	2017	additive	yes	no	TV	gradient
Our proposal	2020	additive	yes	no	TV_p	Laplacian

minimization and PDE techniques [26], we can obtain an equivalent minimization problem of (5) as follows

$$\min_{\{u_k,\Omega_k\}_{k=1}^L} \sum_{k=1}^{L} \Big(\frac{\lambda}{2} \iint_{\Omega_k} G(x-y)(I(x)-u_k(y))^2 dydx + |\partial\Omega_k| \Big),$$

where G is a kernel function with a localization property. The two-phase segmentation and multi-phase segmentation have been well studied in [16] and [28], respectively. Interested readers can read the references therein or [26] for more details.

Another kind of approaches assume that a piecewise smooth function can be approximated by a combination of a piecewise constant function and a smooth function. A straight-forward way is in a multiplicative intrinsic [29], [30]

$$g(x) = b(x) \sum_{k=1}^{L} c_k u_k(x) + n(x), \tag{6}$$

where $b(x) : \Omega \to \mathbb{R}$ is a spatially smooth function, and $n(x)$ denotes the random noises. Relying on (6), Li, Ng and Li [19] reformulated the Mumford-Shah model by using fuzzy membership functions to represent subregions, which reads

$$\min_{\mathbf{c},b,\{\Omega_k\}_{k=1}^L} \frac{\lambda}{2} \sum_{k=1}^{L} \int_{\Omega_k} r_k dx + \mu \int_{\Omega} |\nabla b|^2 dx + \sum_{k=1}^{L} |\partial\Omega_k|, \tag{7}$$

where $r_k = (I - bc_k)^2 + \omega(I - c_k)^2$ with $\omega > 0$ being a tuning parameter. Li *et al.* [21] modified the PC Mumford-Shah model with a local intensity clustering property to segment images with intensity inhomogeneity. Zhang *et al.* [24] rewrote the Mumford-Shah model by modeling the inhomogeneous regions as Gaussian distributions of different means and variances in a level set formulation.

The relationship between the piecewise constant function and the smooth function can be also modeled in an additive way, i.e.,

$$g(x) = \sum_{k=1}^{L} c_k u_k(x) + b(x) + n(x). \tag{8}$$

Le and Vese [15] proposed a piecewise smooth segmentation model by expressing the true intensity as a summation of a piecewise constant component and a smooth component, i.e.,

$$\min_{\mathbf{c},b,\{\Omega_k\}_{k=1}^L} \sum_{k=1}^{L} \int_{\Omega_k} f_k dx + \int_{\Omega} \mu_1 |\nabla b|^2 + \mu_2 |\nabla^2 b|^2 dx + \sum_{k=1}^{L} |\partial\Omega_k|, \tag{9}$$

where $f_k = \frac{\lambda}{2}(I - c_k - b)^2$, $\nabla^2 b$ is the Hessian of b, and μ_1, μ_2 are two positive parameters. Jung [27] used the ℓ^1 data fidelity to reformulate the energy functional for better segmentation of images with low contrast or outliers. Gu *et al.* [22] introduced a new piecewise smooth decomposition model by using a constant parameter to balance the piecewise constant and smooth component.

Until now, we focus on the discussion to guarantee a piecewise smooth solution of the Mumford-Shah model. In fact, how to realize the minimization of the total length of the boundaries is also an important perspective. Since the boundaries of the partition function u agree with the support of its gradient, a straight-forward regularity term for the total boundary length is the ℓ^0 quasi-norm of its gradient [8]. Indeed, as we summarized in Table I, most PS Mumford-Shah models are based on the total variation (TV) regularization, which is defined as the ℓ^1 norm of the image gradient magnitude. It is well-known that minimizing the ℓ^1 norm can not only encourage the sparsity in its arguments, but also preserve the edges of the images and eliminate the noises. However, the boundaries obtained by total variation

minimization are not as sharp as the ones achieved by ℓ^0 regularization. Two exceptions are [23] and [25], which utilize the ℓ^0 regularity term. In order to use the hard thresholding to solve the subproblem with respect to the ℓ^0 quasi-norm, the anisotropic TV is chosen in the modeling which leads to obvious geometric staircasing effects on the boundaries, see the numerical experiments in [23]. Such staircasing effects can be improved by isotropic approximation approaches such as [31]. On the other hand, a direct choice of the smooth regularizer is the ℓ^2 norm of its gradient [13], [14], [17], [19], [22], [23], [26]. As shown in Table I, Hessian operator has also been used to encourage the smoothness of the solution [15], [25]. Besides, several methods use the local estimation to replace the smoother regularizer such as [21], [23], [24], [26].

A. Our contributions

Recently, many studies have demonstrated that non-convex models can yield better edges and contrast preservation for image restoration and reconstruction [32], [33], [34], [35]. Especially, Nikolova [36] analyzed the minimization of anisotropic TV_p regularization can recover the edges of images and signals, and Zeng and Wu [37], [38] recently showed that the isotropic TV_p can also preserve edges. In order to make use of the power of TV_p regularity, we proposed a novel Mumford-Shah model for piecewise smooth image segmentation problems in [39]. In this work, we extend the idea both theoretically and numerically for image labeling and segmentation. More specifically, we approximate the piecewise smooth solution with the additive image model and utilize the TV_p regularization to recover the edges of the piecewise constant component. We also employ the Laplacian operator to enhance the spatial smoothness of the smooth component. An efficient algorithm with convergence verification is developed based on the operator splitting technique and the alternating direction method of multipliers, where all subproblems can be computed efficiently by either the closed-form solution or fast Fourier transform. We conduct numerous experiments to show the efficiency and effectiveness of the proposed TV_p regularized MS model. Compared with state-of-the-art Mumford-Shah models, our proposal presents better convergence speed and labeling/segmentation accuracy. Fig. 1 presents an example of piecewise smooth labeling. As shown, the PC MS model [40] misclassifies the background into three different phases, while the PS model [27] also segments the two small squares into different regions. Our proposal can locate the two squares in the image and segment the background correctly.

(a) Input (b) PC[40] (c) MSL1[27] (d) Our proposal

Fig 1. A three-phase segmentation example.

This paper is organized as follows. In Sect. II, we introduce the notations, the minimization problem of the TV_p regularized Mumford-Shah model and establish the existence of the minimizers. The numerical algorithm and solutions to the sub-minimization problems are presented in Sect. III. We also give the convergence analysis in Sect. III. The proposed model is further extended to image segmentation task in Sect. IV. We present the numerical experiments and comparisons in Sect. V. Finally, Sect. VI concludes the paper.

II. The TV_p regularized Mumford-Shah Model

A. Notations

Let $\Omega = [1, N_1] \times [1, N_2] \subset \mathbb{R}^{N_1 \times N_2}$, denote a regular image grid of $N := |\Omega|$ pixels. We define the following two inner product vector spaces:

$$X = \mathbb{R}^N \quad \text{and} \quad Y = X \times X.$$

The spaces X and Y are equipped with inner products $\langle \cdot, \cdot \rangle_X$, $\langle \cdot, \cdot \rangle_Y$ and norms $\| \cdot \|_X$ and $\| \cdot \|_Y$, respectively. In addition, we mention that, $\boldsymbol{q} \in Y$, at each pixel (i, j), $\boldsymbol{q}_{i,j} = (q^1_{i,j}, q^2_{i,j})$ and $|\boldsymbol{q}_{i,j}| = \sqrt{(q^1_{i,j})^2 + (q^2_{i,j})^2}$.

Motivated by [11], we use the *mimetic finite difference method* by building up a staggered grid scheme as shown in Fig. 2. To be specific, the variables in X are defined on $\bullet$-nodes, while the first and second component of variables in Y are defined on $\circ$-nodes and $\square$-nodes, respectively. The discrete

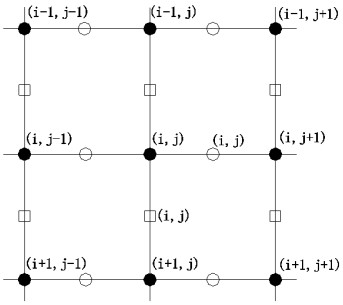


Fig 2. Definitions of the mimetic finite-dimension spaces.

gradient operator is a mapping $\nabla : X \to Y$, defined as

$$(\nabla u)_{i,j} = \big((D_x^+ u)_{i,j}, (D_y^+ u)_{i,j}\big),$$

with

$$(D_x^+ u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & 1 \le i < N_1, \\ u_{1,j} - u_{N_1,j} & i = N_1. \end{cases}$$

$$(D_y^+ u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & 1 \le j < N_2, \\ u_{i,1} - u_{i,N_2} & j = N_2. \end{cases}$$

Using the inner products of X and Y, we can find the adjoint operator of $-\nabla$, i.e., the discrete divergence operator $\text{div} : Y \to X$. Given $\boldsymbol{q} = (q^1, q^2) \in Y$, we have

$$(\text{div}\boldsymbol{q})_{i,j} = q^1_{i,j} - q^1_{i-1,j} + q^2_{i,j} - q^2_{i,j-1} = (D_x^- q^1)_{i,j} + (D_y^- q^2)_{i,j}.$$

Besides, we also introduce the notations for vector-valued variables

$$\boldsymbol{X} = \underbrace{X \times X \times \cdots \times X}_{L} \quad \text{and} \quad \boldsymbol{Y} = \underbrace{Y \times Y \times \cdots \times Y}_{L}.$$

Hence, for $\boldsymbol{u} \in \boldsymbol{X}$, its gradient is $\nabla \boldsymbol{u} = (\nabla u_1, \ldots, \nabla u_L) \in \boldsymbol{Y}$. The usual inner products and norms are also equipped in $\boldsymbol{X}$ and $\boldsymbol{Y}$ as $\langle \cdot, \cdot \rangle_{\boldsymbol{X}}$, $\langle \cdot, \cdot \rangle_{\boldsymbol{Y}}$ and norms $\| \cdot \|_{\boldsymbol{X}}$ and $\| \cdot \|_{\boldsymbol{Y}}$, respectively (see, e.g., [41]).

B. The TV_p *Mumford-Shah model for image labeling*

The non-convex ℓ^p quasi-norm with $0 < p < 1$ has been used for image restoration and reconstruction and gained great success due to its superior in exploring gradient sparsity. We introduce the TV_p regularization for multi-phase image labeling and segmentation problems by defining the isotropic TV_p regularity term for the vector valued variable $\boldsymbol{u} = (u_1, \ldots, u_L)$ as

$$\mathrm{TV}_p^{iso}(\boldsymbol{u}) := \sum_{k=1}^{L} \Bigg(\sum_{\substack{1 \le i \le N_1 \\ 1 \le j \le N_2}} \Big(\sqrt{(D_x u_k)_{i,j}^2 + (D_y u_k)_{i,j}^2} \Big)^p \Bigg), \quad (10)$$

and the anisotropic TV_p regularity term as

$$\mathrm{TV}_p^{ani}(\boldsymbol{u}) := \sum_{k=1}^{L} \Bigg(\sum_{\substack{1 \le i \le N_1 \\ 1 \le j \le N_2}} \Big(|(D_x u_k)_{i,j}|^p + |(D_y u_k)_{i,j}|^p \Big) \Bigg). \quad (11)$$

Based on the additive image decomposition model (8), we propose the novel PS Mumford-Shah model for image labeling by minimizing the ℓ^p quasi-norm of the magnitude of the gradient of $\boldsymbol{u}$, which reads

$$\min_{\boldsymbol{u} \in \mathcal{A}, b} J(\boldsymbol{u}, b) := \langle \boldsymbol{f}, \boldsymbol{u} \rangle_{\boldsymbol{X}} + \mathcal{R}(\nabla \boldsymbol{u}) + \frac{\mu}{2} \|\Delta b\|_X^2 + \frac{\eta}{2} \|b\|_X^2, \quad (12)$$

with $\boldsymbol{f} = (f_1, \ldots, f_L)$, Δb denoting the Laplacian of b, $\mathcal{R}(\nabla \boldsymbol{u})$ being either isotropic TV_p (10) or anisotropic TV_p (11), and λ, μ, η being positive parameters. In addition, the admissible set of $\boldsymbol{u}$ is relaxed by allowing for its value to take from the *probability simplex*, whose elements are nonnegative and sum up to 1, i.e.,

$$\mathcal{A} := \Big\{ \boldsymbol{u} \mid \sum_{k=1}^{L} u_k \equiv 1, \ u_k \ge 0, \ \forall \ k = 1, \ldots, L \Big\}.$$

Although the TV_p regularized MS model is non-convex and non-smooth, we can prove the existence of the minimizer as follows.

Theorem 1. *Assume* $\{c_k\}_{k=1}^L > 0$ *and* $p \in (0, 1)$. *Then, for fixed parameters* λ, μ, $\eta > 0$, *there exists a solution* $(\boldsymbol{u}^*, b^*)$, *which minimizes the proposed model* (12)*, i.e.,*

$$(\boldsymbol{u}^*, b^*) \in \arg\min_{\boldsymbol{u} \in \mathcal{A}, b} \ J(\boldsymbol{u}, b).$$

Proof. Since $\mathcal{R}(\nabla \boldsymbol{u})$ is continuous and $J(\boldsymbol{u}, b)$ is coercive, one can readily prove this theorem. □

III. Description of the ADMM Algorithm

We use the alternating direction method of multipliers (ADMM) to solve the proposed TV_p regularized MS models, which is an efficient algorithm for large-scale and multi-variable optimization problems. By introducing the new variables $\boldsymbol{v} = (v_1, \ldots, v_L)$ and $\boldsymbol{q} = (q_1, \ldots, q_L)$, we can rewrite (12) into a constrained optimization problem as follows

$$\begin{aligned} &\min_{\boldsymbol{q}, \boldsymbol{u}, \boldsymbol{v}, b} \ \langle \boldsymbol{f}, \boldsymbol{u} \rangle_{\boldsymbol{X}} + \mathcal{R}(\boldsymbol{q}) + \frac{\mu}{2} \|\Delta b\|_X^2 + \frac{\eta}{2} \|b\|_X^2, \\ &\text{s.t.,} \quad \boldsymbol{v} = \boldsymbol{u}, \quad \boldsymbol{q} = \nabla \boldsymbol{v}, \quad \boldsymbol{u} \in \mathcal{A}. \end{aligned} \quad (13)$$

Before we discuss the solution to the proposed constrained minimization (13), let us explain how to deal with the probability simplex on the variable $\boldsymbol{u}$. We define a characteristic function $\delta_{\mathcal{A}}(\boldsymbol{u})$ on $\mathcal{A}$

$$\delta_{\mathcal{A}}(\boldsymbol{u}) = \begin{cases} 0, & \text{if } \boldsymbol{u} \in \mathcal{A}, \\ +\infty, & \text{otherwise.} \end{cases} \quad (14)$$

And the optimal solution to the following minimization problem with a given $\boldsymbol{z} \in \boldsymbol{X}$,

$$\min_{\boldsymbol{u}} \ \delta_{\mathcal{A}}(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{z}\|_{\boldsymbol{X}}^2,$$

can be computed by various projection methods such as [42], [43]. Therefore, the associated augmented Lagrangian functional for (13) is defined as

$$\begin{aligned} \mathcal{L}(\boldsymbol{q}, \boldsymbol{u}, \boldsymbol{v}, b; \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2) = \ & \langle \boldsymbol{f}, \boldsymbol{u} \rangle_{\boldsymbol{X}} + \mathcal{R}(\boldsymbol{q}) + \frac{\mu}{2} \|\Delta b\|_X^2 + \frac{\eta}{2} \|b\|_X^2 \\ & - \langle \boldsymbol{\Lambda}_1, \boldsymbol{v} - \boldsymbol{u} \rangle_{\boldsymbol{X}} + \frac{1}{2r_1} \|\boldsymbol{v} - \boldsymbol{u}\|_{\boldsymbol{X}}^2 + \delta_{\mathcal{A}}(\boldsymbol{u}) \\ & - \langle \boldsymbol{\Lambda}_2, \boldsymbol{q} - \nabla \boldsymbol{v} \rangle_{\boldsymbol{Y}} + \frac{1}{2r_2} \|\boldsymbol{q} - \nabla \boldsymbol{v}\|_{\boldsymbol{Y}}^2, \end{aligned} \quad (15)$$

where r_1, r_2 are the positive constants, and $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ are the Lagrange multipliers. A natural choice to solve the above saddle-point problem is to split the variables and solve them alternatively by the ADMM algorithm as follows.

Algorithm I: ADMM for PS image labeling

1: Choose λ, μ, η, r_1, r_2 and let $b = 0$, $\boldsymbol{v}^0 = \boldsymbol{\Lambda}_1^0 = 0$, $\boldsymbol{q}^0 = \boldsymbol{\Lambda}_2^0 = 0$, $\boldsymbol{u}^0 = 1$ in some regions, $\boldsymbol{u}^0 = 0$ otherwise

2: For $n = 0, 1, \ldots,$

(i) Compute $\boldsymbol{q}^{n+1}$ from

$$\boldsymbol{q}^{n+1} = \arg\min_{\boldsymbol{q}} \ \mathcal{L}(\boldsymbol{q}, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n);$$

(ii) Compute $\boldsymbol{u}^{n+1}$ from

$$\boldsymbol{u}^{n+1} = \arg\min_{\boldsymbol{u}} \ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n);$$

(iii) Compute $\boldsymbol{v}^{n+1}$ from

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n);$$

(iv) Compute b^{n+1} from

$$b^{n+1} = \arg\min_{b} \ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n);$$

(v) Update $\boldsymbol{\Lambda}_1^{n+1}$ and $\boldsymbol{\Lambda}_2^{n+1}$ by

$$\boldsymbol{\Lambda}_1^{n+1} = \boldsymbol{\Lambda}_1^n - \frac{1}{r_1}(\boldsymbol{v}^{n+1} - \boldsymbol{u}^{n+1});$$
$$\boldsymbol{\Lambda}_2^{n+1} = \boldsymbol{\Lambda}_2^n - \frac{1}{r_2}(\boldsymbol{q}^{n+1} - \nabla \boldsymbol{v}^{n+1}).$$

3: End until some stopping rules are met.

4: When convergence is achieved, we obtain the binary solution $\boldsymbol{u}^*$ from

$$u_l^* = \begin{cases} 1, & l = \min\{\arg\max_{1 \le k \le L} u_k\}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for all } 1 \le l \le L.$$

A. Sub-minimization with respect to $\boldsymbol{q}$

Since both ℓ^p quasi-norm and ℓ^2 norm are separable, the vector-valued sub-minimization problem of $\boldsymbol{q}$ can be decomposed onto each pixel $(i,j) \in \Omega$ as follows

$$\min_{\boldsymbol{q}} \sum_{k=1}^{L} \sum_{i,j} \mathcal{R}((\boldsymbol{q}_k)_{i,j}) + \frac{1}{2r_2} \big|(\boldsymbol{q}_k)_{i,j} - ((\nabla v_k)_{i,j} + r_2(\boldsymbol{\Lambda}_{2,k})_{i,j})\big|^2. \tag{16}$$

As one can see, the above problem is decomposable and at each pixel (i,j), the problem takes the form as follows

$$\min_{\boldsymbol{q} \in \mathbb{R}^2} \ \mathcal{R}(\boldsymbol{q}) + \frac{1}{2r_2}|\boldsymbol{q} - \boldsymbol{w}|^2, \tag{17}$$

where $\boldsymbol{w} \in \mathbb{R}^2$.

For the isotropic TV_p regularity term, a geometric interpretation of the minimization is provided in Fig. 3 following [44]. The dashed circle with O as the center and $|\boldsymbol{q}|$ as the radius denotes the possible location of $\boldsymbol{q}$. For given w, the potential minimizer should locate in the same quadrant as $\boldsymbol{w}$. According to the triangle inequality of the Euclidean norm $|\cdot|$ in $\mathbb{R}^2$, we have

$$|\boldsymbol{q}| + |\boldsymbol{q} - \boldsymbol{w}| \geq |\boldsymbol{w}| = |\boldsymbol{q}^*| + |\boldsymbol{w} - \boldsymbol{q}^*|.$$

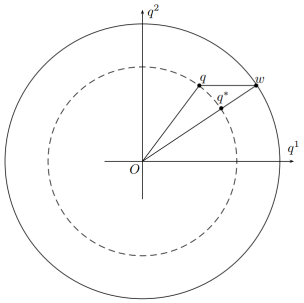


Fig 3. A geometric interpretation of the problem (17).

It is easy to deduce the solution of the problem (16) locates on the line segment Ow. Denote $\boldsymbol{q} = \beta \boldsymbol{w}$ with $0 \leq \beta \leq 1$. Thus, we can simplify (17) as the minimization of β, i.e.,

$$\min_{\beta} \ |\beta|^p + \frac{1}{2r_2}(\beta - 1)^2 |\boldsymbol{w}|^{2-p}, \quad \beta \in [0,1], \tag{18}$$

which is a proximal function of the absolute function raised to the p^{th} power with the box constraint. Such ℓ^p quasi-norm minimization problem has been well studied in [45], [46], [47], [48]. In this work, we use the generalized soft-thresholding (GST) [47] as the solver for the ℓ^p quasi-norm minimization problem, after which a one-step projection of β onto $[0,1]$ is performed.

Proposition 1. *([47]) Given $p \in (0,1)$ and $w \in \mathbb{R}$. Then, the minimizer of the following minimization problem*

$$\mathit{prox}_{r|\cdot|^p}(w) = \arg\min_{w \in \mathbb{R}} \ |q|^p + \frac{1}{2r}(q - w)^2,$$

is defined as

$$\mathit{prox}_{r|\cdot|^p}(w) = \begin{cases} 0, & \text{if } |w| \leq \tau_p(r), \\ \mathit{sgn}(w) S_p(|w|; r), & \text{if } |w| > \tau_p(r), \end{cases} \tag{19}$$

where $\tau_p(r) = (2r(1-p))^{\frac{1}{2-p}} + rp(2r(1-p))^{\frac{p-1}{2-p}}$ and $S_p(w;r)$ is the solution of the following equation

$$S_p(w;r) - w + rp\big(S_p(w;r)\big)^{p-1} = 0. \tag{20}$$

The minimizer of (20) is computed iteratively using the GST algorithm sketched as follows.

Algorithm II: The GST Algorithm

Input: w, r, p and J.

$\tau_p(r) = (2r(1-p))^{\frac{1}{2-p}} + rp(2r(1-p))^{\frac{p-1}{2-p}}$;
if $|w| \leq \tau_p(r)$
  $\mathrm{prox}_{r|\cdot|^p}(w) = 0$;
else
  $l = 0$, $x^l = |w|$;
  Iterate on $l = 0, 1, \ldots, J$
    $x^{l+1} = |w| - rp(x^l)^{p-1}$;
    $l \leftarrow l + 1$;
  $\mathrm{prox}_{r|\cdot|^p}(w) = sgn(w)x^l$;
end

For the anisotropic TV_p regularization, we actually introduce the variables $q^1 = D_x v$ and $q^2 = D_y v$ with $q^1 \in \mathbb{R}$, $q^2 \in \mathbb{R}$. Consequently, we are going to solve a pair of 1-dimensional p^{th} power minimization problem

$$\begin{cases} \min\limits_{q^1} \ |q^1|^p + \frac{1}{2r_2}\big(q^1 - (D_x v + r_2 \Lambda_2^1)\big)^2, \\ \min\limits_{q^2} \ |q^2|^p + \frac{1}{2r_2}\big(q^2 - (D_y v + r_2 \Lambda_2^2)\big)^2, \end{cases} \tag{21}$$

both of which can be efficiently solved by the GST algorithm.

B. Sub-minimization with respect to $\boldsymbol{u}$

The constrained minimization problem w.r.t. $\boldsymbol{u}$ is solved in two steps, an unconstrained minimization and a projection. That is the vector-valued minimization problem

$$\min_{\boldsymbol{u}} \ \delta_{\mathcal{A}}(\boldsymbol{u}) + \frac{1}{2r_1}\Big\|\boldsymbol{u} - (\boldsymbol{v} - r_1 \boldsymbol{\Lambda}_1 - r_1 \boldsymbol{f})\Big\|_{\boldsymbol{X}}^2, \tag{22}$$

is explicitly computed as follows

$$\boldsymbol{u} = \mathrm{proj}_{\mathcal{A}}(\boldsymbol{z}), \ \text{with} \ \ \boldsymbol{z} = \boldsymbol{v} - r_1 \boldsymbol{\Lambda}_1 - r_1 \boldsymbol{f}. \tag{23}$$

Since $\mathcal{A}$ is a closed and convex set, such projection is well defined and unique, which is realized based on sorting the elements of the variable $\boldsymbol{u}$ as described in [42].

C. Sub-minimization with respect to $\boldsymbol{v}$

We come up with the sub-minimization w.r.t. $\boldsymbol{v}$, which is also a vector-valued minimization problem and can be solved individually as

$$\min_{\boldsymbol{v}} \ \frac{1}{2r_1}\big\|\boldsymbol{v} - (\boldsymbol{u} + r_1 \boldsymbol{\Lambda}_1)\big\|_{\boldsymbol{X}}^2 + \frac{1}{2r_2}\big\|\nabla \boldsymbol{v} - \big(\boldsymbol{q} - r_2 \boldsymbol{\Lambda}_2\big)\big\|_{\boldsymbol{Y}}^2. \tag{24}$$

It is straightforward to pursue the Euler-Lagrange equation of (24) and reach at

$$(r_2 \mathcal{I} - r_1 \Delta)\boldsymbol{v} = r_2(\boldsymbol{u} + r_1 \boldsymbol{\Lambda}_1) - r_1 \mathrm{div}(\boldsymbol{q} - r_2 \boldsymbol{\Lambda}_2). \tag{25}$$

The above Poisson equation can be efficiently solved by the fast Fourier transform (FFT) with the periodic boundary condition, which gives

$$\boldsymbol{v} = \mathcal{F}^{-1}\left(\frac{r_2\mathcal{F}(\boldsymbol{u}+r_1\boldsymbol{\Lambda}_1) - r_1(\mathcal{F}(D_x^-)(g^1)+\mathcal{F}(D_y^-)\mathcal{F}(g^2))}{r_2\mathcal{I} - r_1\mathcal{F}(\Delta)}\right), \tag{26}$$

where $\boldsymbol{g} := (g^1, g^2) = (q^1 - r_2\Lambda_2^1, q^2 - r_2\Lambda_2^2)$ and $\mathcal{I}$ denotes the identity matrix.

D. Sub-minimization with respect to b

We look into the sub-minimization problem w.r.t. b, which is defined as

$$\min_b \ \frac{\lambda}{2}\big\langle (I-\boldsymbol{c}-b)^2, \boldsymbol{u}\big\rangle_{\boldsymbol{X}} + \frac{\mu}{2}\|\Delta b\|_X^2 + \frac{\eta}{2}\|b\|_X^2. \tag{27}$$

As the Laplacian operator is self adjoint, the first-order optimal condition of (27) gives a linear equation as

$$\big((\lambda+\eta)\mathcal{I} + \mu\Delta^2\big)b = \lambda\langle \boldsymbol{u}, I-\boldsymbol{c}\rangle_{\boldsymbol{X}}. \tag{28}$$

Similar to the subproblem of $\boldsymbol{v}$, periodic boundary conditions simply the efficient solution of (28) by the FFT as follows

$$b = \mathcal{F}^{-1}\left(\frac{\lambda\mathcal{F}\big(\langle \boldsymbol{u}, I-\boldsymbol{c}\rangle_{\boldsymbol{X}}\big)}{(\lambda+\eta)\mathcal{I} + \mu\mathcal{F}(\Delta)^2}\right). \tag{29}$$

E. Convergence Analysis

The convergence of ADMM for non-convex composite problems is a difficult problem, which requires the subjectiveness of the linear mapping in the composite term, or the linear mappings in the constraints satisfies an image condition [49]. Due to the non-subjectiveness of the gradient operator, we can only achieve a partial convergence result for image labeling Algorithm I.

Theorem 2. *Assume* $\boldsymbol{u}^{n+1} - \boldsymbol{u}^n \to 0$, $\boldsymbol{\Lambda}_1^{n+1} - \boldsymbol{\Lambda}_1^n \to 0$ *and* $\boldsymbol{\Lambda}_2^{n+1} - \boldsymbol{\Lambda}_2^n \to 0$ *as* $n \to \infty$ *in Algorithm I. Then any cluster point of the sequence*$\{(\boldsymbol{q}^n, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n)\}$*, if exists, is a KKT point of the constrained optimization problem* (13).

Proof. The proof uses the same ideas as the ones of Theorem 7.7 in [50], which is given in the appendix. □

IV. Extension to PS Image Segmentation

Another important application of Mumford-Shah model is the image segmentation, where $\{c_k\}_{k=1}^L$ are also unknowns. Thus, the TV$_p$ regularized Mumford-Shah model for image segmentation can be described as

$$\min_{\boldsymbol{u}\in\mathcal{A},b,\boldsymbol{c}} \big\langle \boldsymbol{f}, \boldsymbol{u}\big\rangle_{\boldsymbol{X}} + \mathcal{R}(\nabla\boldsymbol{u}) + \frac{\mu}{2}\|\Delta b\|_X^2 + \frac{\tau}{2}\|b\|_X^2, \tag{30}$$

where $\boldsymbol{c}$ is estimated explicitly as the mean values of the subregions according to

$$c_k = \frac{\int_\Omega (I-b)u_k dx}{\int_\Omega u_k dx}, \ k = 1, \ldots, L. \tag{31}$$

Simultaneously, the data term is updated via the iteration with $\{c_k\}_{k=1}^L$. The ADMM algorithm for image segmentation can be obtained straightforward by introducing the update of $\{c_k\}_{k=1}^L$ as the first step in each iteration.

V. Numerical Evaluation

In this section, we present the numerical results of the proposed TV$_p$ regularized MS (TV$_p$MS) model on various image labeling and segmentation applications. We first give the following remarks regarding to the experiments:

(1) Like most existing variational approaches, our model has parameters, including three model parameters λ, μ, η, and two algorithm parameters r_1, r_2. The model is not sensitive to the choices of η, r_1 and r_2, which are fixed as $r_1 = 10$, $r_2 = 20$, $\eta = 1e-4$ for labeling task and $r_1 = 10$, $r_2 = 200$, $\eta = 1e-4$ for segmentation task. The values of λ and μ vary for different images, which are related to the image structures, noise levels and degrees of inhomogeneity. On the other hand, the parameters of compared algorithms are all fine-tuned and provided for each example.

(2) For labeling applications, our TV$_p$MS model is terminated when either the relative dynamic error is satisfied

$$\|b^{n+1} - b^n\|_X \le \epsilon_1\|b^{n+1}\|_X,$$

with $\epsilon_1 = 2e-3$ or the maximum iteration number is reached, i.e., $Iter_{max} = 800$ for gray examples and $Iter_{max} = 400$ for color examples. For segmentation applications, both our TV$_p$MS model and MSL1 model are terminated when either the relative error is satisfied

$$\|\boldsymbol{c}^{n+1} - \boldsymbol{c}^n\|_{\boldsymbol{X}} \le \epsilon_2\|\boldsymbol{c}^{n+1}\|_{\boldsymbol{X}},$$

with $\epsilon_2 = 5e-5$ or the maximum iteration number is reached, which is the same as labeling applications.

(3) The accuracy of labeling and segmentation is measured by the Jacard Similarity (JS), defined by

$$\mathrm{JS}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|},$$

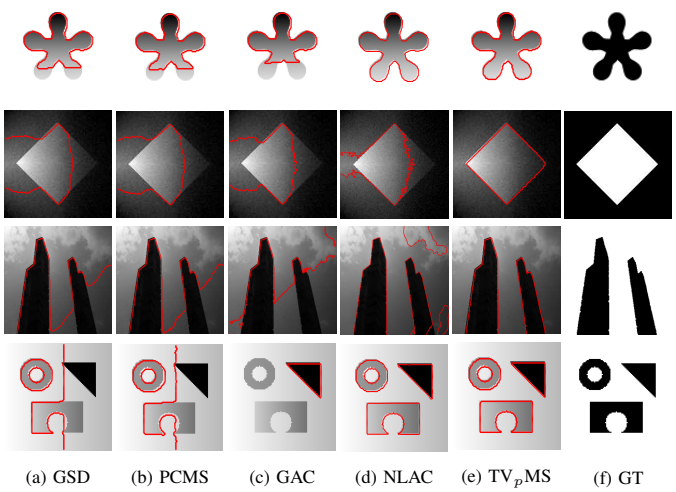
where S_1 is the region segmented by the algorithm and S_2 is the corresponding region in the ground truth.

(4) The initializations of $\boldsymbol{u}^0$ are depicted on the original images, otherwise are randomly generated.

A. Image Labeling Applications

1) Comparison with PC labeling methods: At the first place, we compare the performance of the piecewise constant models and our piecewise smooth model on four general images ('Petal', 'Square', 'Building' and 'Shapes'), including synthetic and real images of single and multiple objects. In Fig. 4, we present the labeling results of the global smoothed-dual (GSD) algorithm [11], piecewise constant MS (PCMS) algorithm [40], global active contour (GAC) [17], nonlocal active contour (NLAC) [51] and our TV$_p$MS model. As shown by both visual and quantitative results, when image intensities are smoothly varying with the location, the PC methods, i.e., GSD, PCMS and GAC algorithm all fail to classify the foreground from the background. The nonlocal model takes advantage of patch similarity, which can deal with certain intensity inhomogeneity and noises, but loses efficiency for images containing strong intensity inhomogeneity. As most real-life images, especially medical images are not piecewise

constant, the PS labeling approaches are obviously more advantageous than PC methods in practice.



(a) GSD (b) PCMS (c) GAC (d) NLAC (e) TV$_p$MS (f) GT

Images	GSD	PCMS	GAC	NLAC	TV$_p$MS
Petal	94.66	94.56	93.35	97.36	97.51
Square	52.49	52.12	52.29	65.72	94.01
Building	79.95	80.27	62.62	88.21	99.01
Shapes	55.35	54.85	82.55	99.91	99.97

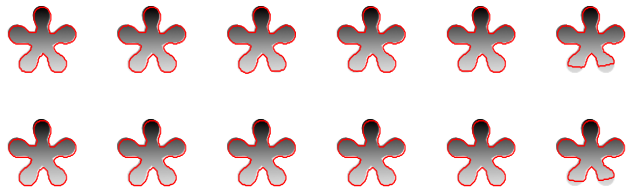
Fig 4. The labeling results of two-phase examples, where $p = 1/3$ for our TV$_p$MS model. 'GT' denotes the ground truth labeling.

In this experiment, the regularization parameter of the GSD model is set as $\lambda = 0.01, 0.05, 0.1, 0.01$ and the time step size is fixed as $\delta = 0.01, 0.05, 0.05, 0.1$ for the four images, respectively. We set the regularization parameters for the PCMS model as $\lambda = 0.005, 0.02, 0.01, 0.01$, and the GAC model as $\lambda = 1$, and $\theta = 0.1$ for image 'Petal' and 'Building', $\theta = 0.5$ for image 'Square' and 'Shapes'. For the NLAC model, we set the parameters as $\tau = 3/n$, and $(\sigma, \xi) = (10/n, 31/n)$ for image 'Petal' and 'Shapes', $(\sigma, \xi) = (10/n, 101/n)$ for image 'Square' and 'Building' with n being the width of image. The parameters of our TV$_p$MS model are selected as $\mu = 200000$ for image 'Petal', 'Building' and 'Shapes', $\mu = 250000$ for image 'Square', and $\lambda = 160, 40, 40, 60$ for the four images, respectively.

2) Sensitivity to model parameters: There are three important model parameters p, λ and μ in our TV$_p$MS model. In the following, we discuss the choices and effects of these parameters on image labeling problems.

- Performance with different p:

One important parameter in our TV$_p$ regularizer is the power index p. We compare the labeling accuracy of the isotropic TV$_p$MS model for different values of $p \in (0, 1]$ on 'Petal' image. For each p, we fine tune the parameters and use both the standard and staggered finite difference in implementation. As shown in Fig. 5, the TV$_p$MS model with the staggered grid scheme always gives better results, and higher JS values are usually achieved by smaller p. In particular, the TV$_p$MS model with $p \in (0, 1)$ gives much more accurate results than the TVMS model, i.e., the case $p = 1$, which demonstrates the advantages of the non-convex regularizer in dealing with image labeling problem. According to the observation, we set the parameter p as $p = 1/3$ in the following experiments.



(a) p=1/4 (b) p=1/3 (c) p=1/2 (d) p=2/3 (e) p=3/4 (f) p=1

	p=1/4	p=1/3	p=1/2	p=2/3	p=3/4	p=1
TV$_p^{std}$	97.53	97.51	97.40	97.47	95.48	96.23
TV$_p^{stg}$	97.56	97.51	97.55	97.53	97.53	96.47

Fig 5. Labeling results of isotropic TV$_p$MS model with standard (1st row) and staggered (2nd row) grid scheme for different p.

- Performance with different λ and μ:

Our TV$_p$MS model has two regularization parameters i.e., λ and μ. We conduct the following experiments to discuss the choices of these two parameters. Of course too large or too small λ and μ will affect the labeling results. More specifically, λ is related to image structures and noise levels. If λ is too small, the algorithm will give incomplete objects, while there will be some outliers and noise left in the results if λ is too large. On the other hand, μ controls the smoothness of the solution of b. The larger μ is, the smoother the bias field is. TABLE II records the JS values of the four images with different combinations of λ and μ, where λ and μ are chosen from $(\lambda, \mu) \in \{40, 60, \ldots, 200\} \times \{2 \times 10^5, 3 \times 10^5, \ldots, 1 \times 10^6\}$. As shown, the JS values are relatively stable for both isotropic and anisotropic TV$_p$MS model such that there are reasonably large internals for the parameters to generate good labeling results. Thus, although we need to manually tuned these two parameters for different images, we can start from moderate values of λ and μ, e.g., $\lambda = 40$ and $\mu = 200000$, to fine tune the results according to above observations.

- Robustness with initializations:

We further evaluate the performance of our model with different contour initializations. Fig. 6 displays five different initializations of $\boldsymbol{u}$ and the corresponding labeling results, where all parameters are fixed the same during the experiments. As shown, the first four initial contours are either insides, outside or across the objects, and the last contour is random generated. Despite the great difference of the initializations, the labeling results are almost the same with correct boundaries. It demonstrates that our TV$_p$MS model is quite robust with different initializations.

- Numerical convergence:

We also record the evolution of $\|b^{n+1} - b^n\|_X^2$ and $\|\boldsymbol{u}^{n+1} - \boldsymbol{u}^n\|_{\boldsymbol{X}}^2$ of the four different images via iteration number n for the isotropic TV$_p$MS model with different values of p in Fig. 7. Although the curves vibrate at the beginning, they all converge to zero as the iteration keep increasing. The smaller the value of p is, the faster the convergence is.

TABLE II
THE PERFORMANCE OF OUR TV$_p$MS MODEL W.R.T. DIFFERENT λ AND μ. THE BOLD VALUES DENOTE THE BEST OR WORST RESULTS.

Images	λ / μ	40		60		80		100		120		140		160		180		200	
		iso	ani	iso	ani	iso	ani	iso	ani	iso	ani	iso	ani	iso	ani	iso	ani	iso	ani
Petal	2×10^5	93.37	93.29	93.70	93.41	95.05	93.49	95.58	95.05	96.98	96.30	95.61	95.69	**97.51**	**97.51**	95.53	95.50	95.50	95.46
	3×10^5	93.21	93.22	93.62	93.60	94.60	93.83	93.22	93.93	94.65	97.13	96.68	97.48	97.29	97.03	97.48	97.47	95.64	95.60
	4×10^5	93.25	93.26	93.75	93.55	94.14	93.72	94.59	94.12	97.06	94.75	94.75	93.44	95.81	94.02	97.03	96.89	97.28	97.12
	5×10^5	93.30	93.22	93.69	93.44	94.08	93.91	94.23	94.15	94.38	94.17	94.82	95.35	94.56	97.06	97.04	94.33	96.92	96.49
	6×10^5	93.36	93.20	93.76	93.39	93.80	93.86	94.14	94.02	94.32	94.11	94.84	93.40	94.83	95.45	94.60	95.85	96.97	95.43
	7×10^5	93.26	93.21	93.49	93.47	94.08	93.65	94.21	94.01	94.26	94.14	94.76	94.64	94.81	94.82	94.62	94.35	97.04	97.05
	8×10^5	93.36	93.22	93.50	93.39	**93.20**	93.63	93.36	93.99	94.26	94.13	93.45	94.29	94.49	94.65	94.63	94.83	94.61	95.00
	9×10^5	93.32	93.21	93.51	93.43	93.87	93.68	93.98	93.82	94.15	94.14	94.24	94.24	94.86	94.24	94.62	94.39	94.33	94.62
	1×10^6	93.40	93.27	93.51	93.45	93.27	**93.00**	94.03	93.28	94.16	93.47	94.24	94.11	93.44	94.22	94.52	94.91	94.63	94.40
Square	2×10^5	93.95	93.80	93.98	93.98	**93.99**	93.91	93.89	93.87	93.79	93.79	93.81	93.81	93.77	93.77	93.75	93.77	93.73	93.73
	3×10^5	93.94	93.77	93.95	93.90	93.95	**93.99**	93.79	93.85	93.77	93.77	93.75	93.75	93.75	93.75	93.71	93.71	93.71	93.71
	4×10^5	93.83	93.85	93.97	93.88	93.93	93.89	93.85	93.85	93.73	93.77	93.77	93.73	93.69	93.69	93.69	93.69	93.69	93.69
	5×10^5	93.94	**93.99**	93.89	93.86	93.87	93.89	93.73	93.73	93.73	93.75	93.73	93.71	93.71	93.71	93.65	93.65	93.67	93.67
	6×10^5	93.82	93.86	93.95	93.83	93.77	93.81	93.77	93.67	93.71	93.71	93.67	93.67	93.65	93.65	93.65	93.65	93.65	93.63
	7×10^5	93.82	93.84	93.95	93.93	93.75	93.75	93.75	93.75	93.69	93.65	93.65	93.65	93.63	93.65	93.65	93.65	93.65	93.65
	8×10^5	93.74	93.80	93.85	93.81	93.83	93.81	93.77	93.75	93.69	93.69	93.65	93.65	93.65	93.65	93.63	93.65	93.61	93.63
	9×10^5	93.70	93.68	93.93	93.91	93.77	93.75	93.69	93.69	93.67	93.65	93.69	93.67	93.65	93.65	93.61	93.61	93.61	93.63
	1×10^6	93.64	93.60	93.79	93.79	93.77	93.75	93.65	93.67	93.69	93.69	93.63	93.63	93.61	93.61	93.61	93.61	**93.59**	**93.59**
Building	2×10^5	99.01	97.49	97.16	97.14	97.14	97.05	97.07	97.04	97.07	97.05	96.76	97.04	95.61	95.60	95.62	95.54	95.61	95.60
	3×10^5	97.43	**99.04**	97.11	97.08	97.07	97.07	96.82	96.92	97.03	96.83	97.01	96.82	95.69	95.47	95.69	95.69	**95.61**	**95.33**
	4×10^5	97.13	99.01	96.83	97.04	96.85	96.83	96.92	96.73	96.76	96.76	96.74	96.74	96.75	96.75	95.68	95.48	95.95	95.67
	5×10^5	97.09	98.99	96.82	96.83	96.84	96.84	96.82	96.75	96.74	96.80	96.75	96.74	96.73	96.73	96.75	96.74	95.70	96.20
	6×10^5	96.81	96.78	96.82	96.81	96.78	96.79	96.77	96.78	96.74	96.74	96.73	96.74	96.74	96.73	95.90	96.74	96.73	96.72
	7×10^5	96.79	96.77	96.80	96.79	96.79	96.75	96.76	96.74	96.74	96.73	96.74	96.73	96.75	96.72	96.73	96.73	95.89	95.88
	8×10^5	96.79	96.78	96.79	96.83	96.77	96.76	96.73	96.72	96.75	96.72	96.73	96.71	96.74	96.72	96.73	96.73	96.73	96.73
	9×10^5	**99.03**	98.99	96.78	96.83	96.75	96.75	96.75	96.74	96.74	96.70	96.72	96.71	96.72	96.73	96.73	96.73	96.72	96.73
	1×10^6	99.01	99.00	96.78	96.76	96.73	6.73	96.75	96.73	96.74	96.72	96.73	96.73	96.72	96.72	96.73	96.73	96.73	96.72
Shapes	2×10^5	99.90	99.90	99.97	99.97	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99	1	1	1	1	1	1
	3×10^5	99.92	99.92	99.97	99.97	99.99	99.99	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1
	4×10^5	99.89	99.89	99.97	99.97	99.99	99.99	99.99	99.99	**1**	**1**	1	1	1	1	1	1	1	1
	5×10^5	99.86	99.86	99.97	99.97	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1	1	1
	6×10^5	99.83	99.83	99.94	99.94	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1	1	1
	7×10^5	99.85	99.85	99.97	99.97	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1	1	1
	8×10^5	99.83	99.83	99.96	99.96	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1	1	1
	9×10^5	99.83	99.83	99.95	99.95	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1	1	1
	1×10^6	**99.81**	**99.81**	99.95	99.95	99.99	99.99	99.99	99.99	1	1	1	1	1	1	1	1	1	1

3) Extension to color image labeling: We extend the proposed model to color image labeling problem. Let us denote a color image by $\boldsymbol{I} = (I_r, I_g, I_b)$, where I_r, I_g, $I_b \in X$. The data fidelity term $\boldsymbol{f} = (f_1, \ldots, f_L)$ is then defined as

$$f_k = \frac{\lambda}{2} \sum_{i=\{r,g,b\}} |I_i - b_i - (c_k)_i|^2, \quad \text{for } k = 1, \ldots, L,$$

where $c_k \in \mathbb{R}^3$ represents a constant color value, and $\boldsymbol{b} = (b_r, b_g, b_b)$ with $b_i : \Omega \to \mathbb{R}$. Thus, the TV$_p$ regularized Mumford-Shah model (12) for color image segmentation still looks like

$$\min_{\boldsymbol{u}\in\mathcal{A},\boldsymbol{b}} \ \langle \boldsymbol{f}, \boldsymbol{u}\rangle_{\boldsymbol{X}} + \mathcal{R}(\nabla \boldsymbol{u}) + \frac{\mu}{2}\|\Delta \boldsymbol{b}\|^2_{X\times X\times X} + \frac{\eta}{2}\|\boldsymbol{b}\|^2_{X\times X\times X}, \tag{32}$$

where $\mathcal{R}(\nabla \boldsymbol{u})$ can be either isotropic or anisotropic TV$_p$ regularity. Similarly, we use the splitting technique and ADMM scheme to solve the minimization problem (32), where we omit the implementation details.

We present two experiments on color image labeling, where $\boldsymbol{c}$ are fixed with the pre-defined values. We apply both isotropic and anisotropic TV$_p$MS model to the *Parrot* image and *Hats* image with $L = 6$ and $L = 10$, and compare the labeling results with the PC labeling approaches[1]. As visible on the dark region in the bottom left part of *Parrot* and the brown region in the top right part of *Hats* in Fig. 8, the anisotropic TV regularized PC model favors horizontal and vertical structures because the Manhattan distance is minimized instead of the Euclidean distance. Due to the mimetic finite difference method, both the anisotropic and isotropic TV$_p$ based PS models give very similar labeling results. In addition, the

[1]The codes were downloaded from http://www.gipsa-lab.grenoble-inp.fr/ laurent.condat/publications.html.

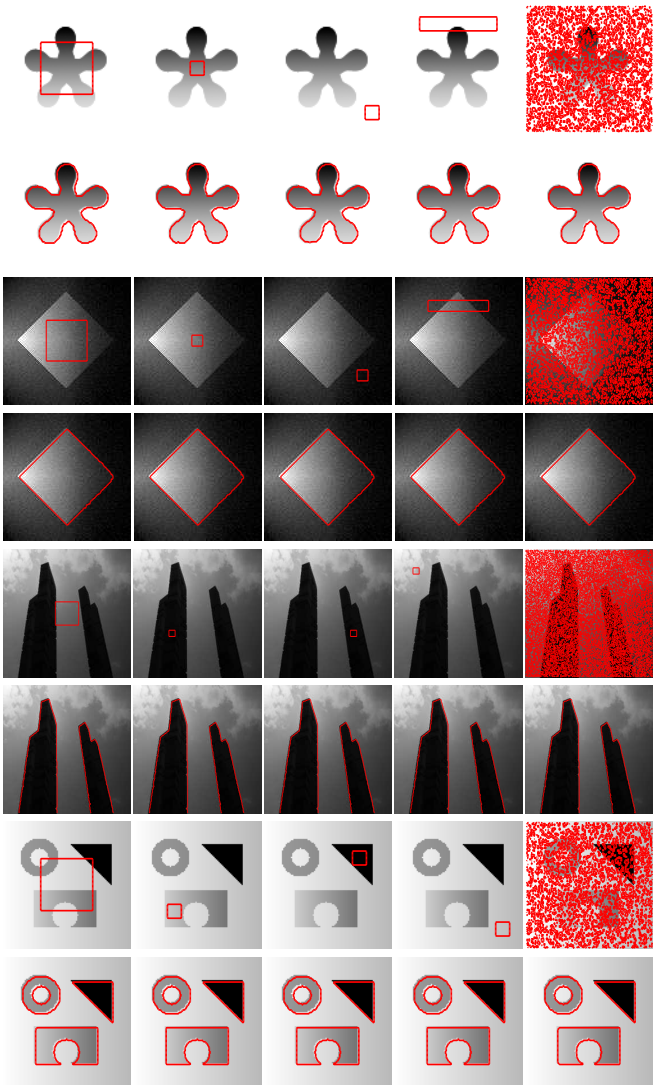
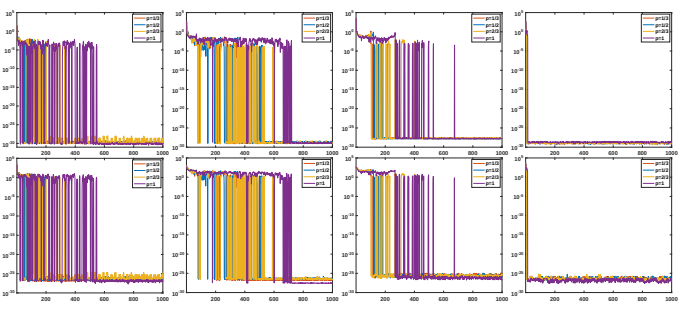


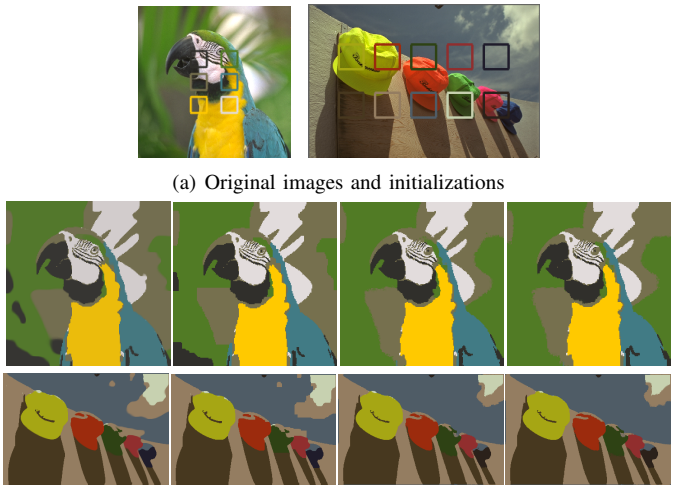
Fig 6. Labeling results of the isotropic TV$_p$MS model with different contour initializations.

comparison between the PC model and PS model demonstrates that the PS model can achieve better labeling results, which classifies the green plants into the same region for *Parrot* and locates the boundary of green hat accurately for *Hats*.



(a) Petal (b) Square (c) Building (d) Shapes

Fig 7. The evolutions of $\|b^{n+1}-b^n\|_2^2$ (top) and $\|\boldsymbol{u}^{n+1}-\boldsymbol{u}^n\|_2^2$ (bottom) via iterations for the four images in Fig. 4.



(a) Original images and initializations

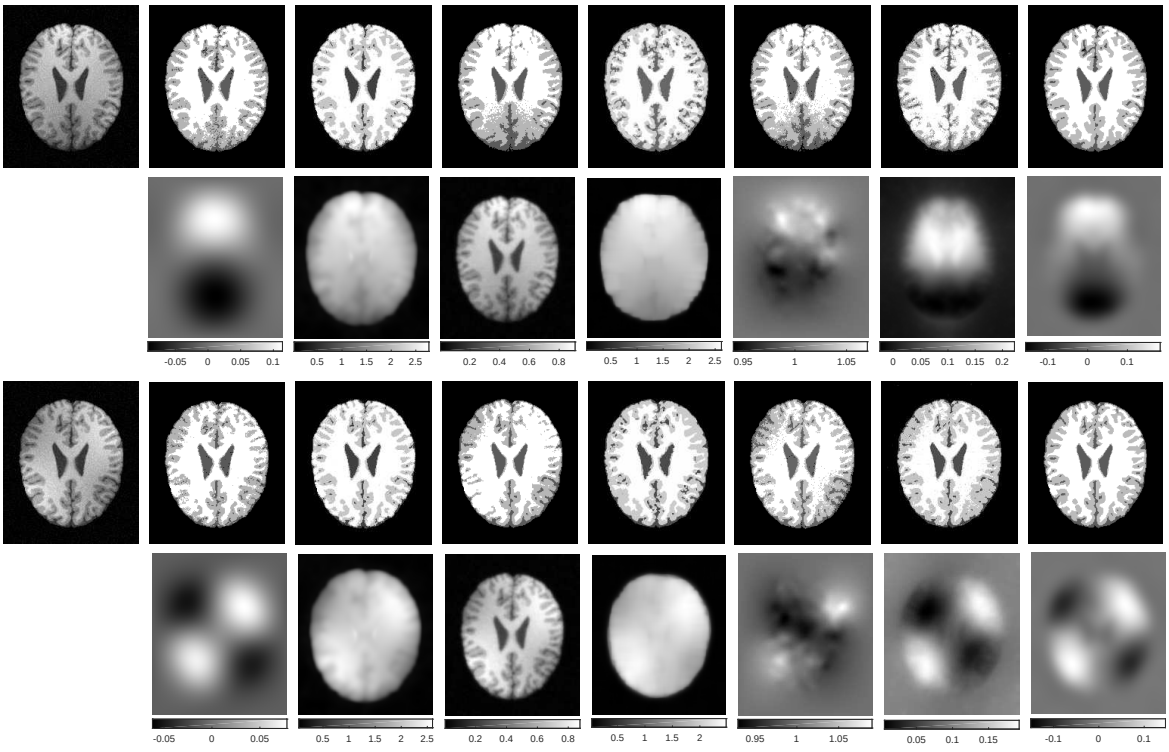
(b) TV^{iso} PC (c) TV^{ani} PC (d) TV$_p^{iso}$ PS (e) TV$_p^{ani}$ PS

Fig 8. Experiments on multi-phase color image labeling problems. The parameters for the PC model are $\lambda = 0.03,\ 0.09$ and our TV$_p$MS model are $\mu = 1000000,\ \lambda = 20,\ 9$ for *Parrot* and *Hats*.

B. Image Segmentation Applications

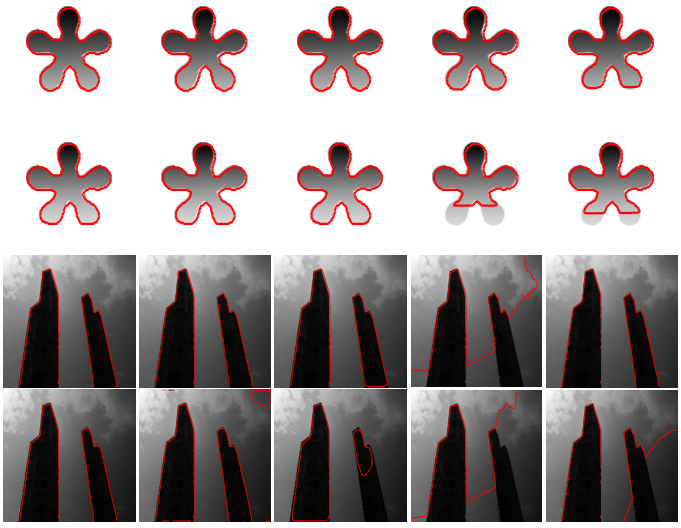
In this subsection, we apply our TV$_p$MS model to image segmentation problem, and compare with the well-known piecewise smooth segmentation approaches, including the piecewise smooth Mumford-Shah (MSL2) model [15], local intensity clustering (LIC) model [21], the additive Mumford-Shah (AMS) model [22], the maximum likelihood (ML) model [24], the ℓ^0 regularized Mumford-Shah (L0MS) model [23] and the piecewise smooth Mumford-Shah model with the L^1 data fidelity (MSL1) model [27]. We omit the discussion on the choices of parameters as it is the same as the labeling problem.

1) Comparison with TV models: Our model has two main differences with the existing MSL2 model and MSL1 model: (i) we use the non-convex TV$_p$ regularizer while they used the TV regularizer; (ii) we adopt the Laplacian operator instead of gradient in the smooth regularizer. We conduct the following experiments to verify the capacity of the non-convex regularizer in smooth segmentation on image 'Petal' and 'Building'. For each image, we compare our TV$_p$MS model with the two TV-based models on images with weak and strong intensity inhomogeneity. We also evaluate the performance of our TV$_p$MS model with different p for $p = 1/3,\ 2/3$ and 1, respectively. The visual segmentation results, JS values and parameters are displayed in Fig. 9. We can observe that our TV$_p$MS model can provide accurate segmentation results when p is small such as $p = 1/3$, while both MSL2 model and MSL1 model fail to identify the correct boundaries for strong intensity inhomogeneity. The comparison between our model with $p = 1$ and the two TV-based models reveals that the Laplacian operator outperforms the gradient operator for piecewise smooth segmentation; see the second row in Fig. 9. Obviously, the TV$_p$ regularizer for the piecewise constant component is more important in improving the segmentation accuracy when images are corrupted by strong intensity inhomogeneity; see the fourth row in Fig. 9.



(a) MSL2 (b) LIC (c) AMS (d) ML (e) L0MS (f) MSL1 (g) TV$_p$MS

Fig 10. Performances of the MSL2, LIC, AMS, ML, L0MS, MSL1 and our TV$_p$MS model on synthetic brain images. The parameters for the MSL2 model are $\mu = 5000$, $\lambda = 1500$; the LIC model are $\mu = 1.0$, $\sigma = 4$; the AMS model are $\mu = 0.005$, $\nu = 1$; the ML model are $\rho = 4.5$; the L0MS model are $\alpha = 0.15$, $\mu = 1.0$, $\sigma = 4$; the MSL1 model are $\mu = 50000$, $\lambda = 100$ and our TV$_p$MS model are $\mu = 200000$, $\lambda = 100$.



(a) p=1/3 (b) p=2/3 (c) p=1 (d) MSL2 (e) MSL1

Images	parameters	p=1/3	p=2/3	p=1	MSL2	MSL1
1^{st} row	JS	**97.41**	97.41	97.41	97.62	97.37
	$\lambda, \mu(\times 10^5)$	60, 4	60, 4	60, 4	800, 0.1	80, 1
2^{nd} row	JS	**97.39**	97.39	97.39	93.54	94.54
	$\lambda, \mu(\times 10^5)$	60, 1	60, 1	60, 1	800, 0.1	80, 6
3^{rd} row	JS	**99.08**	98.92	98.66	66.83	98.97
	$\lambda, \mu(\times 10^5)$	10, 8	10, 9	10, 10	1000, 0.4	70, 5
4^{th} row	JS	**99.04**	96.96	90.68	58.02	82.53
	$\lambda, \mu(\times 10^5)$	10, 6	10, 6	10, 7	1000, 0.1	70, 9

Fig 9. Segmentation results of isotropic TV$_p$MS, MSL2 and MSL1 model with different degrees of inhomogeneity.

2) Comparison with PS segmentation methods: We evaluate the performance of the TV$_p$MS model on a synthetic brain image and compare with state-of-art PS segmentation methods. We generate 10 different images by introducing intensity inhomogeneity of different profiles and additive Gaussian white noise of mean 0 and variance 0.0002. Two of the these images are displayed in Fig. 10 as examples with the corresponding JS values of white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), iteration numbers and the CPU times provided in TABLE III. Once again, we observe that the proposed TV$_p$MS model gives the best segmentation results for all three regions, and enjoys a faster convergence than most comparative models. Meanwhile, we plot the JS values obtained by all methods for 10 images in Fig. 11, which clearly show our model is more stable and robust against intensity inhomogeneity and noises.

TABLE III
JS VALUES AND CPU TIMES (IN SECONDS) OF THE BRAIN EXPERIMENT.

Methods	1^{st} image in Fig. 10					2^{nd} image in Fig. 10				
	WM	GM	CSF	Iter	Time	WM	GM	CSF	Iter	Time
MSL2	87.70	71.79	66.46	800	8.8	84.68	67.62	67.70	800	8.5
LIC	83.72	66.80	72.26	100	3.4	83.27	65.01	72.19	100	3.4
AMS	71.71	53.33	64.86	150	4.1	76.21	57.06	78.02	150	4.1
ML	83.83	67.10	64.09	**8**	**1.56**	85.71	72.61	69.31	**15**	**2.24**
L0MS	75.00	56.31	60.89	775	6.7	80.03	63.36	69.24	775	6.6
MSL1	79.98	62.18	79.83	800	15.0	84.10	75.52	79.30	800	14.8
TV$_p$MS	**93.00**	**83.77**	**83.12**	104	5.9	**91.17**	**82.03**	**86.01**	104	5.7

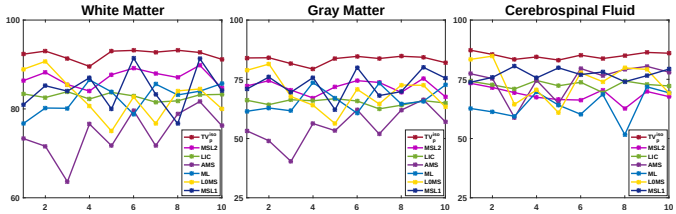
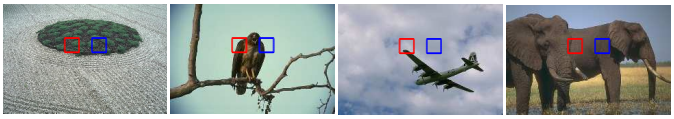
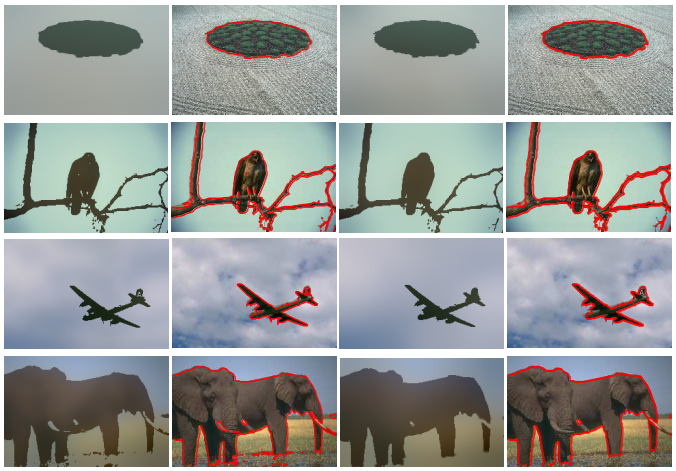


Fig 11. Comparison of the MSL2, LIC, AMS, ML, L0MS, MSL1 and our TV$_p$MS model in terms of JS (%).

3) Color image segmentation examples: In Fig. 12, we apply the two-phase segmentation model to real color images of size $161 \times 241 \times 3$ from Berkeley segmentation dataset[2]. As shown by the solution g, both the MSL1 and our TV$_p$MS models are able to achieve piecewise smooth segmentation results. Indeed, due to the merit of the nonconvex regularity in preserving edges, the proposed TV$_p$MS model gives better results with less outliers. For instance, in the fourth example, the proposed method can isolate the elephant from the background, while the MSL1 classify some grass on the bottom into the foreground. From our experiments, we find that the TV$_p$MS model is relatively less sensitive to the parameters, where the same parameters are used for all four test images and different parameters are required for the MSL1 model to achieve reasonable results.



(a) Original images and initial contours



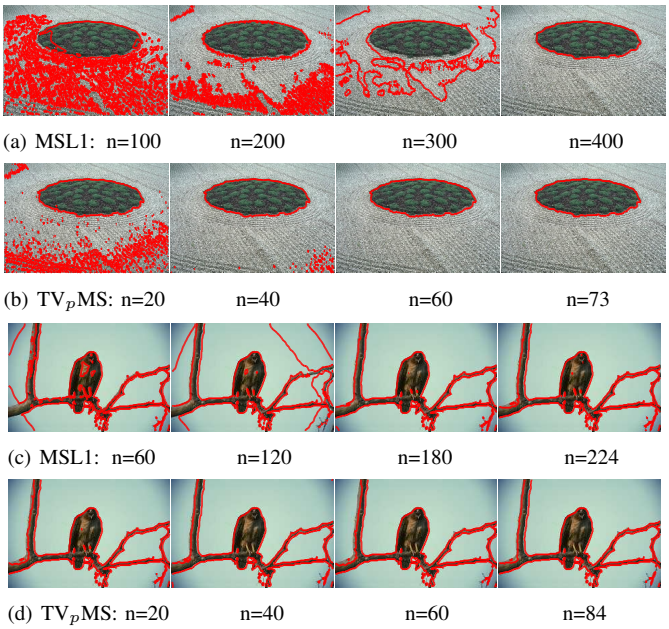
(b) MSL1: g and final curve (c) TV$_p$MS: g and final curve

	1^{st} image		2^{nd} image		3^{rd} image		4^{th} image	
	iter	time	iter	time	iter	time	iter	time
MSL1	400	18.7	224	10.9	150	7.1	400	18.6
TV$_p$MS	73	3.3	84	3.7	128	5.5	95	4.1

Fig 12. Experiments on two-phase color image segmentation. The parameters for the MSL1 model are $r = 10$ (the 1st and 2nd row), $r = 1$ (the 3rd and 4th row), $\mu = 12500$ (the 1st-3rd row), $\mu = 20000$ (the 4th row), $\lambda = 3, 20, 10, 20$, and our TV$_p$MS model are $\mu = 200000$, $\lambda = 2.2$.

[2]The image data were downloaded from https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/.

As shown by the table in Fig. 12, our model is terminated with fewer iterations and saves much computational time. In addition, we track and compare the evolution of the contours along with the iterations of the MSL1 and TV$_p$MS model for the first and second image in Fig. 12. The intermediate contours and the final contours obtained by two approaches are provided in Fig. 13, which demonstrate that our TV$_p$MS model converges faster than the MSL1 model under the same stopping criterion.



(a) MSL1: n=100 n=200 n=300 n=400

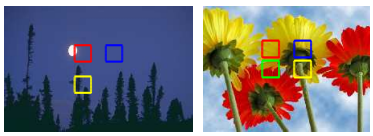
(b) TV$_p$MS: n=20 n=40 n=60 n=73

(c) MSL1: n=60 n=120 n=180 n=224

(d) TV$_p$MS: n=20 n=40 n=60 n=84

Fig 13. The contour evolutions of the MSL1 and our TV$_p$MS model along with iterations.

Fig. 14 depicts the multi-phase color image segmentation results obtained by the MSL1 model and our TV$_p$MS model, where the phase number is $L = 3$ and $L = 4$ for the *Moon* image and *Flowers* image, respectively. As shown by the magnified portions of the segmentation, our proposal yields better results for both images. For example, the MSL1 model segments the flower stem into the petal, while our model can accurately locate the boundaries between the stem and petal.



(a) Original images and initial contours

(b) MSL1 (c) TV$_p$MS (d) MSL1 (e) TV$_p$MS

Fig 14. Experiments on multi-phase color image segmentation. The parameters for the MSL1 model are $r = 2, 1$, $\mu = 100000, 55000$, $\lambda = 1.6, 25$ and our TV$_p$MS model are $\mu = 200000, 50000$, $\lambda = 80, 0.8$.

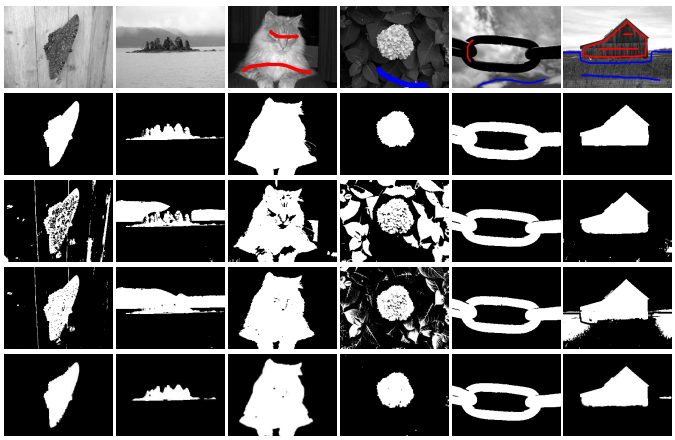
C. Evaluation on segmentation dataset

In this subsection, we evaluate our TV$_p$MS model on two public dataset: Weizmann segmentation dataset (100 images) [52] and 4.7 T anatomy dataset (30 volumes) [53].

The Weizmann dataset contains 100 gray images with objects that differ from their surroundings by either intensity, texture or other low level cues [52]. Because our TV$_p$MS model is a typical intensity based segmentation approach, we integrate our model with priori given supervision constraints. More specifically, we labelled some image pixels as foreground seeds (i.e., Ω_f) and background seeds (i.e., Ω_b) in advance, and reformulate our segmentation model (30) into a constrained minimization as follows

$$\min_{\boldsymbol{u}\in\mathcal{A},b,\boldsymbol{c}} \ \langle \boldsymbol{f},\boldsymbol{u}\rangle_{\boldsymbol{X}} + \mathcal{R}(\nabla \boldsymbol{u}) + \frac{\mu}{2}\|\Delta b\|_X^2 + \frac{\tau}{2}\|b\|_X^2,$$
$$\text{s.t.,} \quad \boldsymbol{u}(\Omega_f) = 1, \quad \boldsymbol{u}(\Omega_b) = 0.$$

Six representative test images and the corresponding segmentation results are displayed in Fig. 15, where two images are without priori, two images are with only foreground or background priori and two images are with both foreground and background priori. We use the same supervision constraints, random initialization and fixe the parameters for the LIC, MSL1 and our TV$_p$MS model in evaluating the Weizmann dataset. As shown, our TV$_p$MS method gives the best overall performance among these three intensity based approaches.

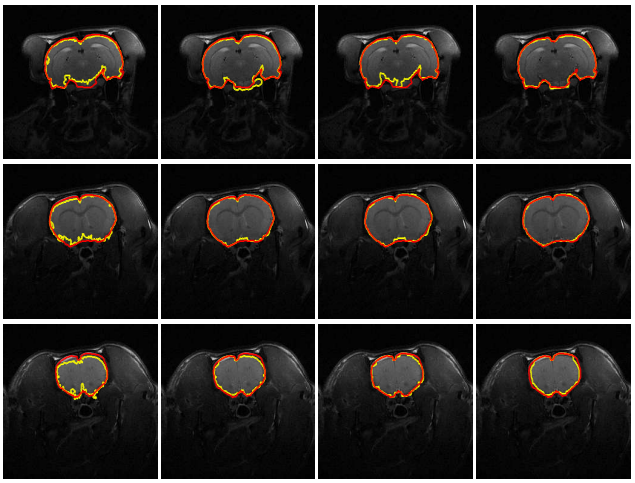


	LIC	MSL1	TV$_p$MS	% of known
JS	63.25±3.8e-2	57.26±4.4e-2	**80.66±6.7e-3**	17.67±1.1

Fig 15. Comparison of segmentation results on Weizmann dataset. 1st row: Selected images from Weizmann dataset with pixels marked as foreground (red) and background (blue). 2nd row: Ground truth of segmentation. 3rd row-5th row: Segmentation results of LIC, MSL1 and TV$_p$MS method, respectively. The parameters for the LIC model are $\mu = 1.0, \sigma = 4$; the MSL1 model are $\mu = 200000$, $\lambda = 500$ and our TV$_p$MS model are $\mu = 400000$, $\lambda = 5$.

On the anatomy dataset, we use the TV$_p$MS model for brain extraction task, which is an important preprocessing step for further analysis of brain MR images. We regard brain extraction as a three phase segmentation problem, which classifies the image domain into the brain region, the surrounding non-brain region and the background outside the body. We not only compare our proposal with the variational approach L0MS method, but also two brain extraction softwares, BSE[3] and 3D-PCNN[4]. We fix the parameters of the comparative algorithms for all volumes and perform the same morphological operations on the resultant brain masks of 3D-PCNN, L0MS and our proposal including erosion, dilation and filling holes. More implementation details can be referred to [54].

Fig. 16 shows examples of the brain boundaries identified by all comparative methods on three slices of a 3D volume. Visual results demonstrate that our TV$_p$MS method can locate the boundary quite accurately, and the estimated boundaries are much smoother than the ℓ^0 regularized model. Simultaneously, our TV$_p$MS method achieves the best JS values among the four methods, which is significantly higher than values obtained by the L0MS method. Both the qualitative and quantitative comparison demonstrate that TV$_p$ regularization can well preserve edges for image segmentation problems.



(a) BSE (b) 3D-PCNN (c) L0MS (d) TV$_p$MS

	BSE (3D)	3D-PCNN	L0MS	TV$_p$MS
JS	82.02±2.6e-4	92.07±1.3e-3	80.16±4.3e-3	**93.25±2.1e-4**

Fig 16. Comparison of brain masks (axial view) and JS values, where red curves denote the manual gold standard and yellow curves denote the automatic method. The parameters for the L0MS model are $\alpha = 0.2$, $\mu = 1.0$, $\sigma = 4$ and our TV$_p$MS model are $\mu = 250000$, $\lambda = 8$.

VI. Conclusion

In this paper, we presented a novel piecewise smooth Mumford-Shah model by regarding the solution as a linear combination of a piecewise constant function and a smooth function. Unlike most existing methods, which rely on the minimization of total variation, we utilized the non-convex and non-smooth TV$_p$ regularization for the piecewise constant component and the Lapacian operator for the smooth component. The optimization algorithm with implementation details and convergence verification was given. Both segmentation and labeling experiments on gray and color images, including medical and natural images, demonstrated the good performance of the proposed model in terms of accuracy and the computational efficiency.

[3]http://brainsuite.org/

[4]http://www.sbic. astar.edu.sg/research/lmi/PCNN3D%20binary.zip

APPENDIX A
PROOF OF THEOREM 2

Proof. By Algorithm I, we have

$$\begin{cases} \partial\mathcal{R}(\boldsymbol{q}^{n+1}) - \boldsymbol{\Lambda}_2^n + \dfrac{1}{r_2}(\boldsymbol{q}^{n+1} - \nabla\boldsymbol{v}^n) \ni 0, \\ \boldsymbol{u}^{n+1} \in \arg\min\limits_{\boldsymbol{u}} \delta_{\mathcal{A}}(\boldsymbol{u}) + \langle \boldsymbol{f}, \boldsymbol{u}\rangle_{\boldsymbol{X}} + \langle \boldsymbol{\Lambda}_1^n, \boldsymbol{u}\rangle_{\boldsymbol{X}} + \dfrac{1}{2r_1}\|\boldsymbol{u} - \boldsymbol{v}^n\|_2^2, \\ (r_2\mathcal{I} - r_1\triangle)\boldsymbol{v}^{n+1} = r_2(\boldsymbol{u}^{n+1} + r_1\boldsymbol{\Lambda}_1^n) - r_1\mathrm{div}(\boldsymbol{q}^{n+1} - r_2\boldsymbol{\Lambda}_2^n), \\ \Big((\lambda+\eta)\mathcal{I} + \mu\Delta^2\Big)b^{n+1} = \lambda\langle \boldsymbol{u}^{n+1}, I - \boldsymbol{c}\rangle_{\boldsymbol{X}}, \\ \boldsymbol{\Lambda}_1^{n+1} = \boldsymbol{\Lambda}_1^n - \dfrac{1}{r_1}(\boldsymbol{v}^{n+1} - \boldsymbol{u}^{n+1}), \\ \boldsymbol{\Lambda}_2^{n+1} = \boldsymbol{\Lambda}_2^n - \dfrac{1}{r_2}(\boldsymbol{q}^{n+1} - \nabla\boldsymbol{v}^{n+1}). \end{cases} \tag{33}$$

The assumption $\boldsymbol{\Lambda}_1^{n+1} - \boldsymbol{\Lambda}_1^n \to 0$ as $n \to \infty$ gives immediately $\lim\limits_{n\to\infty}(\boldsymbol{v}^n - \boldsymbol{u}^n) = 0$. Together with $\boldsymbol{u}^{n+1} - \boldsymbol{u}^n \to 0$, we can express $\boldsymbol{v}^{n+1} - \boldsymbol{v}^n = \boldsymbol{v}^{n+1} - \boldsymbol{u}^{n+1} + \boldsymbol{u}^{n+1} - \boldsymbol{u}^n + \boldsymbol{u}^n - \boldsymbol{v}^n$ and deduce that $\lim\limits_{n\to\infty}(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n) = 0$. Let $A := (\lambda+\eta)\mathcal{I} + \mu\Delta^2$. Due to the invertibility of A and the assumption $\boldsymbol{u}^{n+1} - \boldsymbol{u}^n \to 0$ as $n \to \infty$, there is $\lim\limits_{n\to\infty}(b^{n+1} - b^n) = 0$. The assumption $\boldsymbol{\Lambda}_2^{n+1} - \boldsymbol{\Lambda}_2^n \to 0$ as $n \to \infty$ gives immediately $\lim\limits_{n\to\infty}(\boldsymbol{q}^n - \nabla\boldsymbol{v}^n) = 0$. As $\boldsymbol{q}^{n+1} - \boldsymbol{q}^n = \boldsymbol{q}^{n+1} - \nabla\boldsymbol{v}^{n+1} + \nabla\boldsymbol{v}^{n+1} - \nabla\boldsymbol{v}^n + \nabla\boldsymbol{v}^n - \boldsymbol{q}^n$, we further deduce that $\lim\limits_{n\to\infty}(\boldsymbol{q}^{n+1} - \boldsymbol{q}^n) = 0$.

Assume $(\boldsymbol{q}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, b^*; \boldsymbol{\Lambda}_1^*, \boldsymbol{\Lambda}_2^*)$ be a cluster point of the sequence $\{(\boldsymbol{q}^n, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n)\}$. Thus, there is a subsequence $\{(\boldsymbol{q}^{n_i}, \boldsymbol{u}^{n_i}, \boldsymbol{v}^{n_i}, b^{n_i}; \boldsymbol{\Lambda}_1^{n_i}, \boldsymbol{\Lambda}_2^{n_i})\}$ converging to $(\boldsymbol{q}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, b^*; \boldsymbol{\Lambda}_1^*, \boldsymbol{\Lambda}_2^*)$. The asymptotic regularity of $\{(\boldsymbol{q}^n, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n)\}$ leads to $\lim\limits_{i\to\infty}(\boldsymbol{q}^{n_i+1}, \boldsymbol{u}^{n_i+1}, \boldsymbol{v}^{n_i+1}, b^{n_i+1}; \boldsymbol{\Lambda}_1^{n_i+1}, \boldsymbol{\Lambda}_2^{n_i+1}) = (\boldsymbol{q}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, b^*; \boldsymbol{\Lambda}_1^*, \boldsymbol{\Lambda}_2^*)$. Let n be n_i in (33) and sending i to ∞ except for the first formula, we get

$$\partial\mathcal{R}(\boldsymbol{q}^{n_i+1}) - \boldsymbol{\Lambda}_2^{n_i} + \frac{1}{r_2}(\boldsymbol{q}^{n_i+1} - \nabla\boldsymbol{v}^{n_i+1}) \ni 0, \tag{34}$$

and

$$\begin{cases} \partial\delta_{\mathcal{A}}(\boldsymbol{u}^*) + \boldsymbol{\Lambda}_1^* + \dfrac{\lambda}{2}(I - b^* - \boldsymbol{c})^2 \ni 0, \\ \boldsymbol{\Lambda}_1^* + \mathrm{div}\boldsymbol{\Lambda}_2^* = 0, \\ \Big((\lambda+\eta)\mathcal{I} + \mu\Delta^2\Big)b^* = \lambda\langle \boldsymbol{u}^*, I - \boldsymbol{c}\rangle_{\boldsymbol{X}}, \\ \boldsymbol{v}^* = \boldsymbol{u}^*, \\ \boldsymbol{q}^* = \nabla\boldsymbol{u}^*, \end{cases} \tag{35}$$

As $\lim\limits_{i\to\infty}\boldsymbol{\Lambda}_2^{n_i} = \boldsymbol{\Lambda}_2^*$, $\lim\limits_{i\to\infty}\boldsymbol{q}^{n_i} = \boldsymbol{q}^*$, and $\lim\limits_{i\to\infty}(\boldsymbol{q}^{n_i+1} - \nabla\boldsymbol{v}^{n_i+1}) = 0$, we have

$$\boldsymbol{\Lambda}_2^* \in \partial\mathcal{R}(\boldsymbol{q}^*) \tag{36}$$

provided that $\lim\limits_{i\to\infty}\mathcal{R}(\boldsymbol{q}^{n_i+1}) = \mathcal{R}(\boldsymbol{q}^*)$.

We now show $\lim\limits_{i\to\infty}\mathcal{R}(\boldsymbol{q}^{n_i+1}) = \mathcal{R}(\boldsymbol{q}^*)$. On one hand, by taking $(\boldsymbol{q}^{n_i+1}, \boldsymbol{u}^{n_i+1}, \boldsymbol{v}^{n_i+1}, b^{n_i+1}; \boldsymbol{\Lambda}_1^{n_i+1}, \boldsymbol{\Lambda}_2^{n_i+1})$ into (15) and the lower semi-continuity of $\mathcal{R}$, as well as (35), it holds that

$$\begin{aligned} &\liminf_{i\to\infty}\mathcal{L}(\boldsymbol{q}^{n_i+1}, \boldsymbol{u}^{n_i+1}, \boldsymbol{v}^{n_i+1}, b^{n_i+1}; \boldsymbol{\Lambda}_1^{n_i+1}, \boldsymbol{\Lambda}_2^{n_i+1}) \\ &\geq \frac{\lambda}{2}\sum_{k=1}^{L}\langle |I - b^* - c_k|^2, u_k^*\rangle_X + \mathcal{R}(\boldsymbol{q}^*) + \frac{\mu}{2}\|\triangle b^*\|_X^2 + \frac{\eta}{2}\|b^*\|_X^2. \end{aligned}$$

On the other hand, we have

$$\begin{cases} \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n) \leq \mathcal{L}(\boldsymbol{q}^*, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n), \\ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n) \leq \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n), \\ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n) \leq \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n), \\ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^{n+1}; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n) \leq \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n), \\ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^{n+1}; \boldsymbol{\Lambda}_1^{n+1}, \boldsymbol{\Lambda}_2^n) \\ \quad = \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^{n+1}; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n) + r_1\big\|\boldsymbol{\Lambda}_1^{n+1} - \boldsymbol{\Lambda}_1^n\big\|_{\boldsymbol{X}}^2, \\ \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^{n+1}; \boldsymbol{\Lambda}_1^{n+1}, \boldsymbol{\Lambda}_2^{n+1}) \\ \quad = \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^{n+1}; \boldsymbol{\Lambda}_1^{n+1}, \boldsymbol{\Lambda}_2^n) + r_2\big\|\boldsymbol{\Lambda}_2^{n+1} - \boldsymbol{\Lambda}_2^n\big\|_{\boldsymbol{Y}}^2, \end{cases}$$

which yields

$$\begin{aligned} \mathcal{L}(\boldsymbol{q}^{n+1}, \boldsymbol{u}^{n+1}, \boldsymbol{v}^{n+1}, b^{n+1}; \boldsymbol{\Lambda}_1^{n+1}, \boldsymbol{\Lambda}_2^{n+1}) &\leq \mathcal{L}(\boldsymbol{q}^*, \boldsymbol{u}^n, \boldsymbol{v}^n, b^n; \boldsymbol{\Lambda}_1^n, \boldsymbol{\Lambda}_2^n) \\ &\quad + r_1\big\|\boldsymbol{\Lambda}_1^{n+1} - \boldsymbol{\Lambda}_1^n\big\|_{\boldsymbol{X}}^2 + r_2\big\|\boldsymbol{\Lambda}_2^{n+1} - \boldsymbol{\Lambda}_2^n\big\|_{\boldsymbol{Y}}^2. \end{aligned}$$

Letting $n = n_i$ in the above and passing to the limit, we obtain

$$\begin{aligned} &\limsup_{i\to\infty}\mathcal{L}(\boldsymbol{q}^{n_i+1}, \boldsymbol{u}^{n_i+1}, \boldsymbol{v}^{n_i+1}, b^{n_i+1}; \boldsymbol{\Lambda}_1^{n_i+1}, \boldsymbol{\Lambda}_2^{n_i+1}) \\ &\leq \frac{\lambda}{2}\sum_{k=1}^{L}\langle |I - b^* - c_k|^2, u_k^*\rangle_X + \mathcal{R}(\boldsymbol{q}^*) + \frac{\mu}{2}\|\triangle b^*\|_X^2 + \frac{\eta}{2}\|b^*\|_X^2. \end{aligned}$$

Then we have $\lim\limits_{i\to\infty}\mathcal{R}(\boldsymbol{q}^{n_i+1}) = \mathcal{R}(\boldsymbol{q}^*)$. And (36) holds. (35) and (36) shows that $(\boldsymbol{q}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, b^*; \boldsymbol{\Lambda}_1^*, \boldsymbol{\Lambda}_2^*)$ is a KKT point of the constrained optimization problem (13). □

ACKNOWLEDGMENT

The authors would like to thank Prof. Miyoun Jung from Hankuk University of Foreign Studies for providing us with MATLAB code of [27]. The work is supported by NSFC 11701418, Major Science and Technology Project of Tianjin 18ZXRHSY00160 and Recruitment Program of Global Young Expert. The work of Chunlin Wu was supported by NSFC 11301289, 11531013 and 11871035, Recruitment Program of Global Young Expert, and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
[2] R. A. Adams and J. J. Fournier, "Sobolev spaces," 2003.
[3] L. Ambrosio and V. M. Tortorelli, "Approximation of functional depending on jumps by elliptic functional via γ-convergence," *Communications on Pure and Applied Mathematics*, vol. 43, no. 8, pp. 999–1036, 1990.
[4] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
[5] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.
[6] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.
[7] J. Lie, M. Lysaker, and X.-C. Tai, "A variant of the level set method and applications to image segmentation," *Mathematics of Computation*, vol. 75, no. 255, pp. 1155–1174, 2006.
[8] R. B. Potts, "Some generalized order-disorder transformations," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, no. 1, pp. 106–109, 1952.
[9] T. F. Chan and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM Journal on Applied Mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.
[10] J. Lellmann and C. Schnörr, "Continuous multiclass labeling approaches and algorithms," *SIAM Journal on Imaging Sciences*, vol. 4, no. 4, pp. 1049–1096, 2011.

[11] E. Bae, J. Yuan, and X. C. Tai, "Global minimization for continuous multiphase partitioning problems using a dual approach," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 112–129, 2011.

[12] R. Chan, A. Lanza, S. Morigi, and F. Sgallari, "Convex non-convex image segmentation," *Numerische Mathematik*, vol. 138, no. 3, pp. 635–680, 2018.

[13] A. Tsai, J. Yezzi, Anthony, and A. S. Willsky, "Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1169–1186, 2001.

[14] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.

[15] T. Le and L. Vese, "Additive and multiplicative piecewise-smooth segmentation models in a variational level set approach," *Contemporary Mathematics*, vol. 445, pp. 207–224, 2007.

[16] C. Li, C. Y. Kao, J. C. Gore, and Z. Ding, "Implicit active contours driven by local binary fitting energy," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–7.

[17] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 2, pp. 151–167, 2007.

[18] C. Li, R. Huang, Z. Ding, C. Gatenby, D. Metaxas, and J. Gore, "A variational level set approach to segmentation and bias correction of images with intensity inhomogeneity," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2008, pp. 1083–1091.

[19] F. Li, M. K. Ng, and C. Li, "Variational fuzzy Mumford–Shah model for image segmentation," *SIAM Journal on Applied Mathematics*, vol. 70, no. 7, pp. 2750–2770, 2010.

[20] K. Zhang, H. Song, and L. Zhang, "Active contours driven by local image fitting energy," *Pattern Recognition*, vol. 43, no. 4, pp. 1199–1206, 2010.

[21] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2007–2016, 2011.

[22] Y. Gu, L.-L. Wang, W. Xiong, J. Cheng, W. Huang, and J. Zhou, "Efficient and robust image segmentation with a new piecewise-smooth decomposition model," in *20th IEEE International Conference on Image Processing*, 2013, pp. 2718–2722.

[23] Y. Duan, H. Chang, W. Huang, J. Zhou, Z. Lu, and C. Wu, "The l_0 regularized Mumford–Shah model for bias correction and segmentation of medical images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3927–3938, 2015.

[24] K. Zhang, L. Zhang, K.-M. Lam, and D. Zhang, "A level set approach to image segmentation with intensity inhomogeneity," *IEEE Transactions on Cybernetics*, vol. 46, no. 2, pp. 546–557, 2016.

[25] H. Chang, W. Huang, C. Wu, S. Huang, C. Guan, S. Sekar, K. K. Bhakoo, and Y. Duan, "A new variational method for bias correction and its applications to rodent brain extraction," *IEEE Transactions on Medical Imaging*, vol. 36, no. 3, pp. 721–733, 2017.

[26] Y. Gu, W. Xiong, L.-L. Wang, and J. Cheng, "Generalizing Mumford-Shah model for multiphase piecewise smooth image segmentation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 942–952, 2017.

[27] M. Jung, "Piecewise-smooth image segmentation models with L^1 data-fidelity terms," *Journal of Scientific Computing*, vol. 70, no. 3, pp. 1229–1261, 2017.

[28] Y. Gu, W. Xiong, L.-L. Wang, J. Cheng, W. Huang, and J. Zhou, "A new approach for multiphase piecewise smooth image segmentation," in *21st IEEE International Conference on Image Processing*, 2014, pp. 4417–4421.

[29] W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–442, 1996.

[30] D. L. Pham and J. L. Prince, "An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recognition Letters*, vol. 20, no. 1, pp. 57–68, 1999.

[31] M. Storath, A. Weinmann, J. Frikel, and M. Unser, "Joint image reconstruction and segmentation using the Potts model," *Inverse Problems*, vol. 31, no. 2, p. 025003, 2015.

[32] X. Chen, M. K. Ng, and C. Zhang, "Non-lipschitz ℓ_p-regularization and box constrained model for image restoration," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4709–4721, 2012.

[33] M. Hintermuüller and T. Wu, "Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1385–1415, 2013.

[34] R. Chan, A. Lanza, S. Morigi, and F. Sgallari, "An adaptive strategy for the restoration of textured images using fractional order regularization," *Numerical Mathematics: Theory, Methods and Applications*, vol. 6, no. 1, pp. 276–296, 2013.

[35] A. Lanza, S. Morigi, and F. Sgallari, "Constrained TVp model for image restoration," *Journal of Scientific Computing*, vol. 68, no. 1, pp. 64–91, 2016.

[36] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM Journal on Multiscale Modeling & Simulation*, vol. 4, no. 3, pp. 960–991, 2006.

[37] C. Zeng and C. Wu, "On the edge recovery property of noncovex nonsmooth regularization in image restoration," *SIAM Journal on Numerical Analysis*, vol. 56, no. 2, pp. 1168–1182, 2018.

[38] C. Zeng, R. Jia, and C. Wu, "An iterative support shrinking algorithm for non-Lipschitz optimization in image restoration," *Journal of Mathematical Imaging and Vision*, pp. 1–18, 2018.

[39] Y. Li and Y. Duan, "Piecewise smooth segmentation with sparse prior," in *25th IEEE International Conference on Image Processing*, 2018, pp. 2217–2221.

[40] Y. Gu, L.-L. Wang, and X.-C. Tai, "A direct approach toward global minimization for multiphase labeling and segmentation problems," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2399–2411, 2012.

[41] C. Wu and X.-C. Tai, "Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 300–339, 2010.

[42] M. Held, P. Wolfe, and H. P. Crowder, "Validation of subgradient optimization," *Mathematical Programming*, vol. 6, no. 1, pp. 62–88, 1974.

[43] C. Michelot, "A finite algorithm for finding the projection of a point onto the canonical simplex of n," *Journal of Optimization Theory and Applications*, vol. 50, no. 1, pp. 195–200, 1986.

[44] C. Wu, J. Zhang, and X.-C. Tai, "Augmented Lagrangian method for total variation restoration with non-quadratic fidelity," *Inverse Problems & Imaging*, vol. 5, no. 1, pp. 237–261, 2011.

[45] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *Multiscale Modeling & Simulation*, vol. 4, no. 3, pp. 960–991, 2005.

[46] X. Chen, F. Xu, and Y. Ye, "Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization," *SIAM Journal on Scientific Computing*, vol. 32, no. 5, pp. 2832–2852, 2010.

[47] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 217–224.

[48] K. Bredies, D. A. Lorenz, and S. Reiterer, "Minimization of non-smooth, non-convex functionals by iterative thresholding," *Journal of Optimization Theory and Applications*, vol. 165, no. 1, pp. 78–112, 2015.

[49] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, pp. 1–35, 2018.

[50] C. Wu, Z. Liu, and S. Wen, "A general truncated regularization framework for contrast-preserving variational signal and image restoration: Motivation and implementation," *Science China Mathematics*, vol. 61, no. 9, pp. 1711–1732, 2018.

[51] M. Jung, G. Peyré, and L. D. Cohen, "Nonlocal active contours," *SIAM Journal on Imaging Sciences*, vol. 5, no. 3, pp. 1022–1054, 2012.

[52] S. Alpert, M. Galun, A. Brandt, and R. Basri, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 315–327, 2011.

[53] M. Murugavel and J. M. Sullivan Jr, "Automatic cropping of MRI rat brain volumes using pulse coupled neural networks," *Neuroimage*, vol. 45, no. 3, pp. 845–854, 2009.

[54] N. Chou, J. Wu, J. B. Bingren, A. Qiu, and K.-H. Chuang, "Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2554–2564, 2011.