A Multiphase Image Segmentation Based on Fuzzy Membership Functions and L1-norm Fidelity

**Fang Li · Stanley Osher · Jing Qin ·
Ming Yan**

Abstract In this paper, we propose a variational multiphase image segmentation model based on fuzzy membership functions and L1-norm fidelity. Then we apply the alternating direction method of multipliers to solve an equivalent problem. All the subproblems can be solved efficiently. Specifically, we propose a fast method to calculate the fuzzy median. Experimental results and comparisons show that the L1-norm based method is more robust to outliers such as impulse noise and keeps better contrast than its L2-norm counterpart. Theoretically, we prove the existence of the minimizer and analyze the convergence of the algorithm.

1 Introduction

As a fundamental step in image processing, image segmentation partitions an image into several disjoint regions such that pixels in the same region share some uniform characteristics such as intensity, color, and texture. During the last two decades, image segmentation methods using variational methods and

F. Li (✉)

Department of Mathematics, East China Normal University, and Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice, Shanghai, China.
E-mail: e-mail: fli@math.ecnu.edu.cn

S. Osher, J. Qin

Department of Mathematics, University of California, Los Angeles, CA 90095, USA.
E-mail: sjo@math.ucla.edu,jxq@ucla.edu

and M. Yan

Department of Computational Mathematics, Science and Engineering and Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA.
E-mail: e-mail:yanm@math.msu.edu

partial differential equations are very popular due to their flexibility in problem modeling and algorithm design. There are two key ingredients of variational segmentation models. One is how to describe the regions or boundaries between these regions, and the other is how to model the noise and describe the uniform characteristics of each region.

The Mumford-Shah model [35], a well-known variational segmentation model, penalizes the combination of the total length of the segmentation boundaries and the L2-norm error of approximating the observed image with an unknown piecewise smooth approximation. In other words, the Mumford-Shah model seeks an optimal piecewise smooth function with smooth boundaries to approximate the observed image.

However, the Mumford-Shah model is hard to implement in practice because the discretization of the unknown set of boundaries is very complex. Therefore, a parametric/explicit active contour method is used to represent the segmentation boundaries [49]. In addition, the special Mumford-Shah model with a piecewise constant approximation is studied by Chan and Vese [15], and the level set method [37] is applied to solve this problem. Using an implicit representation of boundaries, the level set method has many advantages over methods using explicit representations of boundaries. For instance, the level set method handles the topological change of zero level set automatically [8, 2,22]. Both the parametric/explicit active contour method and the level set method assume that each pixel belongs to a unique region. An alternative way to represent various regions is to use fuzzy membership functions [14,6, 34,25], which describe the levels of possible membership. Fuzzy membership functions assume that each pixel can be in several regions simultaneously with probability in [0,1]. These probabilities satisfy two constraints: (i) nonnegativity constraint, i.e., the membership functions are nonnegative at all pixels; (ii) sum-to-one constraint, i.e., the sum of all the membership functions at each pixel equals one. Then the length of boundaries can be approximated by the Total Variation (TV) of fuzzy membership functions. The main advantages of using fuzzy membership functions over other methods include: i) the energy functional is convex with respect to fuzzy membership functions, guaranteeing the convergence and stability of many numerical optimization methods. ii) fuzzy membership function has a larger feasible set, and it is possible to find better segmentation results.

For two-phase segmentation, where there are only two regions, we only need one level set function or one fuzzy membership function. Multiphase segmentation is more challenging than two-phase segmentation since more variables and constraints are involved in representing multiple regions and their boundaries effectively. The two-phase Chan-Vese model [15] has been generalized to multiphase segmentation by using multiple level set functions to represent multiple regions [44]. Partitioning an image into N disjoint regions requires $\log_2 N$ level set functions. The advantage of using multiple level set functions is that it automatically avoids the problems of vacuum and overlap of regions. However, the implementation is not easy, and special numerical schemes are needed to ensure stability [32,33,42,27]. For fuzzy membership functions, the sum-to-one

constraint is not satisfied automatically in multiphase segmentation. However, this constraint is easy to deal with in many cases, e.g., Fuzzy C-Mean (FCM) and its variants have closed-form solutions for the membership functions and are widely used in data mining and medical image segmentation [4,1,38,16, 29,31,30,24,7]. Variable splitting schemes are used in both [30] and [31] to get efficient numerical algorithms. The Alternating Direction Method of Multipliers (ADMM) method is used in [24] to derive the algorithm with two sets of extra variables. Primal-dual type algorithm is derived in [10] to solve the TV regularized FCM segmentation model. Both [24] and [10] use projection to simple to handle the constraints of membership functions. Other segmentation approaches include a convex approach [9], two-stage methods [11,43], one single level set function approach [18], et.al.

Noise is unavoidable in images, and it is important to develop segmentation methods that work on noisy images. Among many types of noise, the Gaussian white noise is frequently assumed, and the L2-norm fidelity is adopted. However, when images are corrupted by non-Gaussian noise, in order to obtain a faithful segmentation, one has to model the noise according to its specific type [40,36,12,23,47,48,7]. Particularly, the L1-norm fidelity is used for both salt-and-pepper impulse noise and random-valued impulse noise in image processing [36,12,23]. In addition, it is robust to outliers and able to preserve contrast because the denoising process of L1-norm models is determined by the geometry such as area and length rather than the contrast in the L2-norm case [13].

Inspired by the fact that L1-norm is more robust to impulse noise and outliers and can better preserve contrast, in this paper, we propose a variational *multiphase* fuzzy segmentation model based on *L1-norm fidelity* and *fuzzy membership functions*. This model can also deal with missing data in images. When there are missing pixels in an image, we randomly assign 0 or 255 at these pixels by considering these pixels as corrupted by salt-and-pepper impulse noise. ADMM [20,19], which was rediscovered as split Bregman [21] and found to be very useful for L1 and TV type optimization problems, is applied to solve this nonconvex optimization problem. By introducing two sets of auxiliary variables, we derive an efficient algorithm with all the subproblems having closed-form solutions. In the theoretical aspect, we prove the existence of the minimizer and analyze the convergence of the algorithm. We note that the proposed method is closely related to the method in [24] since both methods use TV regularization and ADMM algorithm. The difference is that L1-norm fidelity is considered in the proposed method, while L2-norm fidelity is used in [24].

The outline of this paper is as follows. In Section 2, we review some closely related existing works. In Section 3, the proposed model is described in detail, and the existence of the minimizer to the model is proved. In Section 4, a numerical algorithm is derived, and its convergence analysis is presented. In Section 5, experimental results and comparisons are presented to illustrate the effectiveness of the proposed method. Finally, the paper is concluded in Section 6.

2 Related works

Let $\Omega \subset \mathbb{R}^2$ be a bounded open subset with Lipschitz boundary, and let $I : \Omega \to \mathbb{R}^s$ be the given clean or noisy image. Let $s = 1$ for grayscale images and $s = 3$ for color images. Our goal is to find an N-phase "optimal" partition $\{\Omega_i\}_{i=1}^N$ such that $\Omega_i \bigcap \Omega_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^N \Omega_i = \Omega$. Define the set of N-phase fuzzy membership functions as

$$\Delta := \left\{ (u_1, ..., u_N) \Big| u_i \in BV(\Omega), u_i(x) \geq 0, \sum_{i=1}^N u_i(x) = 1, \forall x \in \Omega \right\},$$

where $BV(\Omega)$ is the space of functions with bounded variation [2]. The closely related works are listed in the following and will be compared with our proposed method in Section 5. For the sake of simplicity, we use the notations $\mathbf{U} = (u_1, \cdots, u_N)$ and $\mathbf{C} = (c_1, \cdots, c_N) \in \mathbb{R}^{sN}$, where $c_i \in \mathbb{R}$ for grayscale images and $c_i \in \mathbb{R}^3$ for color images.

– FCM [4]– Fuzzy c-means clustering method that solves

$$\min_{(\mathbf{U},\mathbf{C}) \in \Delta \times \mathbb{R}^{sN}} \sum_{i=1}^N \int_\Omega |I(x) - c_i|^2 \, u_i^p(x) \, dx$$

  using the alternating minimization algorithm. Though p can be any number no smaller than 1, it is commonly set to 2.
– FCM_S2 [16] – A variant of FCM that solves

$$\min_{(\mathbf{U},\mathbf{C}) \in \Delta \times \mathbb{R}^{sN}} \sum_{i=1}^N \left\{ \int_\Omega |I(x) - c_i|^2 \, u_i^p(x) \, dx + \alpha \int_\Omega \left| \bar{I}(x) - c_i \right|^2 u_i^p(x) \, dx \right\},$$

  where $\bar{I}$ is obtained by applying the median filter on I and $\alpha > 0$ is a weight parameter. It can also be solved by the alternating minimization algorithm, and it is more robust to impulse noise than FCM.
– FLICM [29] – Fuzzy Local Information C-Means clustering method that solves

$$\min_{(\mathbf{U},\mathbf{C}) \in \Delta \times \mathbb{R}^{sN}} \sum_{i=1}^N \left\{ \int_\Omega |I(x) - c_i|^2 \, u_i^p(x) \, dx \right.$$

$$\left. + \alpha \int_\Omega \int_{y \in \mathcal{N}(x)} (1 - u_i(y))^p |I(y) - c_i|^2 \, u_i^p(x) dy dx \right\},$$

  where $\mathcal{N}(x)$ is a neighborhood of x. FLICM is more robust to both Gaussian noise and impulse noise than FCM.
– L2FS [24] – L2-norm fidelity based Fuzzy Segmentation method that solves

$$\min_{(\mathbf{U},\mathbf{C}) \in \Delta \times \mathbb{R}^{sN}} \sum_{i=1}^N \left\{ \int_\Omega \|\nabla u_i(x)\| \, dx + \lambda \int_\Omega |I(x) - c_i|^2 \, u_i(x) \, dx \right\}, \quad (1)$$

by ADMM. Here $\lambda > 0$ is a parameter and $\int_{\Omega} \|\nabla u_i(x)\| \, dx$ denotes the TV of u_i with $\|\nabla u_i(x)\| := \sqrt{(\nabla_{x_1} u_i(x))^2 + (\nabla_{x_2} u_i(x))^2}$ for $x = (x_1, x_2)$. For fixed $\mathbf{C}$, [24] is related to the popular TV denoising method [39]. Note that the similar model is solved by other fast numerical methods in [30].

– L1SS [26] – L1-norm fidelity based Soft Segmentation method, in which $\log_2 N$ soft functions are introduced to represent N phases. Since the model for multiphase segmentation is complicated for more than four phases, we show the four-phase model as follows:

$$\min_{u_1, u_2 \in [0,1], \mathbf{C} \in \mathbb{R}^{sN}} \left\{ \sum_{i=1}^{2} \int_{\Omega} \|\nabla u_i(x)\| \, dx + \lambda \sum_{j=1}^{4} \int_{\Omega} |I(x) - c_j| \, M_j(x) dx \right\},$$

where the membership functions M_j, $j = 1, \cdots, 4$, are represented by soft functions $u_1(x), u_2(x) \in [0, 1]$ in the following way:

$$M_1(x) = u_1(x)u_2(x), \qquad M_2(x) = u_1(x)(1 - u_2(x)),$$
$$M_3(x) = (1 - u_1(x))u_2(x), \qquad M_4(x) = (1 - u_1(x))(1 - u_2(x)).$$

– L2L0 [43] – L2-norm fidelity and L0-norm regularization based image partition model:

$$\min_u \|\nabla u\|_0 + \lambda \|u - I\|_2^2.$$

This model seeks a piecewise constant approximation of the original image I. Since this model can not specify the number of classes, a second step is applied to combine some classes if this model returns more classes than required. Here we apply FCM on the piecewise constant approximation u to obtain the final segmentation result.

Remark: There are two advantages of our proposed method over L1SS. Firstly, we use fuzzy membership functions to represent regions, where N fuzzy membership functions are required for an N-phase segmentation. Hence, the solution space is much larger than L1SS, which ensures the higher possibility to obtain optimal segmentation. Secondly, the proposed method can take use of other commonly used segmentation methods such as FCM to gain good initialization of fuzzy membership functions. Multiphase segmentation is sensitive to initialization, and a good initialization is very important for a successful segmentation. However, it is hard to use the existing segmentation methods to get a good initialization for soft membership functions in L1SS.

3 The proposed model

In this paper, we assume that the given image can be approximated by a piecewise constant function, i.e.,

$$I(x) = \sum_{i=1}^{N} c_i \chi_{\Omega_i}(x) + n(x).$$

Here χ_{Ω_i} denotes the indicator function of region Ω_i, i.e.,

$$\chi_{\Omega_i}(x) = \begin{cases} 1, & \text{if } x \in \Omega_i; \\ 0, & \text{otherwise}, \end{cases}$$

c_i is a constant that represents the given data in region Ω_i, and n can be outliers, impulse noise or other mixture types rather than Gaussian noise.

Instead of using the L2-norm fidelity to measure the approximation error when the noise is the Gaussian white noise, we use the L1-norm fidelity. Same as in the Mumford-Shah model, we require the segmentation boundaries to be smooth. Then we have the following model

$$\min_{\{\Omega_i\},\mathbf{C}} \sum_{i=1}^{N} \int_{\Omega} \|\nabla \chi_{\Omega_i}(x)\| dx + \lambda \int_{\Omega} \left| I(x) - \sum_{i=1}^{N} c_i \chi_{\Omega_i}(x) \right| dx, \qquad (2)$$

where $\lambda > 0$ is a tuning parameter. Note that the TV of χ_{Ω_i} in the first term is equal to the length of boundary $\partial \Omega_i$. An equivalent form of model (2) is

$$\min_{\{\Omega_i\},\mathbf{C}} \sum_{i=1}^{N} \left\{ \int_{\Omega} \|\nabla \chi_{\Omega_i}(x)\| dx + \lambda \int_{\Omega} |I(x) - c_i| \chi_{\Omega_i}(x) dx \right\}. \qquad (3)$$

Because χ_{Ω_i} can take values 0 and 1 only and $\{\chi_{\Omega_i}\}$ is a partition, $(\chi_{\Omega_1}, \ldots, \chi_{\Omega_N})$ belongs to the set

$$\Delta_0 := \left\{ (u_1, ..., u_N) \big| u_i \in BV(\Omega), u_i(x) \in \{0,1\}, \sum_{i=1}^{N} u_i(x) = 1, \forall x \in \Omega \right\}.$$

At any point $x \in \Omega$, there is only one function having value 1, and all the other functions have value 0. Thus set Δ_0 is not continuous, which leads to difficulties and instabilities in numerical implementations. However, we can relax binary indicator functions $\{\chi_{\Omega_i}\}$ to fuzzy membership functions $\{u_i\}$, which satisfy the nonnegativity constraint and the sum-to-one constraint, i.e., $(u_1, ..., u_N)$ belongs to the set Δ defined in (2). It is obvious that $u_i(x) \in [0, 1]$ and Δ is a simplex at any $x \in \Omega$. So $u_i(x)$ can be understood as the probability of pixel x to be in the ith class. Then we can rewrite our model (3) as

$$\min_{(\mathbf{U},\mathbf{C}) \in \Delta \times \mathbb{R}^{sN}} E(\mathbf{U},\mathbf{C}) = \sum_{i=1}^{N} \left\{ \int_{\Omega} \|\nabla u_i(x)\| \, dx + \lambda \int_{\Omega} |I(x) - c_i| \, u_i(x) \, dx \right\}. \qquad (4)$$

Note that model (4) is convex with respect to $\mathbf{U}$ and $\mathbf{C}$ separately but not jointly. The difference between (4) and (1) is that the L2 fidelity term in (1) is replaced by the L1 fidelity term. The existence of a minimizer for $E(\mathbf{U},\mathbf{C})$ in (4) is stated in Theorem 1.

Theorem 1 (Existence of a minimizer) *Given an image $I \in L^{\infty}(\Omega)$, there exists a minimizer of $E(\mathbf{U},\mathbf{C})$ in $\Delta \times \mathbb{R}^{sN}$ for any parameter $\lambda > 0$.*

Proof Since $E(\mathbf{U}, \mathbf{C})$ is positive and proper, the infimum of $E(\mathbf{U}, \mathbf{C})$ must be finite. Let us pick a minimizing sequence $(\mathbf{U}^n, \mathbf{C}^n) \in \Delta \times \mathbb{R}^{sN}$ that satisfies $E(\mathbf{U}^n, \mathbf{C}^n) \to \inf_{\mathbf{U}, \mathbf{C}} E(\mathbf{U}, \mathbf{C})$ as $n \to \infty$. Then there exists a constant $M > 0$ such that

$$E(\mathbf{U}^n, \mathbf{C}^n) = \sum_{i=1}^{N} \left\{ \int_{\Omega} \|\nabla u_i^n(x)\| \, dx + \lambda \int_{\Omega} |I(x) - c_i^n| u_i^n(x) \, dx \right\} \le M.$$

Then each term in $E(\mathbf{U}^n, \mathbf{C}^n)$ is bounded, i.e., for each $i = 1, \cdots, N$,

$$\int_{\Omega} \|\nabla u_i^n(x)\| dx \le M, \quad \int_{\Omega} |I(x) - c_i| u_i^n(x) dx \le M. \tag{5}$$

Since $u_i^n(x) \in [0,1]$, we have $\int_{\Omega} u_i^n(x) dx \le |\Omega|$, where $|\Omega|$ is the area of Ω. Together with the first equality in (5), we have that u_i^n is uniformly bounded in $BV(\Omega)$ for all $i = 1, ..., N$. By the compactness property of $BV(\Omega)$ and the relative compactness of $BV(\Omega)$ in $L^1(\Omega)$, up to a subsequence also denoted by $\{u_i\}^n$ after relabeling, there exists a function $u_i^* \in BV(\Omega)$ such that

$$u_i^n \to u_i^* \quad \text{strongly in } L^1(\Omega),$$
$$u_i^n \to u_i^* \quad \text{a.e. } x \in \Omega,$$
$$\nabla u_i^n \rightharpoonup \nabla u_i^* \quad \text{in the sense of measure.}$$

Then by the lower semicontinuity of the TV semi-norm,

$$\int_{\Omega} \|\nabla u_i^*(x)\| dx \le \liminf_{n \to \infty} \int_{\Omega} \|\nabla u_i^n(x)\| dx. \tag{6}$$

Meanwhile since $\mathbf{U}^n = (u_1^n, ..., u_N^n) \in \Delta$, we have $\mathbf{U}^* = (u_1^*, ..., u_N^*) \in \Delta$.

It is easy to derive the first order optimality condition about c_i^n, which is

$$0 \in \int_{\Omega} \partial |I(x) - c_i^n| u_i^n(x) dx,$$

where $\partial |\cdot|$ is the subdifferential of $|\cdot|$. Since $u_i^n(x) \ge 0$ and $\int_{\Omega} u_i^n(x) dx > 0$, the above equation implies that the constant c_i^n satisfies

$$|c_i^n| \le \|I\|_{\infty}.$$

By the boundedness of sequence $\{c_i^n\}$, we can extract a subsequence also denoted by $\{c_i^n\}$ such that for some constant c_i^*

$$c_i^n \to c_i^*, \text{ as } n \to \infty.$$

Finally, since $u_i^n(x) \to u_i^*(x)$, a.e. $x \in \Omega$ and $c_i^n \to c_i^*$ as $n \to \infty$, Fatou's lemma yields

$$\int_{\Omega} |I(x) - c_i^*| u_i^*(x) dx \le \liminf_{n \to \infty} \int_{\Omega} |I(x) - c_i^n| u_i^n(x) dx. \tag{7}$$

Combining inequalities (6) and (7) for all i, on a suitable subsequence, we established

$$E(\mathbf{U}^*, \mathbf{C}^*) \leq \liminf_{n \to \infty} E(\mathbf{U}^n, \mathbf{C}^n) = \inf_{\mathbf{U}, \mathbf{C}} E(\mathbf{U}, \mathbf{C}),$$

and hence $(\mathbf{U}^*, \mathbf{C}^*)$ must be a minimizer of the energy E. This completes the proof. □

The minimizer of $E(\mathbf{U}, \mathbf{C})$ is not unique due to the following hidden symmetry property. Denote S_N as the permutation group of $\{1, ..., N\}$, i.e., each permutation $\sigma \in S_N$ is defined as a one-to-one map $\sigma : \{1, ..., N\} \to \{1, ..., N\}$ such that $\{\sigma(1), ..., \sigma(N)\}$ is a rearrangement of $\{1, ..., N\}$. Denote $\mathbf{U}_\sigma = (u_{\sigma(1)}, ..., u_{\sigma(N)})$, $\mathbf{C}_\sigma = (c_{\sigma(1)}, ..., c_{\sigma(N)})$. It is straightforward to show the following theorem.

Theorem 2 (Symmetry of minimizer) *For any* $(\mathbf{U}, \mathbf{C}) \in \Delta \times \mathbb{R}^N$ *and any* $\sigma \in S_N$,

$$E(\mathbf{U}_\sigma, \mathbf{C}_\sigma) = E(\mathbf{U}, \mathbf{C}).$$

In particular, suppose that $(\mathbf{U}^*, \mathbf{C}^*)$ *is a minimizer of* $E(\mathbf{U}, \mathbf{C})$, *i.e.,*

$$(\mathbf{U}^*, \mathbf{C}^*) = \underset{(\mathbf{U}, \mathbf{C}) \in \Delta \times \mathbb{R}^N}{\arg\min} E(\mathbf{U}, \mathbf{C}).$$

Then, for any $\sigma \in S_N$, *we have*

$$(\mathbf{U}_\sigma^*, \mathbf{C}_\sigma^*) = \underset{(\mathbf{U}, \mathbf{C}) \in \Delta \times \mathbb{R}^N}{\arg\min} E(\mathbf{U}, \mathbf{C}).$$

4 The numerical algorithm and its convergence analysis

In this section, we provide an efficient algorithm based on ADMM and discuss its convergence.

4.1 The algorithm

ADMM is applied, in this subsection, to solve the proposed fuzzy segmentation model (4). We introduce two sets of auxiliary variables $\mathbf{D} = (d_1, ..., d_N)$, $\mathbf{W} = (w_1, ..., w_N)$ such that $\nabla u_i = d_i, u_i = w_i$. Then the model (4) is equivalent to the following minimization problem with equality constraints:

$$\min_{\mathbf{D}, \mathbf{W}, \mathbf{C}, \mathbf{U}} \sum_{i=1}^{N} \left\{ \int_\Omega \|d_i(x)\| dx + \lambda \int_\Omega |I(x) - c_i| \, w_i(x) \, dx \right\} + \delta_\Delta(\mathbf{W}), \quad (8)$$
$$\text{subject to } \nabla u_i = d_i, u_i = w_i, \forall i = 1, \ldots, N,$$

where δ_Δ is the indicator function of the set Δ, i.e.,

$$\delta_\Delta(\mathbf{W}) = \begin{cases} 0, & \text{if } \mathbf{W} \in \Delta, \\ +\infty, & \text{otherwise.} \end{cases}$$

The augmented Lagrangian function for problem (8) is:

$$
\begin{aligned}
& \mathscr{L}(\mathbf{D}, \mathbf{W}, \mathbf{C}, \mathbf{U}; \mathbf{\Lambda_D}, \mathbf{\Lambda_W}) \\
& = \sum_{i=1}^{N} \left\{ \int_{\Omega} \|d_i(x)\| \, dx + \lambda \int_{\Omega} |I(x) - c_i| \, w_i(x) \, dx \right\} + \delta_{\Delta}(\mathbf{W}) \\
& + \sum_{i=1}^{N} \left\{ \langle \lambda_{d_i}, \nabla u_i - d_i \rangle + \frac{r}{2} \int_{\Omega} \|\nabla u_i(x) - d_i(x)\|^2 \, dx \right\} \\
& + \sum_{i=1}^{N} \left\{ \langle \lambda_{w_i}, u_i - w_i \rangle + \frac{r}{2} \int_{\Omega} |u_i(x) - w_i(x)|^2 \, dx \right\},
\end{aligned}
$$

where $\Lambda_{\mathbf{D}} = (\lambda_{d_1}, ..., \lambda_{d_N})$, $\Lambda_{\mathbf{W}} = (\lambda_{w_1}, ..., \lambda_{w_N})$ are the Lagrangian multipliers and r is a positive constant. Here $\langle \lambda_{d_i}, \nabla u_i - d_i \rangle = \int_{\Omega} \lambda_{d_i}^T(x)(\nabla u_i(x) - d_i(x)) dx$, and $\langle \lambda_{w_i}, u_i - w_i \rangle = \int_{\Omega} \lambda_{w_i}(x)(u_i(x) - w_i(x)) dx$.

The ADMM solves the primal variables one by one in the Gauss-Seidel style and then updates the dual variables (Lagrangian multipliers). It can be summarized as follows:

$$
\begin{aligned}
\mathbf{D}^{k+1} &= \arg\min_{\mathbf{D}} \mathscr{L}(\mathbf{D}, \mathbf{W}^k, \mathbf{C}^k, \mathbf{U}^k; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k), \\
\mathbf{W}^{k+1} &= \arg\min_{\mathbf{W}} \mathscr{L}(\mathbf{D}^{k+1}, \mathbf{W}, \mathbf{C}^k, \mathbf{U}^k; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k), \\
\mathbf{C}^{k+1} &= \arg\min_{\mathbf{C}} \mathscr{L}(\mathbf{D}^{k+1}, \mathbf{W}^{k+1}, \mathbf{C}, \mathbf{U}^k; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k), \\
\mathbf{U}^{k+1} &= \arg\min_{\mathbf{U}} \mathscr{L}(\mathbf{D}^{k+1}, \mathbf{W}^{k+1}, \mathbf{C}^{k+1}, \mathbf{U}; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k), \\
\lambda_{d_i}^{k+1} &= \lambda_{d_i}^k + r \left(\nabla u_i^{k+1} - d_i^{k+1} \right), \\
\lambda_{w_i}^{k+1} &= \lambda_{w_i}^k + r \left(u_i^{k+1} - w_i^{k+1} \right).
\end{aligned}
$$

In the following, we show how to solve each subproblem and then give the algorithm.

D-subproblem

The subproblem for $\mathbf{D}$ is equivalent to

$$
\mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \sum_{i=1}^{N} \left\{ \int_{\Omega} \|d_i(x)\| dx + \frac{r}{2} \int_{\Omega} \left\| d_i(x) - \nabla u_i^k(x) - \frac{\lambda_{d_i}^k(x)}{r} \right\|^2 dx \right\}.
$$

This is separable and the optimal solution of d_i^{k+1} can be explicitly expressed using shrinkage operators. We simply compute

$$
d_i^{k+1}(x) = \mathcal{S} \left(\nabla u_i^k(x) + \frac{\lambda_{d_i}^k(x)}{r}, \frac{1}{r} \right),
$$

where $\mathcal{S}$ is the shrinkage operator defined as

$$\mathcal{S}(v, \tau) := \frac{v}{\|v\|} * \max\left(\|v\| - \tau, 0\right).$$

For the sake of simplicity, we denote this step as

$$\mathbf{D}^{k+1} = \mathcal{S}\left(\nabla \mathbf{U}^k + \frac{\Lambda_{\mathbf{D}}^k}{r}, \frac{1}{r}\right).$$

W-subproblem

The subproblem for $\mathbf{W}$ is equivalent to

$$\min_{\mathbf{W}} \sum_{i=1}^{N} \left\{ \frac{r}{2} \int_{\Omega} \left| w_i(x) - u_i^k(x) - \frac{\lambda_{w_i}^k(x)}{r} + \frac{\lambda |I(x) - c_i^k|}{r} \right|^2 dx \right\} + \delta_{\Delta}(\mathbf{W}).$$

Since Δ is a convex simplex at any $x \in \Omega$, the solution is given by

$$\mathbf{W}^{k+1} = \Pi_{\Delta}\left(\left[u_i^k + \frac{\lambda_{w_i}^k}{r} - \frac{\lambda |I - c_i^k|}{r}\right]_{i=1}^{N}\right),$$

where Π_{Δ} is the projection onto the simplex Δ, for which more details can be found in [17]. We denote the step as

$$\mathbf{W}^{k+1} = \Pi_{\Delta}\left(\mathbf{U}^k + \frac{\Lambda_{\mathbf{W}}^k}{r} - \lambda \frac{|I - \mathbf{C}^k|}{r}\right).$$

C-subproblem

The subproblem for $\mathbf{C}$ is

$$\mathbf{C}^{k+1} = \arg\min_{\mathbf{C}} \sum_{i=1}^{N} \int_{\Omega} |I(x) - c_i| \, w_i^{k+1}(x) \, dx.$$

It is separable, and c_i^{k+1} can be solved independently. The optimality condition for each c_i^{k+1} is

$$0 \in -\int_{\Omega} \partial |I(x) - c_i^{k+1}| w_i^{k+1}(x) dx. \tag{9}$$

Because the right-hand side of (9) is maximal monotone [3], the bisection method and ADMM are applied to solve it [28, 26]. The next lemma provides a new way to find an optimal solution for discrete cases.

Lemma 1 *Given a finite non-decreasing sequence $\{I_{[j]}\}_{j=1}^{n}$, i.e.,*

$$I_{[1]} \leq I_{[2]} \leq ... \leq I_{[n]},$$

and a non-negative sequence $\{w_{[j]}\}_{j=1}^{n}$ with $A = \sum_{j=1}^{n} w_{[j]} > 0$, the optimal solution set for

$$\min_{c} \sum_{j=1}^{n} \left|I_{[j]} - c\right| w_{[j]}, \tag{10}$$

is $\left[I_{[j^*]}, I_{[j_*+1]}\right]$, where j^* and j_* satisfy

$$A - 2\sum_{j=1}^{j^*} w_{[j]} \leq 0 < A - 2\sum_{j=1}^{j^*-1} w_{[j]},$$

$$A - 2\sum_{j=1}^{j_*+1} w_{[j]} < 0 \leq A - 2\sum_{j=1}^{j_*} w_{[j]}.$$

The fuzzy median of $\{I_j\}_{j=1}^{n}$ with respect to the weight $\{w_{[j]}\}_{j=1}^{n}$ [28], which is defined as $(I_{[j^*]} + I_{[j_*+1]})/2$, is an optimal solution. If, in addition, there exists j^* such that

$$A - 2\sum_{j=1}^{j^*} w_{[j]} < 0 < A - 2\sum_{j=1}^{j^*-1} w_{[j]},$$

then $j_* = j^* - 1$, and the optimal solution is unique.

Proof The optimality condition of (10) is

$$0 \in h(c) := \sum_{j=1}^{n} \partial |I_{[j]} - c| w_{[j]}.$$

We can see that $h(c)$ is non-increasing and it can be expressed as

$$h(c) = \begin{cases} A, & c < I_{[1]}, \\ [A - 2w_{[1]}, A], & c = I_{[1]}, \\ A - 2w_{[1]}, & c \in (I_{[1]}, I_{[2]}), \\ \dots \\ [A - 2\sum_{j=1}^{s} w_{[j]}, A - 2\sum_{j=1}^{s-1} w_{[j]}], & c = I_{[s]}, \\ A - 2\sum_{j=1}^{s} w_{[j]}, & c_i \in (I_{[s]}, I_{[s+1]}), \\ \dots \\ -A, & c > I_{[n]}. \end{cases}$$

The range of h is $[-A, A]$, and $h(c)$ can take multiple values when $c = I_{[j]}$ for any j with $w_{[j]} \neq 0$. Therefore, we can always find j^* such that

$$A - 2\sum_{j=1}^{j^*} w_{[j]} \leq 0 < A - 2\sum_{j=1}^{j^*-1} w_{[j]}.$$

Thus $0 \in h(I_{[j^*]}) = \left[A - 2\sum_{j=1}^{j^*} w_{[j]}, A - 2\sum_{j=1}^{j^*-1} w_{[j]}\right]$, and $I_{[j^*]}$ is an optimal solution. In addition, we have that $h(c) > 0$ when $c < I_{[j^*]}$. Similarly, we can find j_* such that

$$A - 2\sum_{j=1}^{j_*+1} w_{[j]} < 0 \leq A - 2\sum_{j=1}^{j_*} w_{[j]},$$

and $I_{[j_*+1]}$ is an optimal solution. In addition, we have that $h(c) < 0$ when $c > I_{[j_*+1]}$. Then $h(c)$ being non-increasing gives us that the set of optimal solutions for (10) is $\left[I_{[j^*]}, I_{[j_*+1]}\right]$. When $j^* = j_* + 1$, the optimal solution is unique. $\qquad\square$

Remark: When there are missing pixels in images, we can put a mask on the data fidelity term as in image inpainting problems [41] or assign a value to each missing pixel. In [24], the missing pixels are assigned with zero, and **C** changes based on the percentage of missing pixels. While, this lemma tells us that assigning 0 or 255 (for grayscale images) randomly to these missing pixels will not change the value of c with a high probability because c_i is a weighted median. Also, by assigning 0 or 255 randomly, this algorithm is able to deal with cases where more than half of the pixels are missing. See the numerical experiments in Section 5.

Assume that I and w_i^{k+1} are vectors in $\mathbb{R}^n$, where n is the total number of pixels. We can reorder the components of I and w_i^{k+1} such that the monotonicity of I in Lemma 1 is satisfied. If there are multiple optimal solutions for c_i, we choose the smallest one. We denote this step as

$$\mathbf{C}^{k+1} = \psi(\mathbf{W}^{k+1}).$$

U-subproblem

The subproblem for **U** is equivalent to

$$\mathbf{U}^{k+1} = \arg\min_{\mathbf{U}} \sum_{i=1}^{N} \int_{\Omega} \left\|\nabla u_i(x) - d_i^{k+1}(x) + \frac{\lambda_{d_i}^k(x)}{r}\right\|^2 dx$$
$$+ \int_{\Omega} \left|u_i(x) - w_i^{k+1}(x) + \frac{\lambda_{w_i}^k(x)}{r}\right|^2 dx.$$

It is separable for u_i^{k+1}, and, from the following first order optimality condition for each u_i^{k+1}:

$$\nabla^T \left(\nabla u_i^{k+1}(x) - d_i^{k+1}(x) + \frac{\lambda_{d_i}^k(x)}{r}\right) + \left(u_i^{k+1}(x) - w_i^{k+1}(x) + \frac{\lambda_{w_i}^k(x)}{r}\right) = 0,$$

we can derive the closed-form solution of u_i^{k+1} from the equation:

$$\left(\nabla^T \nabla + \mathbf{I}\right) u_i^{k+1}(x) = \nabla^T d_i^{k+1}(x) + w_i^{k+1}(x) - \frac{\nabla^T \lambda_{d_i}^k(x)}{r} - \frac{\lambda_{w_i}^k(x)}{r}.$$

A diagonalization trick can be applied to find u_i^{k+1} efficiently [45]. We denote this step as

$$\mathbf{U}^{k+1} = \left(\nabla^T \nabla + \mathbf{I}\right)^{-1} \left(\nabla^T \mathbf{D}^{k+1} + \mathbf{W}^{k+1} - \frac{\nabla^T \Lambda_{\mathbf{D}}^k}{r} - \frac{\Lambda_{\mathbf{W}}^k}{r}\right).$$

Updating dual variables. We denote the steps as

$$\Lambda_{\mathbf{D}}^{k+1} = \Lambda_{\mathbf{D}}^k + r \left(\nabla \mathbf{U}^{k+1} - \mathbf{D}^{k+1}\right),$$
$$\Lambda_{\mathbf{W}}^{k+1} = \Lambda_{\mathbf{W}}^k + r \left(\mathbf{U}^{k+1} - \mathbf{W}^{k+1}\right).$$

Combining all these steps together, we obtain the proposed L1 Fuzzy Segmentation algorithm (L1FS) in Algorithm 1 for solving (4).

Algorithm 1 The proposed L1FS algorithm

− Initialization: $\mathbf{U}^0$ and $\mathbf{C}^0$ are specified, $\Lambda_{\mathbf{D}}^0 = \mathbf{0}, \Lambda_{\mathbf{W}}^0 = \mathbf{0}$.
− For $k = 0, 1, 2, \cdots$, repeat until the stopping criterion is reached.

$$\mathbf{D}^{k+1} = \mathcal{S}\left(\nabla \mathbf{U}^k + \frac{\Lambda_{\mathbf{D}}^k}{r}, \frac{1}{r}\right),$$

$$\mathbf{W}^{k+1} = \Pi_\Delta \left(\mathbf{U}^k + \frac{\Lambda_{\mathbf{W}}^k}{r} - \frac{\lambda |I - \mathbf{C}^k|}{r}\right),$$

$$\mathbf{C}^{k+1} = \psi(\mathbf{W}^{k+1}),$$

$$\mathbf{U}^{k+1} = \left(\nabla^T \nabla + \mathbf{I}\right)^{-1} \left(\nabla^T \mathbf{D}^{k+1} + \mathbf{W}^{k+1} - \frac{\nabla^T \Lambda_{\mathbf{D}}^k}{r} - \frac{\Lambda_{\mathbf{W}}^k}{r}\right),$$

$$\Lambda_{\mathbf{D}}^{k+1} = \Lambda_{\mathbf{D}}^k + r \left(\nabla \mathbf{U}^{k+1} - \mathbf{D}^{k+1}\right),$$

$$\Lambda_{\mathbf{W}}^{k+1} = \Lambda_{\mathbf{W}}^k + r \left(\mathbf{U}^{k+1} - \mathbf{W}^{k+1}\right).$$

− Output: $\mathbf{C}^{k+1}, \mathbf{U}^{k+1}$.

Remark: Though there are four variables $\mathbf{D}, \mathbf{W}, \mathbf{C}$ and $\mathbf{U}$, they can be grouped into two variables $(\mathbf{D}, \mathbf{W})$ and $(\mathbf{C}, \mathbf{U})$ and the subproblems for these two variables can be decoupled into the four subproblems. So it is essentially a two block ADMM applied on a nonconvex optimization problem.

Because problem (4) is nonconvex, a good initial guess of $\mathbf{U}^0$ and $\mathbf{C}^0$, which can be obtained from other methods using fuzzy membership functions, is helpful for the stability of L1FS. Thus we update $\mathbf{D}^1$ and $\mathbf{W}^1$ first because both of them can use the initial guess of $\mathbf{U}^0$.

4.2 Convergence analysis

If $\mathbf{C}$ is given, problem (4) is convex. Then, the algorithm is a standard ADMM by considering $(\mathbf{D}, \mathbf{W})$ together, and its convergence is guaranteed [5]. In this section, we give a convergence result of Algorithm 1 for unknown $\mathbf{C}$. To simplify notations, let us define the sextuples:

$$\mathbf{Z} := (\mathbf{D}, \mathbf{W}, \mathbf{C}, \mathbf{U}, \varLambda_{\mathbf{D}}, \varLambda_{\mathbf{W}}).$$

A point $\mathbf{Z}$ is a KKT point of (8) if it satisfies the following KKT conditions:

$$\partial \|d_i^*(x)\| - \lambda_{d_i}^*(x) \ni 0, \tag{11a}$$

$$\lambda |I - \mathbf{C}| - \varLambda_{\mathbf{W}} + \partial \delta_\Delta(\mathbf{W}) \ni \mathbf{0}, \tag{11b}$$

$$\int_\Omega \partial |I(x) - c_i^*| w_i^*(x) dx \ni 0, \tag{11c}$$

$$\nabla^T \lambda_{d_i}(x) + \lambda_{w_i}(x) = 0, \tag{11d}$$

$$\nabla u_i(x) - d_i(x) = 0, \tag{11e}$$

$$u_i(x) - w_i(x) = 0. \tag{11f}$$

where $\partial \| \cdot \|$ and $\partial \delta_\Delta(\cdot)$ are subdifferentials of $\| \cdot \|$ and $\delta_\Delta(\cdot)$, respectively.

Theorem 3 (Convergence to stationary points of L1FS) *Let $\left\{\mathbf{Z}^k\right\}_{k=1}^\infty$ be a sequence generated by Algorithm 1 that satisfies the condition*

$$\lim_{k \to \infty} \left(\mathbf{Z}^{k+1} - \mathbf{Z}^k\right) = \mathbf{0}.$$

Then any accumulation point of $\left\{\mathbf{Z}^k\right\}_{k=1}^\infty$ is a KKT point of problem (8). Consequently, any accumulation point of $\left(\mathbf{C}^k, \mathbf{U}^k\right)$ is a stationary point of problem (4).

Proof From the updating formulas in Algorithm 1, we obtain the following inequalities related to the subsequent iteration:

$$\mathbf{D}^{k+1} - \mathbf{D}^k = \mathcal{S}\left(\nabla \mathbf{U}^k + \frac{\varLambda_{\mathbf{D}}^k}{r}, \frac{1}{r}\right) - \mathbf{D}^k, \tag{12a}$$

$$\mathbf{W}^{k+1} - \mathbf{W}^k = \varPi_\Delta\left(\mathbf{U}^k + \frac{\varLambda_{\mathbf{W}}^k}{r} - \frac{\lambda|I - \mathbf{C}^k|}{r}\right) - \mathbf{W}^k, \tag{12b}$$

$$\mathbf{C}^{k+1} - \mathbf{C}^k = \psi(\mathbf{W}^{k+1}) - \mathbf{C}^k, \tag{12c}$$

$$\left(\nabla^T \nabla + I\right)\left(\mathbf{U}^{k+1} - \mathbf{U}^k\right) = \nabla^T(\mathbf{D}^{k+1} - \nabla \mathbf{U}^k) + (\mathbf{W}^{k+1} - \mathbf{U}^k) \tag{12d}$$

$$- \frac{\nabla^T \varLambda_{\mathbf{D}}^k}{r} - \frac{\varLambda_{\mathbf{W}}^k}{r}, \tag{12e}$$

$$\varLambda_{\mathbf{D}}^{k+1} - \varLambda_{\mathbf{D}}^k = r\left(\nabla \mathbf{U}^{k+1} - \mathbf{D}^{k+1}\right), \tag{12f}$$

$$\varLambda_{\mathbf{W}}^{k+1} - \varLambda_{\mathbf{W}}^k = r\left(\mathbf{U}^{k+1} - \mathbf{W}^{k+1}\right). \tag{12g}$$

By the assumption $\lim_{k \to \infty} \left(\mathbf{Z}^{k+1} - \mathbf{Z}^k \right) = \mathbf{0}$, the left-hand side and right-hand side of each equality in (12) go to zero as $k \to \infty$. Then we have

$$\mathcal{S} \left(\nabla \mathbf{U}^k + \frac{\Lambda_{\mathbf{D}}^k}{r}, \frac{1}{r} \right) - \mathbf{D}^k \to \mathbf{0}, \tag{13a}$$

$$\Pi_\Delta \left(\mathbf{U}^k + \frac{\Lambda_{\mathbf{W}}^k}{r} - \frac{\lambda |I - \mathbf{C}^k|}{r} \right) - \mathbf{W}^k \to \mathbf{0}, \tag{13b}$$

$$\psi(\mathbf{W}^{k+1}) - \mathbf{C}^k \to \mathbf{0}, \tag{13c}$$

$$\nabla^T (\mathbf{D}^{k+1} - \nabla \mathbf{U}^k) + (\mathbf{W}^{k+1} - \mathbf{U}^k) - \frac{\nabla^T \Lambda_{\mathbf{D}}^k}{r} - \frac{\Lambda_{\mathbf{W}}^k}{r} \to \mathbf{0}, \tag{13d}$$

$$\nabla \mathbf{U}^{k+1} - \mathbf{D}^{k+1} \to \mathbf{0}, \tag{13e}$$

$$\mathbf{U}^{k+1} - \mathbf{W}^{k+1} \to \mathbf{0}. \tag{13f}$$

Assume $\mathbf{Z}^* = (\mathbf{D}^*, \mathbf{W}^*, \mathbf{C}^*, \mathbf{U}^*, \Lambda_{\mathbf{D}}^*, \Lambda_{\mathbf{W}}^*)$ is an accumulation point of $\mathbf{Z}^k$. (13) gives us

$$\mathcal{S} \left(\nabla \mathbf{U}^* + \frac{\Lambda_{\mathbf{D}}^*}{r}, \frac{1}{r} \right) - \mathbf{D}^* = \mathbf{0}, \tag{14a}$$

$$\Pi_\Delta \left(\mathbf{U}^* + \frac{\Lambda_{\mathbf{W}}^*}{r} - \frac{\lambda |I - \mathbf{C}^*|}{r} \right) - \mathbf{W}^* = \mathbf{0}, \tag{14b}$$

$$\psi(\mathbf{W}^*) - \mathbf{C}^* = \mathbf{0}, \tag{14c}$$

$$\nabla^T \Lambda_{\mathbf{D}}^* + \Lambda_{\mathbf{W}}^* = \mathbf{0}, \tag{14d}$$

$$\nabla \mathbf{U}^* - \mathbf{D}^* = \mathbf{0}, \tag{14e}$$

$$\mathbf{U}^* - \mathbf{W}^* = \mathbf{0}. \tag{14f}$$

By (14a), it follows that $\mathbf{D}^*$ is a solution of the minimization problem

$$\min_{\mathbf{D}} \sum_{i=1}^{N} \left\{ \int_\Omega \|d_i(x)\| dx + \frac{r}{2} \int_\Omega \left\| d_i(x) - \nabla u_i^*(x) - \frac{\lambda_{d_i}^*(x)}{r} \right\|^2 dx \right\}.$$

Thus $\mathbf{D}^*$ satisfies the first order optimality condition

$$0 \in \partial \|d_i^*(x)\| + r \left(d_i^*(x) - \nabla u_i^*(x) - \frac{\lambda_{d_i}^*(x)}{r} \right).$$

Using (14e), we simplify the above condition as

$$0 \in \partial \|d_i^*(x)\| - \lambda_{d_i}^*(x). \tag{15}$$

By (14b), $\mathbf{W}^*$ is a solution of the following minimization problem:

$$\min_{\mathbf{W}} \sum_{i=1}^{N} \left\{ \frac{r}{2} \int_\Omega \left| w_i(x) - u_i^*(x) - \frac{\lambda_{w_i}^*(x)}{r} + \frac{\lambda |I(x) - c_i^*|}{r} \right|^2 dx \right\} + \delta_\Delta(\mathbf{W}).$$

Hence $\mathbf{W}^*$ satisfies the optimality condition

$$\mathbf{0} \in \partial \delta_\Delta(\mathbf{W}^*) + r(\mathbf{W}^* - \mathbf{U}^*) - \Lambda_\mathbf{W}^* + \lambda |I - \mathbf{C}^*|.$$

Together with the equality in (14f), the above equation can be simplified as

$$\mathbf{0} \in \partial \delta_\Delta(\mathbf{W}^*) - \Lambda_\mathbf{W}^* + \lambda |I - \mathbf{C}^*|. \tag{16}$$

The equation (14c) implies that $\mathbf{C}^*$ is a solution of the minimization problem

$$\min_\mathbf{C} \sum_{i=1}^N \int_\Omega |I(x) - c_i|\, w_i^*(x)\ dx,$$

and the optimal condition is

$$0 \in \int_\Omega \partial |I(x) - c_i^*|w_i^*(x)dx. \tag{17}$$

The equivalence of equations (14a)-(14c) with (15)-(17), together with equations (14d)-(14f) shows that the accumulation point $\mathbf{Z}^*$ satisfies the KKT condition in equations (11a)-(11f), thus $\mathbf{Z}^*$ is a KKT point of problem (8).

Since problem (8) and problem (4) are equivalent, the convergence of $(\mathbf{C}^k, \mathbf{U}^k)$ in the statement follows directly.   □

The convergence analysis is motivated by [46]. This theorem implies that whenever $\{\mathbf{Z}^k\}_{k=1}^\infty$ converges, it converges to a KKT point of problem (8). However, since the minimization problem (8) is nonconvex, the KKT point is not necessary to be an optimal solution.

5 Experimental results

In order to demonstrate the effectiveness of the proposed method, we compare our method with some existing methods on both synthetic and real images. These methods (FCM, FCM_S2, FLICM, L2FS, L2L0, and L1SS) are discussed in Section 2. Since the segmentation models in these methods are not jointly convex, and they are easy to get stuck in local minima, "good" initialization is crucial for all algorithms, especially when the given image is noisy. For FCM, FCM_S2, and FLICM, the initial $\mathbf{U}$ is uniformly distributed in $[0, 1]$ and normalized to satisfy the sum-to-one constraint. While for TV based methods L2FS and L1FS, one can also use the results of FCM and FCM_S2 as the initial $\mathbf{U}$ and $\mathbf{C}$. Here we consider three ways for choosing the initial $\mathbf{U}$ and $\mathbf{C}$: the result of FCM, the result of FCM_S2, and $\mathbf{U}$ as functions uniformly distributed in $[0, 1]$ and $\mathbf{C}$ as the result of FCM. In all the experiments, we choose the one with the highest performance among all the three initializations. For L1SS, we use the initialization method as described in the original paper. The stopping criterion, which is the same for all the methods except L1SS, is defined as

$$\frac{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_2}{\|\mathbf{U}^k\|_2} < \epsilon$$

where ϵ is a very small number. For L1SS, this stopping criterion does not work since L1SS leads to contrast loss due to the error in calculating class centers $\{c_i\}_{i=1}^{N}$ at early iterations. However, the contrast of L1SS will be enhanced gradually if the number of iterations increases. To gain satisfactory results, we choose to stop L1SS by setting the maximum number of iterations to be 1000.

The parameters of the methods being compared are set as follows. In FCM, we set $p = 2$. In FCM_S2, we set $p = 2, \alpha = 5$, and the window size for the median filter as 5×5. In FLICM, we set $p = 2$ and the window size of the neighborhoods as 3×3 or 5×5, which depends on the noise level. However, the weight parameter λ for L2FS, L1SS, and L1FS are tuned for each experiment to achieve optimal results. In all experiments, the range of λ for L2FS is $[0.00005, 0.0005]$, for L1SS is $[0.03, 1]$, and for L1FS is $[0.001, 0.05]$. For all methods, $\epsilon = 10^{-6}$ for the two-phase segmentation and $\epsilon = 10^{-4}$ for the multiphase segmentation. We use the default parameters in L2L0 except that we set $\lambda = 10$.

The clustering results of all methods are obtained by checking the maximum value of their membership functions. Then we display the recovered piecewise constant image as the final result. To compare segmentation results quantitatively, we consider the Segmentation Accuracy (SA) defined by

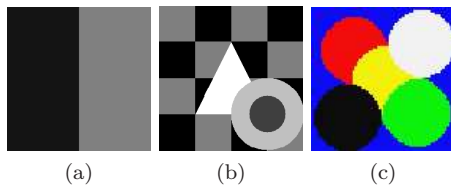$$SA = \frac{\#\text{correctly classified pixels}}{\#\text{all pixels}}.$$

(a)   (b)   (c)

Fig. 1 Synthetic piecewise constant test images. (a) A two-phase grayscale image with intensities 20 and 128, size 128×128; (b) A five-phase grayscale image with intensities 0, 63, 127, 192, and 255, size 235×237; (c) A six-phase color image with color vectors (12 11 242), (242 12 11), (242 241 242), (243 241 12), (12 12 12), and (12 242 12), size 100×100.
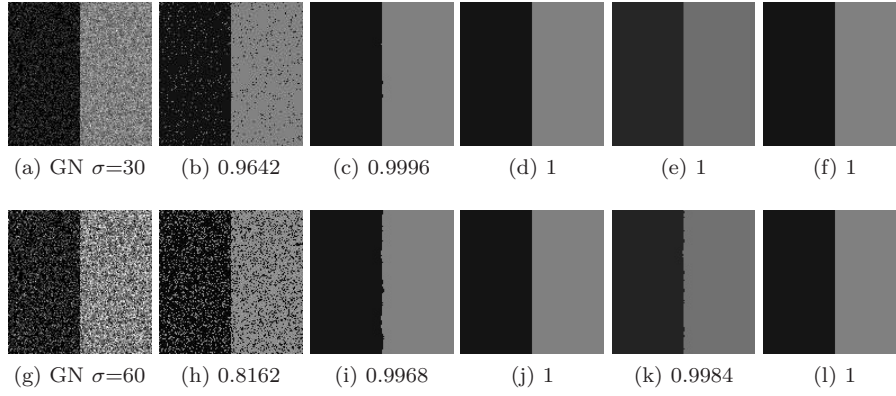
The synthetic piecewise constant test images are displayed in Fig. 1. The noisy images are contaminated by three types of noise: Gaussian Noise (GN) with zero mean and standard deviation varying from 10 to 80, Salt-and-Pepper Impulse Noise (SPIN) with 10% to 60% pixels contaminated, and Random-Valued Impulse Noise (RVIN) with 10% and 60% pixels contaminated.

5.1 Test on Fig. 1a

Tab. 1 provides the SA comparison of these six algorithms, and Figs. 2-4 show some of the corresponding results.

Table 1 SA performance of different methods applied on Fig. 1a contaminated by different levels of GN, SPIN, and RVIN.

GN (σ)	10	20	30	40	50	60	70	80
FCM	1	0.9970	0.9642	0.9146	0.8627	0.8162	0.7811	0.7524
FLICM	1	0.9999	0.9996	0.9990	0.9975	0.9968	0.9960	0.9954
L2FS	1	1	1	1	1	1	0.9999	**0.9998**
L1SS	1	1	1	0.9998	0.9996	0.9984	0.9975	0.9965
L1FS	1	1	1	1	1	1	1	**0.9998**
SPIN (%)	10%	20%	30%	40%	50%	60%		
FCM	0.9480	0.8983	0.8478	0.7979	0.7486	0.6980		
FLICM	0.9984	0.9921	0.9738	0.8002	0.7313	0.6554		
L2FS	0.9999	0.9998	0.9982	0.9983	-	-		
L1SS	0.9998	0.9990	0.9977	0.9967	0.9956	0.9953		
L1FS	1	1	1	1	1	**0.9995**		
RVIN (%)	10%	20%	30%	40%	50%	60%		
FCM_S2	0.9985	0.9972	0.9945	0.9862	0.9630	0.9042		
FLICM	0.9987	0.9985	0.9970	0.9958	0.9948	0.9919		
L2FS	1	1	1	0.9998	0.9995	0.9974		
L1SS	1	0.9995	0.9985	0.9979	0.9966	0.9891		
L1FS	1	1	1	1	1	**0.9976**		

(a) GN σ=30   (b) 0.9642   (c) 0.9996   (d) 1   (e) 1   (f) 1

(g) GN σ=60   (h) 0.8162   (i) 0.9968   (j) 1   (k) 0.9984   (l) 1

Fig. 2 Two-phase segmentation on the synthetic image Fig. 1a with Gaussian noise. First column: images contaminated by Gaussian noise with standard deviations 30 and 60, respectively; Second column to last column: results of FCM, FLICM, L2FS, L1SS, and L1FS, respectively. The SA values are reported below each segmentation result.

For GN, all methods being tested give correct segmentation results for small standard deviations (e.g., $\sigma = 10$). As the standard deviation increases, the SA value of FCM decays very fast. All the other algorithms have very large SA values even when the standard deviation is 80. L1FS has the best performance for all cases, and it is able to give correct segmentation results even when $\sigma \leq 70$. L2FS is the second best algorithm, and it is able to give correct segmentation results when $\sigma \leq 60$.

The results of all methods when $\sigma = 30$ and 60 are displayed in Fig. 2. The results of FCM are relatively "noisy" (the second column). For FLICM (the

(a) SPIN 20%   (b) 0.8983   (c) 0.9921   (d) 0.9998   (e) 0.9990   (f) 1

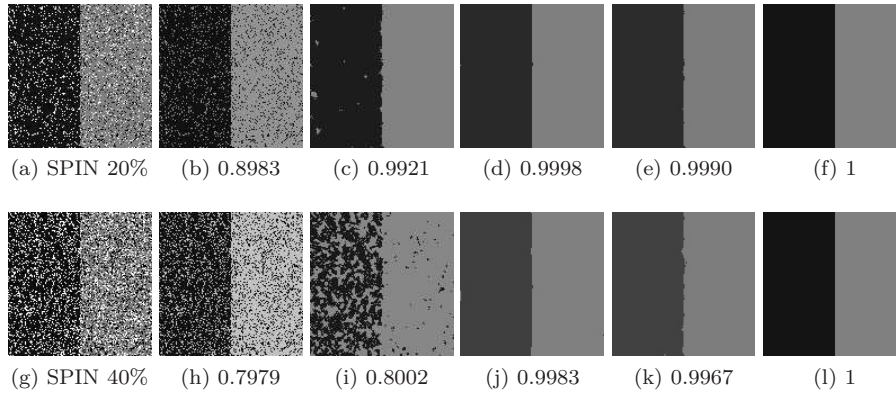(g) SPIN 40%   (h) 0.7979   (i) 0.8002   (j) 0.9983   (k) 0.9967   (l) 1

Fig. 3 Two-phase segmentation on the synthetic image Fig. 1a with SPIN. First column: images contaminated by 20% and 40% SPIN, respectively; Second column to last column: results of FCM, FLICM, L2FS, L1SS, and L1FS, respectively. The SA values are reported below each segmentation result.

(a) RVIN 20%   (b) 0.9972   (c) 0.9985   (d) 1   (e) 0.9995   (f) 1

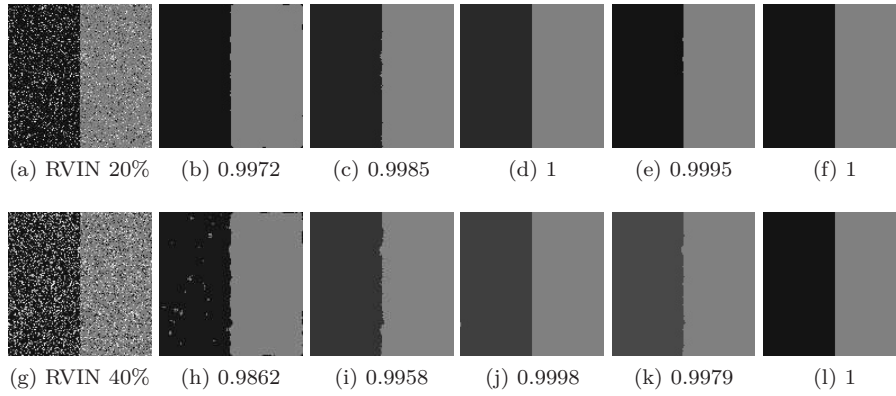(g) RVIN 40%   (h) 0.9862   (i) 0.9958   (j) 0.9998   (k) 0.9979   (l) 1

Fig. 4 Two-phase segmentation on the synthetic image Fig. 1a with RVIN. First column: images contaminated by 20% and 40% RVIN, respectively; Second column to last column: results of FCM_S2, FLICM, L2FS, L1SS, and L1FS, respectively. The SA values are reported below each segmentation result.

third column) and L1SS (the fifth column), the segmentation error occurs on the middle edge.

For SPIN, Tab. 1 shows that FLICM performs much better than FCM for noise levels 10%-30%. However, if the noise level is higher than 30%, both FCM and FLICM have very poor performance. L1SS achieves much better performance than FLICM, even when the noise level is higher than 30%. L2FS is slightly better than L1SS. Meanwhile, if the noise level is higher than 50%, L1SS fails to give a satisfactory result. L1FS has the highest SA among all methods. It gives completely correct segmentation results for noise levels 10%-50%. Fig. 3 shows the results of all methods for noise levels 20% and 40%,

respectively. For L2FS and L1SS, the segmentation errors occur along the middle edge. We also observe that for high noise levels such as 40%, both L2FS and L1SS suffer from slight contrast loss, e.g., Fig. 3j and Fig. 3k. However, L1FS is still able to preserve contrast, e.g., Fig. 3f and Fig. 3l.

For RVIN, FCM_S2 is used to initialize $\mathbf{U}$ and $\mathbf{C}$ for TV based methods L2FS and L1FS. Tab. 1 shows that FCM_S2 has the worst performance among all. L1SS is much better than FCM_S2 especially for high noise levels. FLICM performs slightly better than FCM_S2. L1FS achieves the best performance which is slightly better than L2FS. L2FS gives correct segmentation results at noise levels 10%-30%, while our method L1FS can give the correct segmentation results at noise levels 10%-50%. Fig. 4 shows the results for noise levels 20% and 40%. Again, we find that most of the errors occur along the middle edge for FLICM and L1SS. Moreover, the results of L2FS in Fig. 4j, L1SS in Fig. 4e and Fig. 4k lose some contrast.

Next we analyze the contrast problem for TV based methods. For L2FS, the estimated intensity c_i in each segmented region roughly equals the mean value of the intensities in that region. In the Gaussian noise case, the noise has zero mean and therefore it has almost no impact on the estimation of c_i. However, for both impulse noise cases, the noise has a significant influence on the estimation of c_i by taking the average. More specifically, assuming that the impulse noise follows the uniform distribution, its impact on the estimation of c_i is like this. Given an image with intensity range $[0, 255]$, for the region Ω_i with true intensity $\beta \leq 128$, if there are more noisy pixels with intensity greater than β than those with intensity less than β, then $c_i \geq \beta$ after taking the average and vice versa. Hence, the range of the image will be shrunk by applying L2FS even when the segmentation is correct, and thereby the recovered image will suffer from contrast loss. Note that the contrast loss problem has also been reported for the TVL2 restoration model in impulse noise removal [12]. For L1SS algorithm, one step of ADMM is used to solve the $\mathbf{C}$-subproblem approximately. Thus c_i is not accurate for the first few iterations. However, since L1SS uses the L1-norm fidelity, the loss of contrast becomes more and more subtle as the number of iterations increases. In L1FS, we solve the $\mathbf{C}$-subproblem exactly. Thus, L1FS can preserve contrast well in the segmentation process.

5.2 Test on Fig. 1b

The performance comparison for the multiphase synthetic piecewise constant gray image Fig. 1b is shown in Tab. 2 and Fig. 5.

As shown in Tab. 2, FCM performs poorly for GN, while L2FS and L1FS perform relatively well with similar SA. For the noise level $\sigma = 10$, both L2FS and L1FS give a correct segmentation result. For SPIN, FCM also gives the worst performance in terms of SA, while L1FS achieves the best performance. L2FS fails to yield a correct segmentation result when noise levels $\sigma \geq 50$. For RVIN, FCM_S2 achieves high SA values since it can smooth out some noise

Table 2 SA comparison of different methods applied on Fig. 1b with different levels of GN, SPIN, and RVIN.

GN (σ)	10	20	30	40	50	60	70	80
FCM	0.9987	0.8191	0.6634	0.5849	0.5233	0.4718	0.4319	0.4017
L2FS	**1**	**0.9999**	**0.9994**	0.9978	0.9959	**0.9950**	**0.9931**	**0.9918**
L1FS	**1**	**0.9999**	0.9993	**0.9980**	**0.9964**	**0.9950**	**0.9931**	0.9905
SPIN (%)	10%	20%	30%	40%	50%	60%	-	-
FCM	0.9202	0.8431	0.7638	0.6847	0.6096	0.5296	-	-
L2FS	0.9926	0.9877	0.9713	0.9673	-	-	-	-
L1FS	**0.9977**	**0.9948**	**0.9923**	**0.9894**	**0.9848**	**0.9782**	-	-
RVIN (%)	10%	20%	30%	40%	-	-	-	-
FCM	0.9922	0.9809	0.96672	0.9248	-	-	-	-
L2FS	0.9949	0.9923	0.9880	0.9731	-	-	-	-
L1FS	**0.9976**	**0.9957**	**0.9922**	**0.9868**-	-	-	-	-

Table 3 SA comparison of different methods applied on Fig. 1c with different levels of GN, SPIN, and RVIN.

GN (σ)	10	20	30	40	50	60	70	80
FCM	**1**	**1**	0.9998	0.9958	0.7927	0.7772	0.7488	0.7205
L2FS	**1**	**1**	**1**	**1**	0.9996	0.9992	**0.9989**	**0.9978**
L2L0	**1**	**1**	**1**	0.9999	0.9996	0.9991	0.9983	0.9967
L1FS	**1**	**1**	**1**	**1**	**0.9998**	**0.9994**	0.9985	0.9973
SPIN (%)	10%	20%	30%	40%	50%	60%	-	-
FCM	0.8498	0.7294	0.6128	0.5092	0.4248	0.3501	-	-
L2FS	0.9960	0.9925	0.9883	0.9822	0.9772	-	-	-
L2L0	0.9951	0.9880	0.9819	0.9740	0.9401	0.8752	-	-
L1FS	**0.9973**	**0.9937**	**0.9897**	**0.9854**	**0.9810**	**0.9732**	-	-
RVIN (%)	10%	20%	30%	40%	50%	60%	-	-
FCM	0.8971	0.7992	0.6967	0.6085	0.5196	0.4294	-	-
L2FS	0.9974	0.9957	0.9899	0.9853	0.9841	0.8988	-	-
L2L0	0.9971	0.9955	0.9906	0.9856	0.9727	0.9488	-	-
L1FS	**0.9988**	**0.9963**	**0.9939**	**0.9906**	**0.9881**	**0.9777**	-	-

in the segmentation process. L2FS performs much better than FCM_S2, and L1FS performs the best among all methods.

Fig. 5 shows some results corresponding to Tab.2. The first row is the noisy images being tested. The second row shows the results of FCM (Fig. 5g-5j) and FCM_S2 (Fig. 5k-5l). Most of them looks very "noisy" except Fig. 5k. L2FS and L1FS give very clean results in the third row and last row, respectively. Comparing the results by L2FS and L1FS for Gaussian noise, both of them have high visual qualities. However, for SPIN and RVIN, it is obvious that L1FS preserves the contrast much better than L2FS and has better segmentation results.
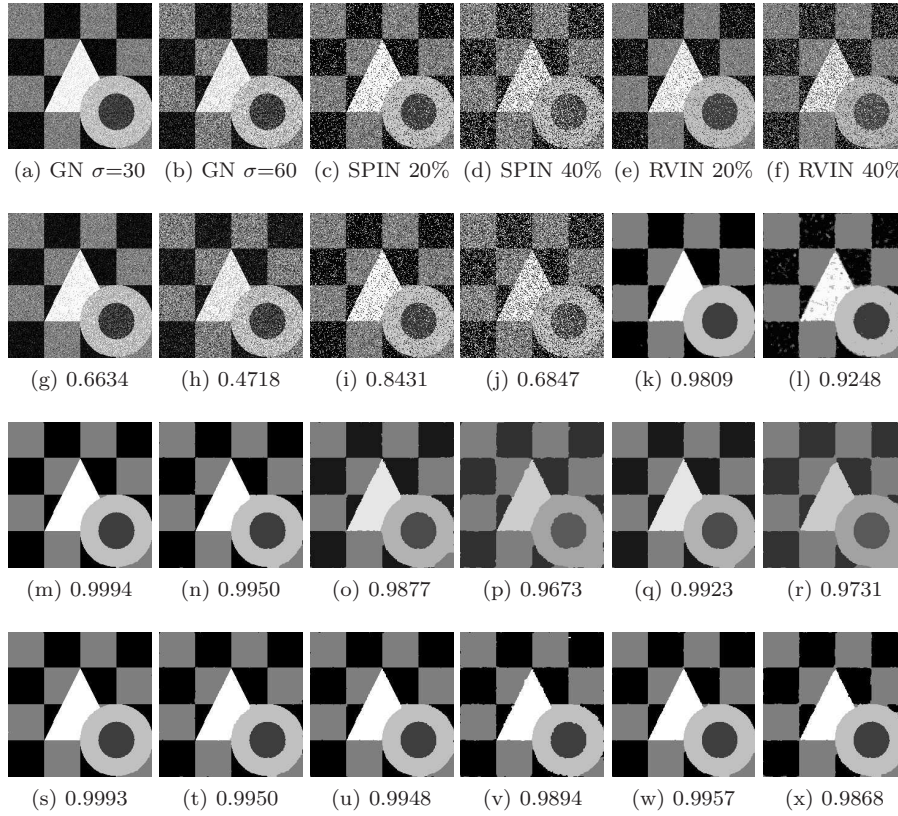
(a) GN σ=30 (b) GN σ=60 (c) SPIN 20% (d) SPIN 40% (e) RVIN 20% (f) RVIN 40%

(g) 0.6634 (h) 0.4718 (i) 0.8431 (j) 0.6847 (k) 0.9809 (l) 0.9248

(m) 0.9994 (n) 0.9950 (o) 0.9877 (p) 0.9673 (q) 0.9923 (r) 0.9731

(s) 0.9993 (t) 0.9950 (u) 0.9948 (v) 0.9894 (w) 0.9957 (x) 0.9868

Fig. 5 Multiphase segmentation on the synthetic image Fig. 1b with different levels of GN, SPIN, and RVIN. First row: images contaminated different types of noise with different levels. Second row to last row: results of FCM, L2FS and L1FS, respectively. The SA values are reported below each segmentation result.

5.3 Test on Fig. 1c

In Tab. 3 and Fig. 6, we test the multiphase synthetic piecewise constant color image Fig. 1c with various levels of GN, SPIN, and RVIN.

From Tab. 3, in the GN case, when the standard deviation of noise $\sigma \leq 20$, all the four methods, including FCM, L2FS, L2L0 and L1FS, give correct segmentation results. Moreover, both L2FS and L1FS yield correct segmentation results when $\sigma \leq 40$. When $\sigma \geq 50$, the performance of FCM decreases rapidly, while L2FS, L2L0, and L1FS still achieve very large SA values. Note that we initialize **U** randomly for GN in this test. For SPIN and RVIN, FCM has the worst performance which is far lower than that of the other three methods. It is also obvious that L1FS outperforms L2L0 and L2FS.

Fig. 6 shows some results corresponding to Tab. 3. The first row shows the tested noisy images. The results of FCM in the second row seems to be
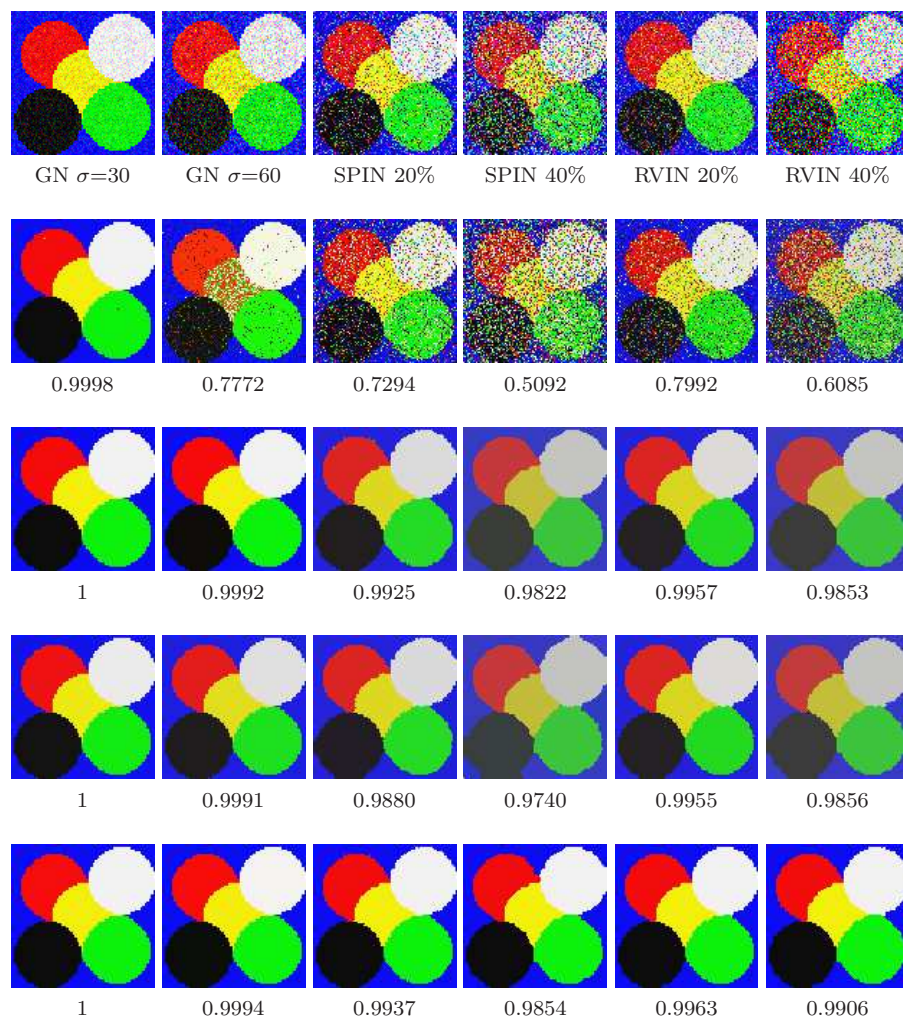
Fig. 6 Multiphase segmentation on the synthetic color image Fig. 1c with different levels of GN, SPIN, and RVIN. First row: images contaminated different types of noise with different levels. Second row to last row: results of FCM, L2FS, L2L0, and L1FS, respectively. The SA values are reported below each segmentation result.

"noisy" in most cases. The results of L2FS (the third row), L2L0 (the fourth row), and L1FS (the last row) are very clean. However, in terms of contrast, it is obvious that L1FS outperforms L2L0 and L2FS particularly for SPIN and RVIN.

5.4 Test on real images

We test some real images including two medical images and six natural images in Fig. 7 without artificial noise. However, the images can be regarded as containing some types of noise due to the acquisition and transmission processes.

The results of FCM and L1FS are displayed for comparison. One can see that FCM tends to produce some tiny components and irregular segmentation boundaries. By contrast, L1FS tends to smooth out tiny components to generate large ones and smooth boundaries between regions, which is more natural for human vision system and good for postprocessing. This smoothing effect is mainly achieved by total variation regularization in the L1FS model. Moreover, L1FS preserves slightly better contrast in the piecewise constant approximation than FCM, which is mainly achieved by the use of the L1-norm fidelity.

In Figs. 7j-7l, FCM and L1FS give quite different segmentation results. Obviously, FCM fails to segment the blue color part of the clothes in the original image, while the proposed L1FS works well. To illustrate the difference of these two methods, we display, in Fig. 8, the corresponding six segmented regions of FCM and L1FS for the women image Fig. 7j, respectively. We find that the segmented regions of FCM in Fig. 8a-8e are somehow "noisy". In particular, the background lattice pattern is partitioned into five regions as shown in Fig. 8a-8e. Compared with FCM, the proposed L1FS gives quite clean segmented regions in the second row in Fig. 8. Especially, the background lattice pattern are classified into only two classes as shown in Fig. 8g-8h. We further compare the blue parts of the clothes corresponding to Fig. 8c and Fig. 8i. In Fig. 8c, some background pattern heavily affects the estimation of $\mathbf{C}$, and therefore the color is not blue any more. However, in Fig. 8i, since the background is clean, the correct color can be obtained. To sum up, Fig. 8 demonstrates that the proposed L1FS gives smoother segmentations than FCM.

6 Conclusion

This paper presents a novel piecewise constant image segmentation model based on fuzzy membership functions and L1-norm fidelity. ADMM is applied to derive an efficient numerical algorithm, in which each subproblem has a closed-form solution. In particular, an efficient algorithm is proposed to find the fuzzy median. The numerical results on both synthetic and real piecewise constant images demonstrate that the proposed method is superior to some existing state-of-the-art methods since it is more robust to impulse noise and can preserve better contrast. Even in the case of Gaussian noise, the proposed method can achieve similar results as its L2-fidelity counterpart. In this work, we assume that the image to be dealt with is piecewise constant, which works well on images with homogeneous image features. The future work is to extend this framework to piecewise smooth image segmentation.
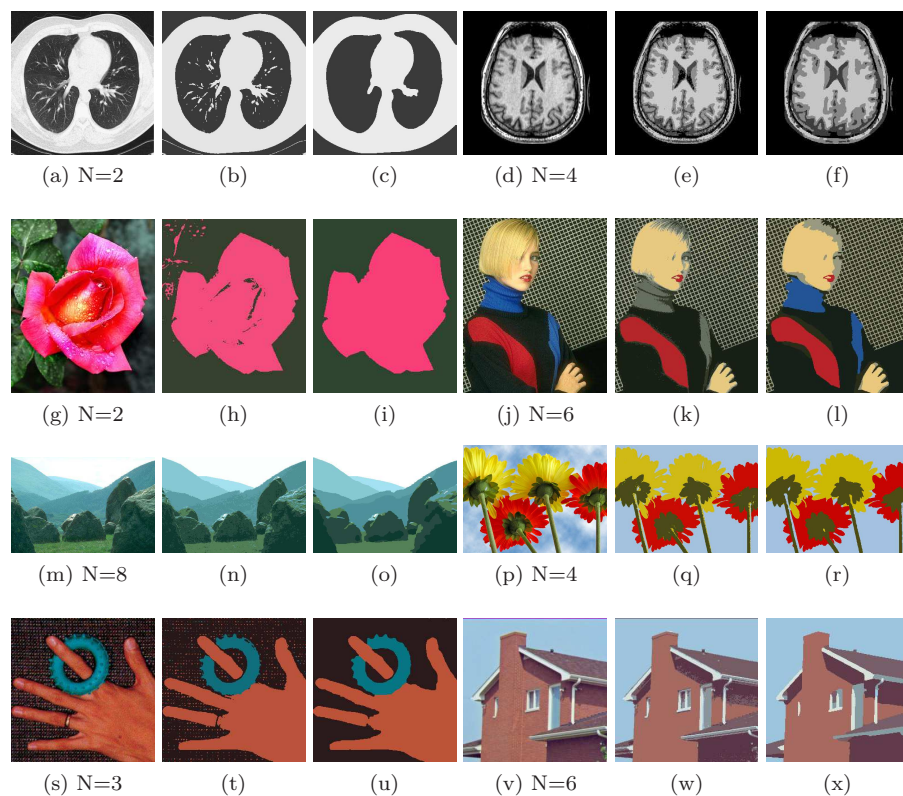
(a) N=2 (b) (c) (d) N=4 (e) (f)

(g) N=2 (h) (i) (j) N=6 (k) (l)

(m) N=8 (n) (o) (p) N=4 (q) (r)

(s) N=3 (t) (u) (v) N=6 (w) (x)

Fig. 7 Segmentation on real images. First column and fourth column: real color images; Second column and fifth column: results of FCM; Third column and last column: results of L1FS.
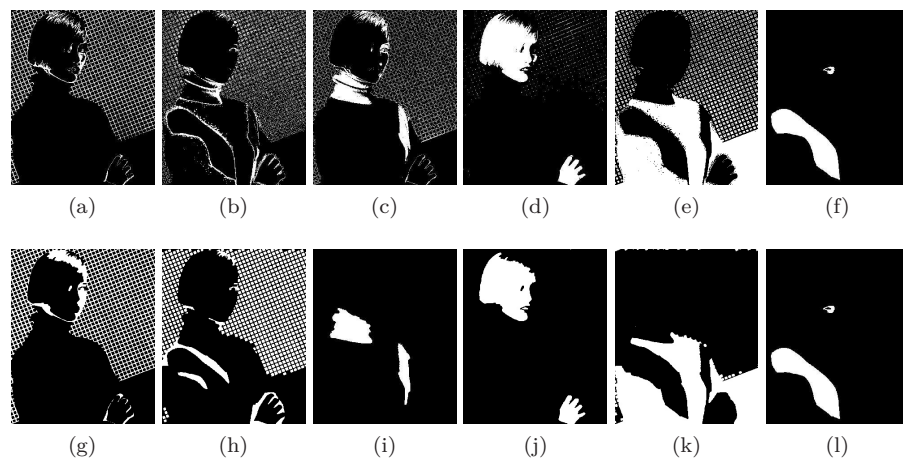

(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Fig. 8 Detailed comparison of FCM and L1FS on the woman image. First row: six segmented regions of FCM. Second row: six segmented regions of L1FS.

Acknowledgment

References

1. Mohamed N Ahmed, Sameh M Yamany, Nevin Mohamed, Aly A Farag, and Thomas Moriarty. A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *Medical Imaging, IEEE Transactions on*, 21(3):193–199, 2002.
2. Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147. Springer, 2006.
3. H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer New York, 2011.
4. James C Bezdek, LO Hall, and LP Clarke. Review of MR image segmentation techniques using pattern recognition. *Medical Physics*, 20(4):1033–1048, 1992.
5. Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
6. Xavier Bresson, Selim Esedoḡlu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and vision*, 28(2):151–167, 2007.
7. Xiaohao Cai. Variational image segmentation model coupled with image restoration achievements. *Pattern Recognition*, 48(6):2029–2042, 2015.
8. Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.
9. Antonin Chambolle, Daniel Cremers, and Thomas Pock. A convex approach to minimal partitions. *SIAM Journal on Imaging Sciences*, 5(4):1113–1158, 2012.
10. Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
11. Raymond Chan, Hongfei Yang, and Tieyong Zeng. A two-stage image segmentation method for blurry images with poisson or multiplicative gamma noise. *SIAM Journal on Imaging Sciences*, 7(1):98–127, 2014.
12. Raymond H Chan, Chung-Wa Ho, and Mila Nikolova. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *Image Processing, IEEE Transactions on*, 14(10):1479–1485, 2005.
13. Tony F Chan and Selim Esedoglu. Aspects of total variation regularized L1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.
14. Tony F Chan, Selim Esedoglu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.
15. Tony F Chan and Luminita A Vese. Active contours without edges. *Image Processing, IEEE transactions on*, 10(2):266–277, 2001.
16. Songcan Chen and Daoqiang Zhang. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(4):1907–1916, 2004.
17. Yunmei Chen and Xiaojing Ye. Projection onto a simplex. *arXiv preprint arXiv:1101.6081*, 2011.

18. Anastasia Dubrovina, Guy Rosman, and Ron Kimmel. Multi-region active contours with a single level set function. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(8):1585-1601, 2015.
19. Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.
20. Roland Glowinski and A Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.
21. Tom Goldstein and Stanley Osher. The split bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.
22. Weihong Guo, Jing Qin, and Sibel Tari. Automatic prior shape selection for image segmentation. *Research in Shape Modeling*, 2015.
23. Xiaoxia Guo, Fang Li, and Michael K Ng. A fast ℓ_1-TV algorithm for image restoration. *SIAM Journal on Scientific Computing*, 31(3):2322–2341, 2009.
24. Yanyan He, M Yousuff Hussaini, Jianwei Ma, Behrang Shafei, and Gabriele Steidl. A new fuzzy c-means method with total variation regularization for segmentation of images with noisy and incomplete data. *Pattern Recognition*, 45(9):3463–3471, 2012.
25. Nawal Houhou, J Thiran, and Xavier Bresson. Fast texture segmentation model based on the shape operator and active contour. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
26. Miyoun Jung, Myeongmin Kang, and Myungjoo Kang. Variational image segmentation models involving non-smooth data-fidelity terms. *Journal of Scientific Computing*, 59(2):277–308, 2014.
27. Yoon Mo Jung, Sung Ha Kang, and Jianhong Shen. Multiphase image segmentation via Modica-Mortola phase transition. *SIAM Journal on Applied Mathematics*, 67(5):1213–1232, 2007.
28. P.R. Kersten. Fuzzy order statistics and their application to fuzzy clustering. *Fuzzy Systems, IEEE Transactions on*, 7(6):708–712, 1999.
29. Stelios Krinidis and Vassilios Chatzis. A robust fuzzy local information c-means clustering algorithm. *Image Processing, IEEE Transactions on*, 19(5):1328–1337, 2010.
30. Fang Li, Michael K Ng, Tie Yong Zeng, and Chunli Shen. A multiphase image segmentation method based on fuzzy region competition. *SIAM Journal on Imaging Sciences*, 3(3):277–299, 2010.
31. Fang Li, Chaomin Shen, and Chunming Li. Multiphase soft segmentation with total variation and H1 regularization. *Journal of Mathematical Imaging and Vision*, 37(2):98–111, 2010.
32. Johan Lie, Marius Lysaker, and Xue-Cheng Tai. A piecewise constant level set framework. *International Journal of Numerical Analysis and Modeling*, 2(4):422–438, 2005.
33. Johan Lie, Marius Lysaker, and Xue-Cheng Tai. A binary level set model and some applications to Mumford-Shah image segmentation. *Image Processing, IEEE Transactions on*, 15(5):1171–1181, 2006.
34. Benoit Mory and Roberto Ardon. Fuzzy region competition: a convex two-phase segmentation framework. In *Scale Space and Variational Methods in Computer Vision*, pages 214–226. Springer, 2007.
35. David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mthematics*, 42(5):577–685, 1989.
36. Mila Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1-2):99–120, 2004.
37. Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.
38. Dzung L Pham. Fuzzy clustering with spatial constraints. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 2, pages II–65. IEEE, 2002.
39. Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

40. Alex Sawatzky, Daniel Tenbrinck, Xiaoyi Jiang, and Martin Burger. A variational framework for region-based segmentation incorporating physical noise models. *Journal of Mathematical Imaging and Vision*, 47(3):179–209, 2013.
41. Jianhong Shen and Tony F Chan. Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62(3):1019–1043, 2002.
42. Jianhong Jackie Shen. A stochastic-variational model for soft Mumford-Shah segmentation. *International Journal of Biomedical Imaging*, 2006:92329, 2006.
43. Martin Storath and Andreas Weinmann. Fast partitioning of vector-valued images. *SIAM Journal on Imaging Sciences*, 7(3):1826–1852, 2014.
44. Luminita A Vese and Tony F Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.
45. Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.
46. Yangyang Xu, Wotao Yin, Zaiwen Wen, and Yin Zhang. An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers of Mathematics in China*, 7(2):365–384, 2012.
47. Ming Yan. Restoration of images corrupted by impulse noise and mixed Gaussian impulse noise using blind inpainting. *SIAM J. Imaging Sciences*, 6(3):1227–1245, 2013.
48. Ming Yan, Alex A. T. Bui, Jason Cong, and Luminita A. Vese. General convergent expectation maximization (EM)-type algorithms for image reconstruction. *Inverse Problems and Imaging*, 7(3):1007–1029, 2013.
49. Song Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(9):884–900, 1996.