Multiclass Segmentation by Iterated ROF Thresholding

Xiaohao Cai and Gabriele Steidl

Department of Mathematics,
University of Kaiserslautern, Kaiserslautern, Germany
{cai,steidl}@mathematik.uni-kl.de

Abstract. Variational models as the Mumford-Shah model and the active contour model have many applications in image segmentation. In this paper, we propose a new multiclass segmentation model by combining the Rudin-Osher-Fatemi model with an iterative thresholding procedure. We show that our new model for two classes is indeed equivalent to the Chan-Vese model but with an adapted regularization parameter which allows to segment classes with similar gray values. We propose an efficient algorithm and discuss its convergence under certain conditions. Experiments on cartoon, texture and medical images demonstrate that our algorithm is not only fast but provides very good segmentation results in comparison with other state-of-the-art segmentation models in particular for images containing classes of similar gray values.

1   Introduction

Throughout this paper, let $\Omega \subset \mathbb{R}^2$ be a bounded, open set and $f : \bar{\Omega} \to [0,1]$ a given image. In 1989, Mumford and Shah in [23] proposed to solve segmentation problems by minimizing over $(\Gamma, u) \in \Omega \times W_2^1(\Omega \setminus \Gamma)$ the energy

$$E_{\mathrm{MS}}(\Gamma, u) := \mathcal{H}^1(\Gamma) + \mu \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \lambda \int_{\Omega} (u - f)^2 dx, \quad \lambda, \mu > 0,$$

where $\mathcal{H}^1$ denotes the $1D$ Hausdorff measure. The functional E_{MS} contains three terms: the regularity term on Γ in terms of its length, the regularity term imposing smoothness of u on areas $\Omega \setminus \Gamma$, and the data fidelity term. Related approaches in a spatially discrete setting were proposed in [8,18]. An early attempt to solve the challenging task of finding a minimizer of the non-convex, non-smooth Mumford-Shah functional was done by approximating it by a sequence of simpler elliptic problems in [3]. Many approaches to simplify the model were meanwhile proposed in the literature as, e.g., the convex relaxation of the piecewise smooth Mumford-Shah functional by functional lifting in [26]. A frequently applied strategy is to restrict the model to $\nabla u = 0$ on $\Omega \setminus \Gamma$ which results in the piecewise constant Mumford-Shah model

$$E_{\mathrm{PCMS}}(\Gamma, u) := \mathcal{H}^1(\Gamma) + \lambda \int_{\Omega} (u - f)^2 dx. \tag{1}$$

Assuming that $\Omega = \bigcup_{i=0}^{K-1} \Omega_i$ with pairwise disjoint sets Ω_i and $u(x) := m_i$ for $x \in \Omega_i$, $i = 0, \ldots K - 1$, the above functional can be rewritten as

$$E_{\mathrm{PCMS}}(\boldsymbol{\Omega}, \boldsymbol{m}) = \frac{1}{2} \sum_{i=0}^{K-1} \mathrm{Per}(\Omega_i; \Omega) + \lambda \sum_{i=0}^{K-1} \int_{\Omega_i} (m_i - f)^2 \, dx, \qquad (2)$$

where $\mathrm{Per}(\Omega_i; \Omega)$ denotes the perimeter of Ω_i in Ω and $\boldsymbol{m} := (m_i)_{i=0}^{K-1}$, $\boldsymbol{\Omega} := (\Omega_i)_{i=0}^{K-1}$. For $K = 2$, the piecewise constant Mumford-Shah model is actually the model of the active contours without edges (Chan-Vese model) [16], i.e.

$$E_{\mathrm{CV}}(\Omega_1, m_0, m_1) = \mathrm{Per}(\Omega_1; \Omega) + \lambda \Big(\int_{\Omega_1} (m_1 - f)^2 \, dx + \int_{\Omega \backslash \Omega_1} (m_0 - f)^2 \, dx \Big). \quad (3)$$

One of the model's drawbacks is that it can easily get stuck in local minima. To overcome this drawback, a convex relaxation approach was proposed in [15]. More precisely, it was shown that the global minimizer of $E_{\mathrm{CV}}(\cdot, m_0, m_1)$ for fixed m_0, m_1 can be found by solving

$$\min_{0 \le u \le 1} \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} \big((m_0 - f)^2 - (m_1 - f)^2 \big) u(x) dx, \qquad (4)$$

and setting $\Omega_1 := \{x \in \Omega : u(x) > \rho\}$ for any $\rho \in (0, 1]$, see also [6,9]. In other words, (4) is a tight relaxation of the Chan-Vese model with fixed m_i, $i = 0, 1$. There are many other approaches for two-phase image segmentation based on the Chan-Vese model and its convex version, see, e.g., [31], [9] and [17].

In [28], Chan and Vese proposed a multiphase segmentation model using level sets. Convex (non-tight) relaxation approaches for the model with fixed $\boldsymbol{m}$ were proposed, e.g., in [21,22,25,29,30] and for the full model in [10]. For more details see also [5].

In [12] a two-stage image segmentation method which finds the solution of a convex variant of the Mumford-Shah model in the first stage followed by *one* thresholding step in the second stage was proposed. The applied functional was the Rudin-Osher-Fatemi (ROF) functional [27] supplemented by the *essential additional term* $\int |\nabla u|^2 \, dx$.

In this paper, we propose a new multiphase segmentation model based on *iteratively* thresholding the minimizer of the *original* ROF functional. In contrast to [12] we propose a strategy to update the thresholds and prove its convergence under certain conditions. There exists a clear relationship of our new model to the Chan-Vese model (3) which shows that a solution of (3) for a certain regularization parameter can actually be given by iteratively thresholding the ROF minimizer. Numerical examples demonstrate that our algorithm is not only fast but produces also very good results for images whose classes are close to each other. In particular it outperforms the algorithm in [12].

The paper is organized as follows: In Section 2, we introduce our segmentation model and discuss its the properties. We propose an efficient solution algorithm and provide a convergence analysis in Section 3. Finally, in Section 4, we test our algorithm on various synthetic and real-world images and compare it with other state-of-the-art segmentation algorithms.

2   Continuous Model

2.1   Notation

We briefly introduce the basic notation and relations which can be found, e.g., in
[2,4]. In the following a 'set' is understood as a Lebesgue measurable set in $\mathbb{R}^2$,
where we will consider equivalence classes of sets which are equal up to Lebesgue
negligible sets. By $|A|$ we denote the Lebesgue measure of a set A. By $BV(\Omega)$
we denote the *space of functions of bounded variation*, i.e., the Banach space of
functions $u : \Omega \to \mathbb{R}$ with finite norm $\|u\|_{BV} := \|u\|_{L_1(\Omega)} + TV(u)$, where

$$TV(u) := \sup \big\{ \int_\Omega u(x)\mathrm{div}\varphi \, dx : \varphi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^2), \|\varphi\|_\infty \leq 1 \big\}.$$

The distributional first order derivative Du of u is a vector-valued Radon mea-
sure with total variation $|Du| = TV(u)$. In particular, we have for $u \in W_1^1(\Omega)$
that $Du = \nabla u \in L_1$ so that in this case $TV(u) = \int_\Omega |\nabla u| \, dx$. For a Lebesgue
measurable set $A \subset \Omega$, the *perimeter* of A in Ω is defined by $\mathrm{Per}(A; \Omega) :=
TV(\chi_A)$, where χ_A denotes the *characteristic function* of A. Hence A is of finite
perimeter, if its characteristic function has bounded variation. If A has a $\mathcal{C}^1$
boundary, then $\mathrm{Per}(A; \Omega)$ coincides with $\mathcal{H}^1(\partial A \cap \Omega)$. We define the mean of f
on $A \subset \mathbb{R}^2$ by

$$\mathrm{mean}_f(A) := \begin{cases} \frac{1}{|A|} \int_A f \, dx & \text{if } |A| > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We call (u^*, c^*) a *partial minimizer* of some objective function $E(u, c)$ if

$$\begin{aligned} E(u^*, c^*) &\leq E(u^*, c) \quad \text{for all feasible } c, \quad \text{and} \\ E(u^*, c^*) &\leq E(u, c^*) \quad \text{for all feasible } u. \end{aligned} \tag{5}$$

In case E is differentiable on its domain every partial minimizer contained in
the interior of the domain is stationary, see, e.g., [19]. For example we see that a
partial minimizer $(\boldsymbol{\Omega}^*, \boldsymbol{m}^*)$ of the piecewise constant Mumford-Shah model (2)
with $\boldsymbol{\Omega}^* = (\Omega_i^*)_{i=0}^{K-1}$, $\boldsymbol{m}^* = (m_i^*)_{i=0}^{K-1}$ has to fulfill

$$m_i^* = \mathrm{mean}_f(\Omega_i^*), \quad i = 0, \dots, K-1. \tag{6}$$

2.2   Model

We start by considering the segmentation into $K = 2$ classes. Let

$$\mathcal{E}(\Sigma, \tau) := \mathrm{Per}(\Sigma; \Omega) + \mu \int_\Sigma \tau - f \, dx, \quad \mu > 0. \tag{7}$$

Note that $\mathcal{E}(\emptyset, \tau) = 0$ and $\mathcal{E}(\Omega, \tau) = \mu \int_\Omega \tau - f \, dx$. Since f maps into $[0, 1]$, the
global minimizer of $\mathcal{E}(\cdot, \tau)$ for fixed $\tau \leq 0$ is Ω and for $\tau \geq 1$ it is $\emptyset$. Therefore we
restrict ourselves to $\tau \in (0, 1)$. We intend to find $(\Sigma^*, \tau^*) \in \Omega \times (0, 1)$ fulfilling

$$\mathcal{E}(\Sigma^*, \tau^*) \leq \mathcal{E}(\Sigma, \tau^*) \quad \forall \Sigma \subset \Omega, \tag{8}$$

$$\tau^* = \frac{1}{2} \big(\mathrm{mean}_f(\Sigma^*) + \mathrm{mean}_f(\Omega \backslash \Sigma^*) \big).$$

Remark 1. Note that solving (8) is different from minimizing

$$\min_{\Sigma, \tau} \mathcal{E}(\Sigma, \tau) \quad \text{subject to} \quad \tau = \frac{1}{2}\left(\text{mean}_f(\Sigma) + \text{mean}_f(\Omega \backslash \Sigma)\right). \tag{9}$$

Consider the 1D example with the function $f(x) = x$ on $\Omega = (0, 1)$ and restrict the attention to $\Sigma \in \{(0, b), (b, 1)\}$. Then $\tau = \frac{1}{4}(1 + 2b)$ in (9) and we are searching for b. Now

$$\mathcal{E}(\Sigma, \tau) = 1 + \mu \int_I \tau - x\, dx = \begin{cases} 1 + \frac{1}{4}\mu b & \text{if } I = (0, b), \\ 1 - \frac{1}{4}\mu(1 - b) & \text{if } I = (b, 1) \end{cases}$$

which has no minimizer in $(0, 1)$. On the other hand, let $\tau^* = \frac{1}{2}$. Then it can easily be checked that (Σ^*, τ^*) with $\Sigma^* = (\frac{1}{2}, 1)$ fulfills (8).

The following proposition ensures the existence of a global minimizer of $\mathcal{E}(\cdot, \tau)$ for fixed τ.

Proposition 1. *For any fixed $\tau \in (0, 1)$, a global minimizer Σ of $\mathcal{E}(\cdot, \tau)$ in (7) can be found by solving the convex minimization problem*

$$\min_{u \in BV(\Omega), u \in [0,1]} TV(u) + \mu \int_\Omega (\tau - f)\, u\, dx \tag{10}$$

and then setting $\Sigma := \{x \in \Omega : u(x) > \rho\}$ for any $\rho \in [0, 1)$.

For a proof we refer to Proposition 2.1 in the review paper [14]. This proof uses the same ideas as in [6,24] where the claim was shown for a.e. $\rho \in [0, 1]$. Based on the next lemma, cf., [1, Lemma 4i)] and a smoothness argument, an explanation that the minimizing set Σ is unique for fixed τ was given in [14].

Lemma 1. *For fixed $0 < \tau_1 < \tau_2 < 1$, let Σ_i be minimizers of $\mathcal{E}(\cdot, \tau_i)$, $i = 1, 2$. Then $|\Sigma_2 \backslash \Sigma_1| = 0$ is fulfilled, i.e., $\Sigma_1 \supseteq \Sigma_2$ up to a negligible set.*

The relationship between our model (8) and the Chan-Vese model (3) is explained in the following proposition.

Proposition 2. (Relation between the Chan-Vese model and (8))
Assume that (Σ^*, τ^*), $\Sigma^* \notin \{\emptyset, \Omega\}$ is a solution of (8). Set $m_0^* := \text{mean}_f(\Sigma^*)$ and $m_1^* := \text{mean}_f(\Omega \setminus \Sigma^*)$. Let $\mathcal{O}$ be the set of partial minimizers of the Chan-Vese model (3) with parameter $\lambda := \frac{\mu}{2(m_1^* - m_0^*)}$. Then $(\Sigma^*, m_0^*, m_1^*) \in \mathcal{O}$.

Proof. Since $\Sigma^* \neq \emptyset$ is a minimizer of $\mathcal{E}(\cdot, \tau^*)$ we conclude $\int_{\Sigma^*} \tau^* - f\, dx < 0$ which implies $\tau^* < \text{mean}_f(\Sigma^*) = m_1^*$. Similarly, since $\Sigma^* \neq \Omega$, we see that

$$\text{Per}(\Sigma^*; \Omega) + \mu \int_{\Sigma^*} \tau^* - f\, dx \leq \mu \int_\Omega \tau^* - f\, dx,$$

$$0 < \text{Per}(\Sigma^*; \Omega) \leq \mu \int_{\Omega \backslash \Sigma^*} \tau^* - f\, dx$$

and consequently $m_0^* = \text{mean}_f(\Omega \setminus \Sigma^*) < \tau^*$. Therefore $m_0^* < m_1^*$. The set Σ^* is also a minimizer of $\mathcal{E}(\cdot, \tau^*) + C$ with the constant $C := \lambda \int_\Omega (m_0^* - f)^2 \, dx$. Regarding that $\tau^* = \frac{m_1^* + m_0^*}{2}$ we obtain

$$\mathcal{E}(\Sigma, \tau^*) + C = \text{Per}(\Sigma; \Omega) + \mu \int_\Sigma \tau^* - f \, dx + C$$

$$= \text{Per}(\Sigma; \Omega) + \frac{\mu}{2(m_1^* - m_0^*)} \int_\Sigma (m_1^* - f)^2 - (m_0^* - f)^2 \, dx$$

$$+ \lambda \int_\Omega (m_0^* - f)^2 \, dx$$

$$= \text{Per}(\Sigma; \Omega) + \lambda \left(\int_\Sigma (m_1^* - f)^2 dx + \int_{\Omega \setminus \Sigma} (m_0^* - f)^2 \right).$$

By definition of m_i^*, $i = 0, 1$ and (6) we get the assertion. $\square$

Since $0 < m_1^* - m_0^* \leq 1$ the parameter $\lambda = \frac{\mu}{2(m_1^* - m_0^*)}$ in the Chan-Vese model (3) is larger than μ and increases if $m_1^* - m_0^*$ becomes smaller. Hence, it is adapted to the difference between m_1^*, m_0^* and penalizes the data term more if this difference becomes smaller.

The following proposition has the important consequence that we can obtain a minimizer Σ of $\mathcal{E}(\cdot, \tau)$ by minimizing the ROF functional and subsequent thresholding of the minimizing function by τ.

Proposition 3. *The set $\{x \in \Omega : u(x) > \tau\}$ solves (7) if and only if the function $u \in BV(\Omega)$ solves the ROF model*

$$\min_{u \in BV(\Omega)} TV(u) + \frac{\mu}{2} \int_\Omega (u - f)^2 \, dx. \tag{11}$$

For the proof see [14, Proposition 2.6].

We generalize (7) and (8) to the multiclass case $K \geq 2$ by setting $\boldsymbol{\Sigma} := \{\Sigma_i\}_{i=1}^{K-1}$ and $\boldsymbol{\tau} := \{\tau_i\}_{i=1}^{K-1}$ with $0 < \tau_1 \leq \tau_2 \leq \ldots \leq \tau_{K-1} < 1$, and

$$\mathcal{E}(\boldsymbol{\Sigma}, \boldsymbol{\tau}) := \sum_{i=1}^{K-1} \left(\text{Per}(\Sigma_i; \Omega) + \mu \int_{\Sigma_i} \tau_i - f \, dx \right), \quad \mu > 0. \tag{12}$$

For fixed $\boldsymbol{\tau}$, we know by Lemma 1 that

$$\Omega \supseteq \Sigma_{\tau_1} \supseteq \Sigma_{\tau_2} \supseteq \cdots \supseteq \Sigma_{\tau_{K-1}} \supseteq \emptyset \tag{13}$$

and we see that the corresponding wanted segments

$$\Omega_i := \Sigma_i \setminus \Sigma_{i+1}, \ i = 0, \ldots, K-1, \ \Sigma_0 := \Omega, \ \Sigma_K := \emptyset \tag{14}$$

are pairwise disjoint and fulfill $\cup_{i=0}^{K-1} \Omega_i = \Omega$. We aim to find an ordered vector $\boldsymbol{\tau}^*$ and a corresponding nested set $\boldsymbol{\Sigma}^*$ with

$$\tau_i^* = \frac{1}{2}(m_{i-1}^* + m_i^*), \quad m_i^* := \text{mean}_f(\Omega_i^*), \quad i = 1, \ldots, K-1 \tag{15}$$

which minimizes $\mathcal{E}(\cdot, \boldsymbol{\tau}^*)$ among all sequences of nested sets.

3   Algorithmic Aspects

Our algorithm alternates the minimization of $\mathcal{E}(\cdot, \boldsymbol{\tau})$ in (12) for fixed $\boldsymbol{\tau}$ with $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_{K-1}$ and the computation of $\boldsymbol{\tau}$ as in (15) for fixed sequences $\boldsymbol{\Sigma}$ of nested sets. By Proposition 3 the minimization of $\mathcal{E}(\cdot, \boldsymbol{\tau})$ in (12) for fixed $\boldsymbol{\tau}$ can be obtained by $K-1$ times thresholding the minimizer of the ROF functional. This is in particular efficient since the minimizer of the ROF functional remains the same during the whole thresholding process. We call the algorithm *thresholded ROF* (T-ROF).

Algorithm (T-ROF)

Initialization: $\tau^{(0)} = \left(\tau_i^{(0)}\right)_{i=1}^{K-1}$ with $0 < \tau_1^{(0)} < \cdots < \tau_{K-1}^{(0)} < 1$.

1. Compute the solution u of the ROF model (11).
2. For $k = 0, 1, \ldots$, repeat
    2.1. Compute $\boldsymbol{\Sigma}^{(k)} = \left(\Sigma_i^{(k)}\right)_{i=1}^{K-1}$ by $\Sigma_i^{(k)} := \{x \in \Omega : u(x) > \tau_i^{(k)}\}$.
    2.2. Find $\Omega_i^{(k)} := \Sigma_i^{(k)} \backslash \Sigma_{i+1}^{(k)}$, $i = 0, \ldots, K-1$ with $\Sigma_0^{(k)} := \Omega$ and $\Sigma_K^{(k)} := \emptyset$.
    2.3. Compute $m_i^{(k)} := \text{mean}_f(\Omega_i^{(k)})$, $i = 0, \ldots, K-1$.
    2.4. Update $\tau_i^{(k+1)} := \frac{1}{2}(m_{i-1}^{(k)} + m_i^{(k)})$, $i = 1, \ldots, K-1$.

We will prove the convergence of our algorithm under the following assumption:

(A) If $\Sigma_\tau, \Sigma_{\tilde{\tau}}$ are the minimizers of $\mathcal{E}(\cdot, \tau)$ and $\mathcal{E}(\cdot, \tilde{\tau})$ for any $0 < \tau < \tilde{\tau} < 1$ appearing in the algorithm, then

$$\tau \leq \text{mean}_f(\Sigma_\tau \backslash \Sigma_{\tilde{\tau}}) \leq \tilde{\tau}. \tag{16}$$

The right-hand inequality in (16) is for example fulfilled if $\Sigma_{\tilde{\tau}}$ is also a minimizer of $\text{Per}(\Sigma; \Sigma_\tau) + \mu \int_\Sigma \tilde{\tau} - f \, dx$. The left-hand inequality holds if $\Sigma_\tau \backslash \Sigma_{\tilde{\tau}}$ is also a minimizer of $\text{Per}(\Sigma; \Omega \backslash \Sigma_{\tilde{\tau}}) + \mu \int_\Sigma \tau - f \, dx$. Using the above assumption we can prove the following lemma, see [11]:

Lemma 2. *Under the assumption* (A) *our T-ROF algorithm produces sequences* $(\boldsymbol{\tau}^{(k)})_k$ *and* $(\boldsymbol{m}^{(k)})_k$ *with the following properties:*

i) $0 \leq m_0^{(k)} \leq \tau_1^{(k)} \leq m_1^{(k)} \leq \cdots \leq m_{K-2}^{(k)} \leq \tau_{K-1}^{(k)} \leq m_{K-1}^{(k)}$
ii) *Set* $\tau_0^{(k)} := 0$ *and* $\tau_K^{(k)} := 1$. *If* $\tau_i^{(k)} \geq \tau_i^{(k-1)}$ *and* $\tau_{i+1}^{(k)} \geq \tau_{i+1}^{(k-1)}$, *then* $m_i^{(k)} \geq m_i^{(k-1)}$, $i = 0, \ldots, K-1$ *and this also holds true if* $\leq$ *is replaced everywhere by* $\geq$.

To prove the convergence of the sequence $(\boldsymbol{\tau}^{(k)})_k$, we define a sign sequence $\boldsymbol{\zeta}^{(k)} = (\zeta_i^{(k)})_{i=1}^{K-1}$ as follows: If $\tau_i^{(k)} \neq \tau_i^{(k-1)}$,

$$\zeta_i^{(k)} := \begin{cases} +1 & \text{if } \tau_i^{(k)} > \tau_i^{(k-1)}, \\ -1 & \text{if } \tau_i^{(k)} < \tau_i^{(k-1)}, \end{cases}$$

and otherwise

$$\zeta_i^{(k)} := \begin{cases} \zeta_j^{(k)} & \text{if } i = 1, \\ \zeta_{i-1}^{(k)} & \text{if } i \neq 1, \end{cases} \tag{17}$$

where $j = \min\{l \mid \tau_l^{(k)} \neq \tau_l^{(k-1)}\}$. By s_k we denote the number of sign changes in $\zeta^{(k)}$, for example, if $\boldsymbol{\zeta}^{(k)} = (\overbrace{+1, +1, +1}, \overbrace{-1, -1}, \overbrace{+1}, \overbrace{-1})$, then $s_k = 3$.

Lemma 3. i) *The number of sign changes s_k is monotone decreasing in k.*
ii) *If $\zeta_1^{(k+1)} \neq \zeta_1^{(k)}$, then we have the strict decrease $s_{k+1} < s_k$.*

Proof. i) Let $s_k = N$ and rewrite $(\boldsymbol{\tau}^{(k)})_k$ as

$$(\underbrace{\tau_0^{(k)}, \cdots, \tau_{l_1}^{(k)}}_{v_1^{(k)}}, \cdots, \underbrace{\tau_{i_j}^{(k)}, \cdots, \tau_{l_j}^{(k)}}_{v_j^{(k)}}, \cdots, \underbrace{\tau_{i_N}^{(k)}, \cdots, \tau_K^{(k)}}_{v_N^{(k)}}),$$

where $v_j^{(k)}$ contains those successive components of $(\boldsymbol{\tau}^{(k)})_k$ with the same sign.
1. If $\#v_j^{(k)} \geq 3$, we consider $\zeta_{i^*}^{(k+1)}$ with $i_j \leq i^* - 1 \leq i^* \leq i^* + 1 \leq l_j$, i.e., $\zeta_{i^*-1}^{(k)} = \zeta_{i^*}^{(k)} = \zeta_{i^*+1}^{(k)}$. WLOG let $\zeta_{i^*}^{(k)} = -1$. Then we obtain by Lemma 2 ii) that $m_{i^*-1}^{(k)} \leq m_{i^*-1}^{(k-1)}$ and $m_{i^*}^{(k)} \leq m_{i^*}^{(k-1)}$. Therefore

$$\tau_{i^*}^{(k+1)} = \frac{m_{i^*-1}^{(k)} + m_{i^*}^{(k)}}{2} \leq \frac{m_{i^*-1}^{(k-1)} + m_{i^*}^{(k-1)}}{2} = \tau_{i^*}^{(k)}$$

and consequently $\zeta_{i^*}^{(k+1)} = -1$ or $\zeta_{i_j}^{(k+1)} = \zeta_{i_j+1}^{(k+1)} = \cdots = \zeta_{i^*}^{(k+1)} = 1$. The case $\zeta_{i^*}^{(k)} = 1$ can be handled in the same way.
2. If there is no j such that $\#v_j^{(k)} = 1$, we consider $\zeta_{l_j}^{(k+1)}$ and $\zeta_{i_{j+1}}^{(k+1)}$ which are different by definition. WLOG let $\zeta_{l_j}^{(k)} = -1$ and $\zeta_{i_{j+1}}^{(k)} = +1$. Then, from Lemma 2 ii), we have

$$m_{l_j-1}^{(k)} \leq m_{l_j-1}^{(k-1)}, \quad m_{i_{j+1}}^{(k)} \geq m_{i_{j+1}}^{(k-1)}.$$

If $m_{l_j}^{(k)} \leq m_{l_j}^{(k-1)}$ (or $m_{l_j}^{(k)} \geq m_{l_j}^{(k-1)}$), then $\tau_{l_j}^{(k+1)} \leq \tau_{l_j}^{(k)}$ (or $\tau_{i_{j+1}}^{(k+1)} \geq \tau_{i_{j+1}}^{(k)}$). This means that $\zeta_{l_j}^{(k+1)} \neq \zeta_{l_j}^{(k)}$ and $\zeta_{i_{j+1}}^{(k+1)} \neq \zeta_{i_{j+1}}^{(k)}$ is not possible at the same time.
3. Finally, we consider the case $\#v_j^{(k)} = 1$ for all $j_1 \leq j \leq j_2$, where $\#v_{j_1-1}^{(k)} > 1$ and $\#v_{j_2+1}^{(k)} > 1$. We prove that from iteration $k \to k+1$ the signs of $\zeta_{i_{j_1}-1}^{(\cdot)}, \zeta_j^{(\cdot)}, \ldots, \zeta_{j_2}^{(\cdot)}, \zeta_{i_{j_2}+1}^{(\cdot)}$ can not change at the same time. WLOG assume that $\zeta_{i_{j_1}}^{(k)} = -1$ so that $\zeta_{i_j}^{(k)} = (-1)^{j-j_1+1}$ for $j_1 \leq j \leq j_2$, and $\zeta_{i_{j_1}-1}^{(k)} = \zeta_{i_{j_1}-2}^{(k)} = +1$ and $\zeta_{i_{j_2}+1}^{(k)} = \zeta_{i_{j_2}+2}^{(k)} = (-1)^{j_2-j_1}$. ¿From Lemma 2 ii), we know that

$$m_{i_{j_1}-2}^{(k)} \geq m_{i_{j_1}-2}^{(k-1)}, \quad \text{and} \quad \begin{cases} m_{i_{j_2}+1}^{(k)} \geq m_{i_{j_2}+1}^{(k-1)} & \text{if } j_2 - j_1 \text{ is even,} \\ m_{i_{j_2}+1}^{(k)} \leq m_{i_{j_2}+1}^{(k-1)} & \text{if } j_2 - j_1 \text{ is odd.} \end{cases} \tag{18}$$

If in $\zeta^{(\cdot)}_{i_{j_1}-1}, \zeta^{(\cdot)}_{i_{j_1}}, \cdots, \zeta^{(\cdot)}_{i_{j_2}}, \zeta^{(\cdot)}_{i_{j_2}+1}$ the signs change at the same time, we can deduce by Lemma 2 ii) that

$$\begin{cases} m^{(k)}_{i_{j_2}+1} < m^{(k-1)}_{i_{j_2}+1} & \text{if } j_2 - j_1 \text{ is even,} \\ m^{(k)}_{i_{j_2}+1} > m^{(k-1)}_{i_{j_2}+1} & \text{if } j_2 - j_1 \text{ is odd,} \end{cases}$$

which contradicts (18).

By parts 1–3, we see that $s_{k+1} \leq s_k$ for any $k \in \mathbb{N}$.

ii) If $\zeta^{(k+1)}_1 \neq \zeta^{(k)}_1$, then $v^{(k)}_1 = \tau^{(k)}_1$ and $\tau^{(k)}_2 \in v^{(k)}_2$. By parts 1–3 of the proof we get $s_{k+1} < s_k$. This completes the proof. $\square$

Now we can prove the convergence of our T-ROF algorithm with a slight modification. We divide the interval $[0, 1)$ into $n > 1$ disjoint subintervals $[\frac{i}{n}, \frac{i+1}{n})$, and define a projector $\mathcal{P}_n : [0, 1) \to \{\frac{i}{n} : i = 0, \ldots, n - 1\}$ by $\mathcal{P}_n(x) := \frac{i}{n}$ if $x \in [\frac{i}{n}, \frac{i+1}{n})$. Clearly, $\mathcal{P}_n(x_1) \geq \mathcal{P}_n(x_2)$ if $x_1 \geq x_2$. We choose n large enough (say machine precision). Instead of $\boldsymbol{\tau}^{(k)}$ we compute in each step of the T-ROF algorithm the projection $\mathcal{P}_n(\boldsymbol{\tau}^{(k)})$ and continue the algorithm with these projected thresholds. Clearly, the statements of Lemma 2 and 3 remain true. For convenience we write again $\boldsymbol{\tau}^{(k)}$ for the output of the projected algorithm.

Theorem 1. *Under the assumption* (A), *the sequence* $(\boldsymbol{\tau}^{(k)})_{k \in \mathbb{N}}$ *produced by the projected T-ROF algorithm converges to a vector* $\boldsymbol{\tau}^*$.

Proof. We prove the assertion by induction on the number of sign changes s_k in some iteration step k.

Assume that $s_k = 0$. WLOG let $\zeta^{(k)}_i = +1, i = 1, \ldots, K-1$, i.e., $\tau^{(k)}_i \geq \tau^{(k-1)}_i$. From Lemma 2 ii), we obtain $m^{(k)}_i \geq m^{(k-1)}_i$ and consequently $\tau^{(k+1)}_i \geq \tau^{(k)}_i$, $i = 1, \ldots, K - 1$. Therefore $s_{k+1} = 0$ and $\zeta^{(k+1)}_i = +1, i = 1, \ldots, K - 1$. This means that each sequence $(\tau^{(k)}_i)_k$ is monotone increasing. Since the sequences are moreover bounded in $[0, 1]$, we conclude that $(\boldsymbol{\tau}^{(k)})_k$ converges.

Assume that $(\boldsymbol{\tau}^{(k)})_k$ converges if $s_k \leq N - 1$ for some $k \in \mathbb{N}$.

We prove that in case $s_k = N$, there exists $\hat{k} > k$ such that $s_{\hat{k}} \leq N - 1$. If there exists $\hat{k} > k$ such that $\zeta^{(\hat{k})}_1 \neq \zeta^{(k)}_1$, we get $s_{\hat{k}} \leq N - 1$ directly from Lemma 3 ii). If $\zeta^{(\hat{k})}_1 = \zeta^{(k)}_1$ for all $\hat{k} > k$, then $(\tau^{(\hat{k})}_1)_{\hat{k}>k}$ is monotone and bounded and converges consequently to some threshold τ^*_1. This threshold must be attained in the projected algorithm for some $k_1 > k$. Now we can repeat the same arguments with k_1 instead of k and $\zeta^{(\cdot)}_2$ instead of $\zeta^{(\cdot)}_1$ to see that $(\tau^{(k)}_2)_k$ converges to a threshold τ^*_2 which must be attained for some $k_2 > k_1$. Moreover, we have $\zeta^{(j)}_2 = \zeta^{(j)}_1$ for all $j > k_1$. Repeating this procedure up to the final index $K - 1$ we obtain the assertion. $\square$

4   Numerical Results

In this section, we test our method on various images. We actually use the T-ROF Algorithm with a discrete ROF model, see, e.g. [13], which minimizer is
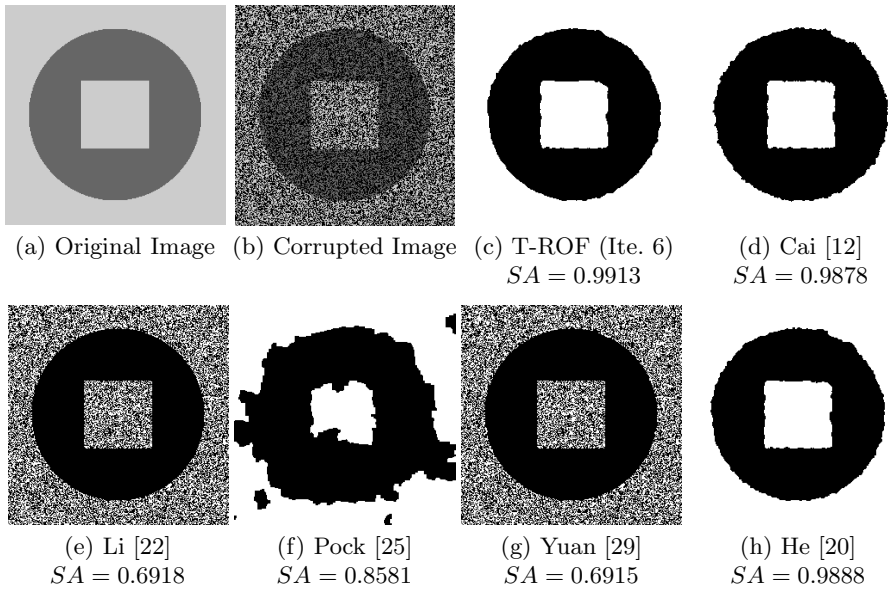
(a) Original Image   (b) Corrupted Image   (c) T-ROF (Ite. 6)   (d) Cai [12]
$SA = 0.9913$ $SA = 0.9878$

(e) Li [22]   (f) Pock [25]   (g) Yuan [29]   (h) He [20]
$SA = 0.6918$ $SA = 0.8581$ $SA = 0.6915$ $SA = 0.9888$

Fig. 1. Segmentation of two-class cartoon image (256×256) with some missing pixel values

(a) Original image   (b) Noisy image   (c) T-ROF (Ite. 6)   (d) Cai [12]
$SA = 0.9845$ $SA = 0.9816$

(e) Li [22]   (f) Pock [25]   (g) Yuan [29]   (h) He [20]
$SA = 0.7867$ $SA = 0.9658$ $SA = 0.9598$ $SA = 0.9663$

Fig. 2. Segmentation of two-class image (128×128) with close intensities

(a) Original Image   (b) T-ROF (Ite. 11)   (c) Cai [12]

(d) Li [22]   (e) Pock [25]   (f) Yuan [29]   (g) He [20]
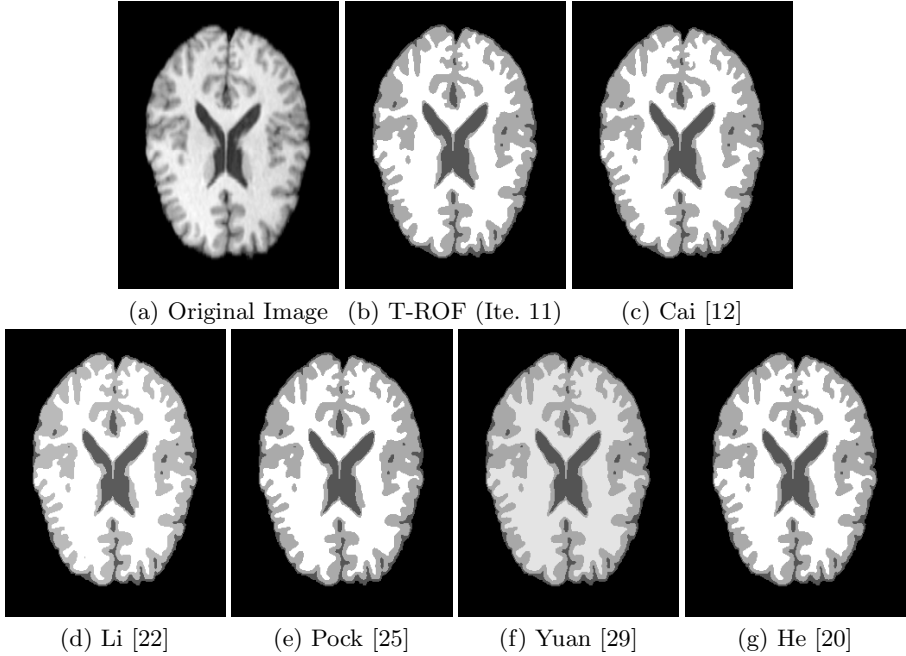
Fig. 3. Four-phase gray and white matter segmentation for a brain MRI image (319×256)
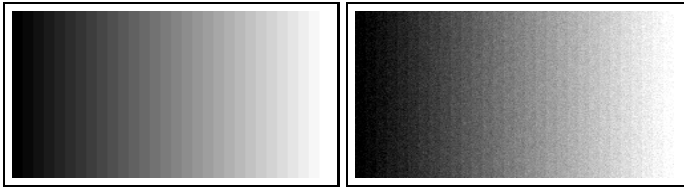

Fig. 4. Stripe image with 30 stripes (140×240) and its noisy version for the segmentation in Tab. 1

computed numerically by an ADMM algorithm with fixed inner parameter 2. Speedups by using more sophisticated methods will be considered in a future paper. The stopping criteria in our T-ROF algorithm for u and $\boldsymbol{\tau}$ are $\left(\|u^{(i)} - u^{(i-1)}\|_2\right)/\|u^{(i)}\|_2 \le \epsilon_u$ and $\|\boldsymbol{\tau}^{(k)} - \boldsymbol{\tau}^{(k-1)}\|_2 \le \epsilon_\tau$, where ϵ_u and ϵ_τ are fixed to 10^{-4} and 10^{-5}, respectively. The initialization of $(\tau_i^{(0)})_{i=1}^{K-1}$ was computed by the fuzzy C-means method [7] with 100 iteration steps.

We will compare our method with the recently proposed multiclass segmentation methods [12,20,22,25,29]. Note that the methods [25,29] work with the fixed fuzzy C-means codebook and we do not update the codebook. Such update is however involved in [20]. The default stopping criterion used in [20,22,25] is the maximum iteration steps; the default stopping criterion used in [29] is 10^{-4} and

Table 1. Parameter μ, iteration steps, CPU time in seconds, and SA for example 4

		Li [22]	Pock [25]	Yuan [29]	He [20]	Cai [12]	T-ROF
Five phases	μ	80	100	10	50	10	8
	Ite.	100	100	87	100	41	84 (4)
	Time	3.87	6.25	4.33	16.75	1.33	1.39
	SA	0.9946	0.9965	0.9867	0.9968	0.9770	0.9986
Ten phases	μ	80	100	10	50	10	8
	Ite.	100	100	102	100	41	84 (5)
	Time	7.71	15.41	9.79	38.52	2.11	2.33
	SA	0.8545	0.9984	0.9715	0.9848	0.8900	0.9967
Fifteen phases	μ	80	100	10	50	10	8
	Ite.	100	100	208	100	41	84 (5)
	Time	11.56	28.21	33.21	63.67	3.06	3.74
	SA	0.7715	0.9993	0.9730	0.9904	0.5280	0.9933

(a) Original image   (b) Noisy image   (c) T-ROF (Ite. 6)   (d) Cai [12]
                                                $SA = 0.9550$ $SA = 0.9232$

(d) Li [22]   (e) Pock [25]   (f) Yuan [29]   (h) He [20]
$SA = 0.4420$ $SA = 0.9485$ $SA = 0.9557$ $SA = 0.9637$
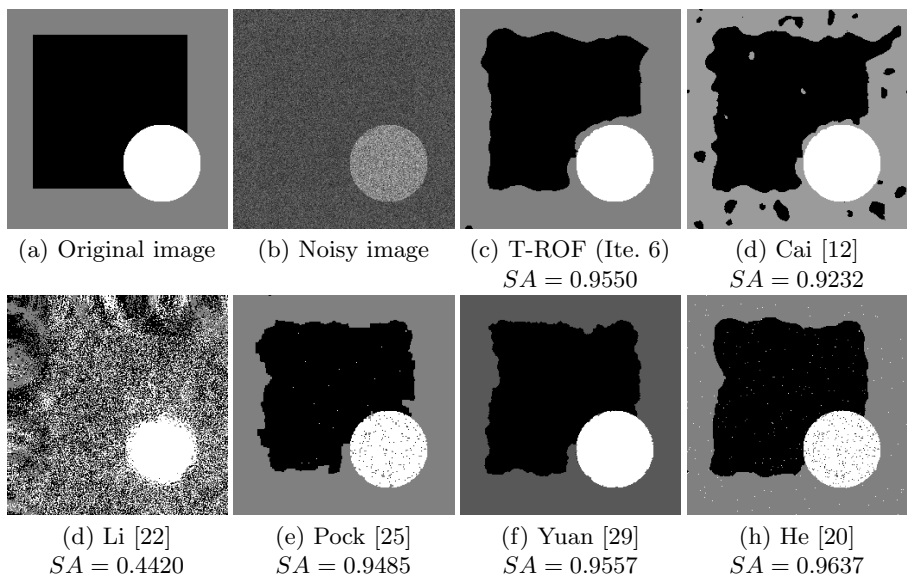
Fig. 5. Segmentation of three-class image (256×256) containing phases with close intensities

maximum iteration steps 300; the default stopping criterion used in [12] is 10^{-4}. We choose the regularization parameter μ of the fidelity term in all the methods by judging the *segmentation accuracy* (SA) defined as

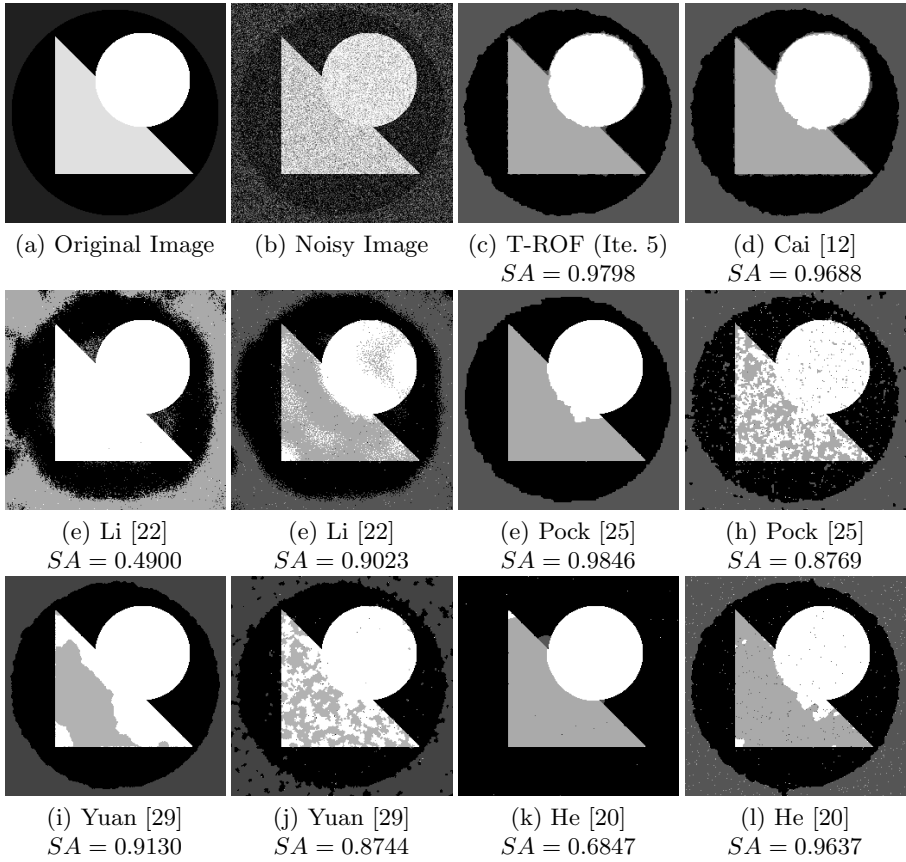$$\text{SA} = \frac{\#\text{correctly classified pixels}}{\#\text{all pixels}}.$$

(a) Original Image　　(b) Noisy Image　　(c) T-ROF (Ite. 5)　　(d) Cai [12]
$SA = 0.9798$　　$SA = 0.9688$

(e) Li [22]　　(e) Li [22]　　(e) Pock [25]　　(h) Pock [25]
$SA = 0.4900$　　$SA = 0.9023$　　$SA = 0.9846$　　$SA = 0.8769$

(i) Yuan [29]　　(j) Yuan [29]　　(k) He [20]　　(l) He [20]
$SA = 0.9130$　　$SA = 0.8744$　　$SA = 0.6847$　　$SA = 0.9637$

Fig. 6. Segmentation of four-class image (256×256) with close gray values

We show the results for two two-class and four multiclass images, where all computations were done on a MacBook with 2.4 GHz processor and 4GB RAM. For further examples we refer to [11].

Example 1 is a two-class cartoon image with some missing pixel values. The segmentation results are shown in Fig. 1. We see that only methods [12,20] and our method gives good results. Indeed a codebook update is required here.

Example 2 is a two-class image with close intensities generated as follows: We have added Gaussian noise with mean 0 and variance 10^{-8} to constant image with gray value 0.5. The noisy image is obtained by keeping the pixel values belonging to the white parts of the original image and reducing the pixel values belonging to the black parts by a factor of 2×10^{-4}. Fig. 2 shows the results of the various algorithms. Except the method [22], all models produce reasonable results.

Example 3 is a four-class gray and white matter segmentation for a brain MRI image from [25]. Fig. 3 gives the results. We can see that, all the methods work well for this kind of image. However our method with 11 τ-value updates is faster

than the other methods, e.g., three times faster than the algorithm of Pock et al. [25] with the assigned parameters.

Example 4 segments the noisy stripe image in Fig. 4 (b), which is generated by imposing Gaussion noise with mean 0 and variance 10^{-3} on the clean image Fig. 4 (a) with 30 stripes. The results for a 5, 10 and 15 class segmentation are listed in Table 1.

Example 5 is a three-class image with close intensities. The test in image Fig. 5 is generated using the same way as those in Example 2 with Gaussian noise of mean 0 and variance 10^{-2}, the scalars used in the black and white parts are 0.1 and 0.6, respectively. For the results of the different methods see Fig. 5.

Example 6 shows a four-class image with close intensities. Fig. 6 (a) and (b) are the original image and the noisy image generated by using Gaussian noise with mean 0 and variance 3×10^{-2}. Figs. 6 (c)–(l) provide two results for each method using different representative parameters μ, where one parameter is optimal with respect to the SA.

References

1. Alter, F., Caselles, V., Chambolle, A.: A characterization of convex calibrable sets in R^N. Mathematische Annalen 332, 329–366 (2005)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, Oxford (2000)
3. Ambrosio, L., Tortorelli, V.: Approximation of functions depending on jumps by elliptic functionals via t-convergence. Communications in Pure and Applied Mathematics 43, 999–1036 (1990)
4. Attouch, H., Buttazzo, G., Michaille, G.: Variational Analysis in Sobolev and BV Spaces. SIAM, Philadelphia (2006)
5. Bar, L., Chan, T., Chung, G., Jung, M., Kiryati, N., Mohieddine, R., Sochen, N., Vese, L.: Mumford and Shah model and its applications to image segmentation and image restoration. In: Handbook of Mathematical Imaging, pp. 1095–1157. Springer (2011)
6. Bellettini, G., Paolini, M., Verdi, C.: Convex approximations of functionals with curvature. Mat. Appl. 2, 297–306 (1991)
7. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences 10, 191–203 (1984)
8. Blake, A., Zisserman: Visual Reconstruction. MIT Press (1987)
9. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J., Osher, S.: Fast global minimization of the active contour/snake model. Journal of Mathematical Imaging and Vision 28, 151–167 (2007)
10. Brown, E., Chan, T., Bresson, X.: Completely convex formulation of the Chan-Vese image segmentation model. International Journal of Computer Vision 98, 103–121 (2012)
11. Cai, X.: A new multiclass image segmentation model by ROF thresholding. Preprint University of Kaiserslautern (2013)
12. Cai, X., Chan, R., Zeng, T.: A two-stage image segmentation method using a convex variant of the Mumford-Shah model and thresholding. SIAM Journal on Imaging Sciences 6, 368–390 (2013)

13. Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision 20, 89–97 (2004)
14. Chambolle, A., Caselles, V., Novaga, M., Cremers, D., Pock, T.: An introduction to total variation for image analysis. HAL Preprint, hal-00437581 (2009)
15. Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics 66, 1632–1648 (2006)
16. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing 10, 266–277 (2001)
17. Dong, B., Chien, A., Shen, Z.: Frame based segmentation for medical images. Commun. Math. Sci. 32, 1724–1739 (2010)
18. Geman, S., Geman, D.: Stochastic relation, Gibbs distributions, and the Baysian resoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(6), 721–741 (1984)
19. Gorski, J., Pfeiffer, F., Klamroth, K.: Biconvex sets and optimization with biconvex functions - a survey and extensions. Mathematical Methods of Optimization Research 66, 373–407 (2007)
20. He, Y., Hussaini, M.Y., Ma, J., Shafei, B., Steidl, G.: A new fuzzy c-means method with total variation regularization for image segmentation of images with noisy and incomplete data. Pattern Recognition 45, 3463–3471 (2012)
21. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. SIAM Journal on Imaging Science (to appear)
22. Li, F., Ng, M., Zeng, T., Shen, C.: A multiphase image segmentation method based on fuzzy region competition. SIAM Journal on Imaging Sciences 3, 277–299 (2010)
23. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics, 577–685 (1989)
24. Nikolova, M., Esedoglu, S., Chan, T.F.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics 66, 1632–1648 (2006)
25. Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach for computing minimal partitions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 810–817 (2009)
26. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In: ICCV 2009 (2009)
27. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992)
28. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. International Journal of Computer Vision 50, 271–293 (2002)
29. Yuan, J., Bae, E., Tai, X.-C., Boykov, Y.: A continuous max-flow approach to Potts model. In: European Conference on Computer Vision, pp. 379–392 (2010)
30. Zach, C., Gallup, D., Frahm, J., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: Vision, Modeling, and Visualization Workshop (2008)
31. Zhang, Y., Matuszewski, B., Shark, L., Moore, C.: Medical image segmentation using new hybrid level-set method. In: Fifth International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics, 71–76 (2008)