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

Fang Li\(^*\), Stanley Osher\(^†\), Jing Qin\(^†\), and 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.

Key words. image segmentation, fuzzy membership function, L1-norm, ADMM, segmentation accuracy.

AMS subject classifications. 65K10, 68U10, 90-08

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 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 [31], 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 [43]. In addition, the special Mumford-Shah model with a piecewise constant approximation is studied by Chan and Vese [12], and the level set method [33] 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 [7, 2, 18]. 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 [11, 6, 30, 21], which describe the levels of possible membership. Fuzzy membership functions assume that...
each pixel can be in several regions simultaneously with probability in \([0,1]\), and they 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 includes: 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/or constraints are involved in representing multiple regions and their boundaries effectively. The two-phase Chan-Vese model [12] has been generalized to multiphase segmentation by using multiple level set functions to represent multiple regions [38]. 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 [28, 29, 37, 23]. 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, 34, 13, 25, 27, 26, 20, 8].

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 [36, 32, 9, 19, 41, 42]. Particularly, the L1-norm fidelity is used for both salt-and-pepper impulse noise and random-valued impulse noise in image processing [32, 9, 19]. 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 contrast in the L2-norm case [10].

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. Alternating Direction Method of Multipliers (ADMM) [16, 15] is applied to solve this nonconvex optimization problem, which was rediscovered as split Bregman [17] and found to be very useful for L1 and TV type optimization problems. 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.

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, 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 $I : \Omega \to \mathbb{R}$ be the given clean or noisy image. Our goal is to find an $N$-phase “optimal” partition $\{\Omega_i\}_{i=1}^N$ such that $\Omega_i \cap \Omega_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^N \Omega_i = \Omega$. Define the set of $N$-phase fuzzy membership functions as

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

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 $U = (u_1, \ldots, u_N)$ and $C = (c_1, \ldots, c_N)$.

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

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

using the alternating minimization algorithm. Though $p$ can be any number no smaller than 1, it is commonly set to 2.

- FCM_S2 [13] – A variant of FCM which solves

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

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 [25] – Fuzzy Local Information C-Means clustering method which solves

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

where $\mathcal{N}(x)$ is a neighborhood of $x$. FLICM is more robust to both Gaussian noise and impulse noise than FCM.

- L2FS [20] – L2-norm fidelity based Fuzzy Segmentation method which solves

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

by ADMM. Here $\lambda > 0$ is a parameter, $\|\nabla u_i(x)\| := \sqrt{(\nabla x_1 u_i(x))^2 + (\nabla x_2 u_i(x))^2}$ for $x = (x_1, x_2)$ and $\int_{\Omega} \|\nabla u_i(x)\| \, dx$ denotes the TV of $u_i$. For fixed $c_i$ [20] is related to the popular TV denoising method [35].
L1SS [22] – 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 complicate for more than four phases, we show the four-phase model as follows:

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

where the membership functions \( M_i \), \( i = 1, \ldots, 4 \) are represented by soft functions \( u_1(x), u_2(x) \in [0,1] \):

\[
M_1(x) = u_1(x)u_2(x), \quad M_2(x) = u_1(x)(1 - u_2(x)),
\]

\[
M_3(x) = (1 - u_1(x))u_2(x), \quad M_4(x) = (1 - u_1(x))(1 - u_2(x)).
\]

**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 \( N \)-phase segmentation. Hence the solution space is much larger than L1SS, which ensures 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 generally 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 \( \chi_{\Omega_i} \) denotes the indicator function of region \( \Omega_i \), 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 which represents the given data in region \( \Omega_i \), 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

\[
(3.1) \quad \min_{\{\Omega_i\}, C} \sum_{i=1}^{N} \int_\Omega \| \nabla \chi_{\Omega_i}(x) \| \, dx + \lambda \int_\Omega \left| I(x) - \sum_{i=1}^{N} c_i \chi_{\Omega_i}(x) \right| \, dx,
\]

where \( \lambda > 0 \) is a tuning parameter. Note that the TV of \( \chi_{\Omega_i} \) in the first term is equal to the length of boundary \( \partial \Omega_i \). An equivalent form of model (3.1) is

\[
(3.2) \quad \min_{\{\Omega_i\}, C} \sum_{i=1}^{N} \left\{ \int_\Omega \| \nabla \chi_{\Omega_i}(x) \| \, dx + \lambda \int_\Omega \left| I(x) - c_i \chi_{\Omega_i}(x) \right| \, dx \right\}.
\]
Because $\chi_{\Omega_i}$ 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
\[
\Delta_0 := \left\{(u_1, \ldots, u_N) | u_i \in BV(\Omega), u_i(x) \in \{0, 1\}, \sum_{i=1}^{N} u_i(x) = 1, \forall x \in \Omega \right\}.
\]

At any point $x \in \Omega$, there is only one function having value 1, and all the other functions have value 0. Thus set $\Delta_0$ is not continuous, which leads to difficulties and instabilities in numerical implementations. However, we can relax binary indicator functions $\Delta$ value 0. Thus set
\[
\Omega_i \in \Delta.
\]

Note that model (3.3) is convex with respect to $U$ and $C$ respectively but not jointly. The difference between (3.3) and (2.1) is that the L2 fidelity term in (2.1) is replaced by L1 fidelity term. The existence of a minimizer for $E(U, C)$ in (3.3) is stated in Theorem 3.1.

**Theorem 3.1 (Existence of a minimizer).** Given an image $I \in L^\infty(\Omega)$, there exists a minimizer of $E(U, C)$ in $\Delta \times \mathbb{R}^N$ for any parameter $\lambda > 0$.

**Proof.** Since $E(U, C)$ is positive and proper, the infimum of $E(U, C)$ must be finite. Let us pick a minimizing sequence $(U^n, C^n) \in \Delta \times \mathbb{R}^N$ which satisfies $E(U^n, C^n) \rightarrow \inf_{U, C} E(U, C)$ as $n \rightarrow \infty$. Then there exists a constant $M > 0$ such that
\[
E(U^n, C^n) = \sum_{i=1}^{N} \left\{ \int_{\Omega} \|\nabla u^n_i(x)\| dx + \lambda \int_{\Omega} |I(x) - c_i| u^n_i(x) dx \right\} \leq M.
\]

Then each term in $E(U^n, C^n)$ is bounded, i.e., for each $i = 1, \ldots, N$,
\[
\int_{\Omega} \|\nabla u^n_i(x)\| dx \leq M, \quad \int_{\Omega} |I(x) - c_i| u^n_i(x) dx \leq M.
\]

Since $u^n_i(x) \in [0, 1]$, we have $\int_{\Omega} u^n_i(x) dx \leq |\Omega|$, where $|\Omega|$ is the area of $\Omega$. Together with the first equality in (3.4), we have that $u^n_i$ is uniformly bounded in $BV(\Omega)$ for all $i = 1, \ldots, N$. 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 $(u^n_i)$ after relabeling, there exists a function $u^*_i \in BV(\Omega)$ such that
\[
\begin{align*}
u^n_i \rightarrow & u^*_i \text{ strongly in } L^1(\Omega), \\
u^n_i \rightarrow & u^*_i \text{ a.e. } x \in \Omega, \\
abla u^n_i \rightarrow & \nabla u^*_i \text{ in the sense of measure.}
\end{align*}
\]
Then by the lower semicontinuity of the TV semi-norm,

\[(3.5)\]
\[
\int_{\Omega} \|\nabla u_i^n(x)\|dx \leq \liminf_{n \to \infty} \int_{\Omega} \|\nabla u_i^n(x)\|dx.
\]

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

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 \( u_i^n(x) \geq 0 \) and \( \int_{\Omega} u_i^n(x) dx > 0 \), the above equation implies that the constant \( c_i^n \) satisfies

\[|c_i^n| \leq \|I\|_{\infty}.
\]

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

\[(3.6)\]
\[
\int_{\Omega} |I(x) - c_i^*| u_i^*(x) dx \leq \liminf_{n \to \infty} \int_{\Omega} |I(x) - c_i^n| u_i^n(x) dx.
\]

Combining inequalities (3.5) and (3.6) for all \( i \), on a suitable subsequence, we established

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

and hence \((U^*, C^*)\) must be a minimizer of the energy \( E \). This completes the proof. \( \blacksquare \)

The minimizer of \( E(U, C) \) 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 \( U_\sigma = (u_{\sigma(1)}, ..., u_{\sigma(N)}) \), \( C_\sigma = (c_{\sigma(1)}, ..., c_{\sigma(N)}) \). It is straightforward to show the following theorem.

**Theorem 3.2 (Symmetry of minimizer).** For any \((U, C) \in \Delta \times \mathbb{R}^N\) and any \( \sigma \in S_N \),

\[E(U_\sigma, C_\sigma) = E(U, C).
\]

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

\[\begin{align*}
(U^*, C^*) &= \arg \min_{(U, C) \in \Delta \times \mathbb{R}^N} E(U, C), \\
(U^*_\sigma, C^*_\sigma) &= \arg \min_{(U, C) \in \Delta \times \mathbb{R}^N} E(U, C).
\end{align*}
\]

Then, for any \( \sigma \in S_N \), we have

\[\begin{align*}
(U^*_\sigma, C^*_\sigma) &= \arg \min_{(U, C) \in \Delta \times \mathbb{R}^N} E(U, C).
\end{align*}
\]
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 (3.3). 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 (3.3) is equivalent to the following minimization problem with equality constraints:

\[
\min_{\mathbf{D}, \mathbf{W}, \mathbf{C}, \mathbf{U}; \Lambda_{\mathbf{D}}, \Lambda_{\mathbf{W}}} \sum_{i=1}^{N} \left\{ \int_{\Omega} \|d_i(x)\| dx + \lambda \int_{\Omega} |I(x) - c_i| w_i(x) dx \right\} + \delta_{\Delta}(\mathbf{W}),
\]

subject to \( \nabla u_i = d_i, u_i = w_i, \forall i = 1, ..., N \),

where \( \delta_{\Delta} \) is the indicator function of the set \( \Delta \), 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 (4.1) is:

\[
\mathcal{L}(\mathbf{D}, \mathbf{W}, \mathbf{C}, \mathbf{U}; \Lambda_{\mathbf{D}}, \Lambda_{\mathbf{W}}) = \sum_{i=1}^{N} \left\{ \int_{\Omega} \|d_i(x)\| dx + \lambda \int_{\Omega} |I(x) - c_i| w_i(x) dx \right\} + \delta_{\Delta}(\mathbf{W})
\]

\[+ \sum_{i=1}^{N} \left\{ \langle \lambda_{d_i}, \nabla u_i - d_i \rangle + \frac{r}{2} \int_{\Omega} \|\nabla u_i(x) - d_i(x)\|^2 dx + \langle \lambda_{w_i}, u_i - w_i \rangle + \frac{r}{2} \int_{\Omega} |u_i(x) - w_i(x)|^2 dx \right\},\]

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:

\[
\mathbf{D}^{k+1} = \arg \min_{\mathbf{D}} \mathcal{L}(\mathbf{D}, \mathbf{W}^k, \mathbf{C}^k, \mathbf{U}^k; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k),
\]

\[
\mathbf{W}^{k+1} = \arg \min_{\mathbf{W}} \mathcal{L}(\mathbf{D}^{k+1}, \mathbf{W}, \mathbf{C}^k, \mathbf{U}^k; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k),
\]

\[
\mathbf{C}^{k+1} = \arg \min_{\mathbf{C}} \mathcal{L}(\mathbf{D}^{k+1}, \mathbf{W}^{k+1}, \mathbf{C}, \mathbf{U}; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k),
\]

\[
\mathbf{U}^{k+1} = \arg \min_{\mathbf{U}} \mathcal{L}(\mathbf{D}^{k+1}, \mathbf{W}^{k+1}, \mathbf{C}^{k+1}, \mathbf{U}; \Lambda_{\mathbf{D}}^k, \Lambda_{\mathbf{W}}^k),
\]

\[
\lambda_{d_i}^{k+1} = \lambda_{d_i}^k + r \left( \nabla u_i^{k+1} - d_i^{k+1} \right),
\]

\[
\lambda_{w_i}^{k+1} = \lambda_{w_i}^k + r \left( u_i^{k+1} - w_i^{k+1} \right).
\]

D-subproblem. The subproblem for 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 \( d_{ik}^{k+1} \) can be explicitly expressed using shrinkage operators. We simply compute

\[
d_{ik}^{k+1}(x) = S \left( \nabla u_i^k(x) + \frac{\lambda_{ik}^k(x)}{r}, \frac{1}{r} \right),
\]

where \( S \) is the shrinkage operator defined as

\[
S(v, \tau) := v \|v\| \max(\|v\| - \tau, 0).
\]

For the sake of simplicity, we denote this step as

\[
D^{k+1} = S \left( \nabla U^k + \frac{\Lambda_{k}}{r} \frac{1}{r} \right).
\]

**W-subproblem.** The subproblem for \( W \) is equivalent to

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

Since \( \Delta \) is a convex simplex at any \( x \in \Omega \), the solution is given by

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

where \( \Pi_{\Delta} \) is the projection onto the simplex \( \Delta \), for which more details can be found in [14]. We denote the step as

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

**C-subproblem.** The subproblem for \( C \) is

\[
C^{k+1} = \arg \min_{C} \sum_{i=1}^{N} \int_{\Omega} \left| I(x) - c_i \right| w_{ik}^{k+1}(x) dx,
\]

which is separable and \( c_{ik}^{k+1} \) can be solved independently. The optimality condition for each \( c_{ik}^{k+1} \) is

\[
0 \in - \int_{\Omega} \partial |I(x) - c_{ik}^{k+1}| w_{ik}^{k+1}(x) dx.
\]

Because the right-hand side of (4.2) is maximal monotone [3], the bisection method and ADMM are applied to solve it [24, 22]. The next lemma provides a new way to find an optimal solution for the discrete case.
Lemma 4.1. Given a finite non-decreasing sequence \( \{I_j\}_{j=1}^n \), i.e.,
\[
I_1 \leq I_2 \leq \cdots \leq I_n,
\]
and a non-negative sequence \( \{w_j\}_{j=1}^n \) with \( A = \sum_{j=1}^n w_j > 0 \), the optimal solution set for
\[
\min_c \sum_{j=1}^n |I_j - c| w_j,
\]
is \([I_{j^*}, I_{j^*+1}]\), where \( j^* \) and \( j_* \) satisfy
\[
A - 2 \sum_{j=1}^{j^*} w_j \leq 0 < A - 2 \sum_{j=1}^{j^*-1} w_j,
\]
\[
A - 2 \sum_{j=1}^{j_*} w_j < 0 \leq A - 2 \sum_{j=1}^{j_*-1} w_j.
\]
The fuzzy median of \( \{I_j\}_{j=1}^n \) with respect to the weight \( \{w_j\}_{j=1}^n \) [24], 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 (4.3) 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
\[
h(c) = \begin{cases}
A, & c < I_1, \\
[A - 2w_1, A], & c = I_1, \\
A - 2w_1, & c \in (I_1, I_2), \\
\cdots \\
[A - 2 \sum_{j=1}^s w_j, A - 2 \sum_{j=1}^{s-1} w_j], & c = I_s, \\
A - 2 \sum_{j=1}^s w_j, & c \in (I_s, I_{s+1}), \\
\cdots \\
-A, & c > I_n.
\end{cases}
\]
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 \( 0 \in h(I_{[\gamma^*]}) = [A - 2 \sum_{j=1}^{j^*} w_j, A - 2 \sum_{j=1}^{j^*+1} w_j] \) and \( I_{[\gamma^*]} \) is an optimal solution. In addition, we have that \( h(c) > 0 \) when \( c < I_{[\gamma^*]} \). 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 (4.3) is \( [I_{[\gamma^*]}, I_{[j_*+1]}] \).

When \( j^* = j_* + 1 \), the optimal solution is unique.

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 4.1 is satisfied. If there are multiple optimal solutions for \( c_i \), we choose the smallest one. We denote this step as

\[
C^{k+1} = \psi(W^{k+1}).
\]

**U-subproblem.** The subproblem for \( U \) is equivalent to

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

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

\[
\nabla^T \left( \nabla u_i^{k+1}(x) - d_i^{k+1}(x) + \frac{\lambda_i^k(x)}{r} \right) + \left( u_i^{k+1}(x) - u_i^{k+1}(x) + \frac{\lambda_i^k(x)}{r} \right) = 0,
\]

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

\[
(\nabla^T \nabla + I) u_i^{k+1}(x) = \nabla^T d_i^{k+1}(x) + u_i^{k+1}(x) - \frac{\nabla^T \lambda_i^k(x)}{r} - \frac{\lambda_i^k(x)}{r}.
\]

A diagonalization trick can be applied to find \( u_i^{k+1} \) efficiently [39]. We denote this step as

\[
U^{k+1} = (\nabla^T \nabla + I)^{-1} \left( \nabla^T D^{k+1} + W^{k+1} - \frac{\nabla^T \lambda_D^k}{r} - \frac{\lambda_W^k}{r} \right).
\]

**Updating dual variables.** We denote the steps as

\[
\lambda_D^{k+1} = \lambda_D^k + r \left( \nabla U^{k+1} - D^{k+1} \right),
\]

\[
\lambda_W^{k+1} = \lambda_W^k + r \left( U^{k+1} - W^{k+1} \right).
\]

Combining all these steps together, we obtain the proposed L1 Fuzzy Segmentation algorithm (L1FS) for solving (3.3) in Algorithm 1.

Remark: Though there are four variables \( D, W, C \) and \( U \), they can be grouped into two variables \( (D, W) \) and \( (C, U) \) and the sub-problems 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 (3.3) is nonconvex and a good initial guess of \( U^0 \) and \( C^0 \), which can be obtained from other methods using fuzzy membership functions, is helpful for the stability of L1FS. Thus we update \( D^1 \) and \( W^1 \) first because both of them can use the initial guess of \( U^0 \).
**Algorithm 1** The proposed L1FS algorithm

- Initialization: $U^0$ and $C^0$ are specified, $\Lambda_D^0 = 0$, $\Lambda_W^0 = 0$.
- For $k = 0, 1, 2, \ldots$, repeat until the stopping criterion is reached

\[
D^{k+1} = S \left( \nabla U^k + \frac{\Lambda_D^k}{r}, \frac{1}{r} \right),
\]
\[
W^{k+1} = \Pi_\Delta \left( U^k + \frac{\Lambda_W^k}{r} - \frac{\lambda |I - C^k|}{r} \right),
\]
\[
C^{k+1} = \psi(W^{k+1}),
\]
\[
U^{k+1} = \left( \nabla^T \nabla + 1 \right)^{-1} \left( \nabla^T D^{k+1} + W^{k+1} - \frac{\nabla^T \Lambda_D^k}{r} - \frac{\Lambda_W^k}{r} \right),
\]
\[
\Lambda_D^{k+1} = \Lambda_D^k + r \left( \nabla U^{k+1} - D^{k+1} \right),
\]
\[
\Lambda_W^{k+1} = \Lambda_W^k + r \left( U^{k+1} - W^{k+1} \right).
\]

- Output: $C^{k+1}, U^{k+1}$.

4.2. **Convergence analysis.** In this section, we give two convergence results of the proposed Algorithm 1. To simplify notation, let us define the sextuples

\[ Z := (D, W, C, U, \Lambda_D, \Lambda_W). \]

A point $Z$ is a KKT point of (4.1) if it satisfies the following KKT conditions for (4.1):

\[
\partial \|d_i^*(x)\| - \frac{\lambda_{d_i}(x)}{2} \geq 0,
\]
\[
\lambda |I - C| - \Lambda_W + \partial \delta_\Delta(W) \geq 0,
\]
\[
\int_{\Omega} \partial |I(x) - e_i^*| w_i^*(x) dx \geq 0,
\]
\[
\nabla^T \lambda_{d_i}(x) + \lambda_{w_i}(x) = 0,
\]
\[
\nabla u_i(x) - d_i(x) = 0,
\]
\[
u_i(x) - w_i(x) = 0.
\]

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

Theorem 4.2 (Convergence to stationary points of L1FS). Assume $\{Z^k\}_{k=1}^\infty$ is a sequence generated by Algorithm 1 that satisfies the condition

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

Then any accumulation point of $\{Z^k\}_{k=1}^\infty$ is a KKT point of problem (4.1). Consequently, any accumulation point of $(C^k, U^k)$ is a stationary point of problem (3.3).
Proof. From the updating formulas in Algorithm 1, we obtain the following inequalities related to the subsequent iteration:

\begin{align}
D^{k+1} - D^k &= S \left( \nabla U^k + \frac{\Lambda_D^k}{r}, \frac{1}{r} \right) - D^k, \\
W^{k+1} - W^k &= \Pi_\Delta \left( U^k + \frac{\Lambda_W^k}{r} - \frac{\lambda|I - C^k|}{r} \right) - W^k, \\
C^{k+1} - C^k &= \psi(W^{k+1}) - C^k, \\
\nabla^T (\nabla + I) \left( U^{k+1} - U^k \right) &= \nabla^T (D^{k+1} - \nabla U^k) + (W^{k+1} - U^k) - \frac{\nabla^T \Lambda_D^k}{r} - \frac{\Lambda_W^k}{r}, \\
\Lambda_D^{k+1} - \Lambda_D^k &= r \left( \nabla U^{k+1} - D^{k+1} \right), \\
\Lambda_W^{k+1} - \Lambda_W^k &= r \left( U^{k+1} - W^{k+1} \right).
\end{align}

By the assumption \( \lim_{k \to \infty} (Z^{k+1} - Z^k) = 0 \), the left-hand side and right-hand side of each equality in (4.5) go to zero as \( k \to \infty \). Then we have

\begin{align}
S \left( \nabla U^k + \frac{\Lambda_D^k}{r}, \frac{1}{r} \right) - D^k &\to 0, \\
\Pi_\Delta \left( U^k + \frac{\Lambda_W^k}{r} - \frac{\lambda|I - C^k|}{r} \right) - W^k &\to 0, \\
\psi(W^{k+1}) - C^k &\to 0, \\
\nabla^T (D^{k+1} - \nabla U^k) + (W^{k+1} - U^k) - \frac{\nabla^T \Lambda_D^k}{r} - \frac{\Lambda_W^k}{r} &\to 0, \\
\nabla U^{k+1} - D^{k+1} &\to 0, \\
U^{k+1} - W^{k+1} &\to 0.
\end{align}

Assume \( Z^* = (D^*, W^*, C^*, U^*, \Lambda_D^*, \Lambda_W^*) \) is an accumulation point of \( Z^k \). (4.6) gives us

\begin{align}
S \left( \nabla U^* + \frac{\Lambda_D^*}{r}, \frac{1}{r} \right) - D^* &= 0, \\
\Pi_\Delta \left( U^* + \frac{\Lambda_W^*}{r} - \frac{\lambda|I - C^*|}{r} \right) - W^* &= 0, \\
\psi(W^*) - C^* &= 0, \\
\nabla^T \Lambda_D^* + \Lambda_W^* &= 0, \\
\nabla U^* - D^* &= 0, \\
U^* - W^* &= 0.
\end{align}

By (4.7a), it follows that \( D^* \) is a solution of the minimization problem

\[
\min_D \sum_{i=1}^N \left\{ \int_{\Omega} \|d_i(x)\| dx + \frac{r}{2} \int_{\Omega} \left\| d_i(x) - \nabla u_i^*(x) - \frac{\lambda^*_d(x)}{r} \right\|^2 dx \right\}.
\]
Thus $D^*$ satisfies the first order optimality condition

$$0 \in \partial\|d^*_i(x)\| + \frac{r}{2} \left( d^*_i(x) - \nabla u^*_i(x) - \frac{\lambda^*_i(x)}{r} \right).$$

Using (4.7c), we simplify the above condition as

(4.8) $$0 \in \partial\|d^*_i(x)\| - \frac{\lambda^*_i(x)}{2}.$$

By (4.7b), $W^*$ is a solution of the following minimization problem

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

Hence $W^*$ satisfies the optimality condition

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

Together with the equality in (4.7f), the above equation can be simplified as

(4.9) $$0 \in \partial \delta_\Delta(W^*) - \Lambda_W^* + \lambda |I - C^*|.$$

The equation (4.7c) implies that $C^*$ is a solution of the minimization problem

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

and the optimal condition is

(4.10) $$0 \in \int_{\Omega} \partial |I(x) - c_i| w^*_i(x) \, dx.$$

The equivalence of equations (4.7a)-(4.7c) with (4.8)-(4.10), together with equations (4.7d)-(4.7f) shows that the accumulation point $Z^*$ satisfies the KKT condition in equations (4.4d)-(4.4f), thus $Z^*$ is a KKT point of problem (4.1).

Since problem (4.1) and problem (3.3) are equivalent, the convergence of $(C^k, U^k)$ in the statement follows directly.

The convergence analysis is motivated by [40]. This theorem implies that whenever $\{Z^k\}_{k=1}^{\infty}$ converges, it converges to a KKT point of problem (4.1). However, since the minimization problem (4.1) is nonconvex, the KKT point is not necessary to be an optimal solution.

**Remark:** If $C$ is fixed, problem (3.3) is convex. In addition $(D, W)$ can be considered as one variable, and the algorithm is a standard ADMM. Thus the convergence of this algorithm is guaranteed [5].
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, 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, a “good” initialization is crucial for all algorithms, especially when the given image is noisy. For FCM, FCM_S2, and FLICM, the initial $U$ 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 $U$ and $C$. Here we consider three ways for choosing the initial $U$ and $C$: the result of FCM, the result of FCM_S2, and $U$ as functions uniformly distributed in $[0, 1]$ and $C$ as the result of FCM. In all the experiments, we choose the one with 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

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

where $\epsilon$ 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 $5 \times 5$. In FLICM, we set $p = 2$ and the window size of the neighborhoods as $3 \times 3$ or $5 \times 5$, which depends on the noise level. However, the weight parameter $\lambda$ for L2FS, L1SS, and L1FS are tuned for each experiment to achieve optimal results. In all experiments, the range of $\lambda$ 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.

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

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

The synthetic piecewise constant test images are displayed in Fig. 1. Fig. 1(a) is a two-phase gray image, Fig. 1(b) is a five-phase gray image, and Fig. 1(c) is a six-phase color image. 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. 1(a).** Tab. 1 provides the SA comparison of these six algorithms, and Figs. 2-4 show some of the corresponding results.
Figure 1. Synthetic piecewise constant test images. (a) Two-phase grayscale image with intensities 20 and 128, size $128 \times 128$; (b) Five-phase grayscale image with intensities 0, 63, 127, 192, and 255, size $235 \times 237$; (c) Six-phase color image with color vectors $(12, 11, 242), (242, 2, 12, 11), (242, 241, 242), (243, 241, 12), (12, 12, 12),$ and $(12, 242, 12)$, size $100 \times 100$.

Table 1

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

| GN ($\sigma$) | 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       | 0.9937 | 0.9840 | 0.9664 | 0.9697 | 0.9711 | 0.9563 |
| FLICM        | 0.9985 | 0.9974 | 0.9989 | 0.9979 | 0.9970 | 0.9919 |
| L2FS         | 1     | 1     | 1     | 1     | 1     | 1     |
| L1SS         | 0.9984 | 0.9964 | 0.9955 | 0.9952 | 0.9949 | 0.9917 |
| L1FS         | 1     | 1     | 1     | 1     | 1     | 0.9996 |

For GN, all methods being tested give correct segmentation results for small standard deviation (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 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 level such as 40%, both L2FS and L1SS suffer from slight contrast loss, e.g., Fig. 3(j) and Fig. 3(k). However, L1FS is still able to preserve contrast, e.g., Fig. 3(f) and Fig. 3(l).

For RVIN, FCM_S2 is used to initialize $U$ and $C$ 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. L2FS and L1FS achieve the best performance. Both of them can give 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. 4(j), L1SS in Fig. 4(e) and Fig. 4(k) 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 significant influence on the estimation of $c_i$ by taking 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 $\Omega_i$ with true intensity $\beta \leq 128$, if there are more noisy pixels with intensity greater than $\beta$ than those with intensity less than $\beta$, then $c_i \geq \beta$ after taking average and vice versa. Hence, the range of 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 TVL2 restoration model for impulse noise removal [9]. For L1SS algorithm, one step of ADMM is used to solve the $C$-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

![Figure 4. Two-phase segmentation on synthetic image Fig. 1(a) 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.](image)
and more subtle as the number of iterations increases. In L1FS, we solve the \( C \)-subproblem exactly. Thus L1FS can preserve contrast well in the segmentation process.

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

| GN (\( \sigma \)) | 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% | 50% | 60% | - | - |
| FCM              | 0.9899 | 0.9735 | 0.9439 | 0.8541 | - | - | - | - |
| L2FS             | 0.9934 | 0.9906 | 0.9784 | 0.9754 | - | - | - | - |
| L1FS             | **0.9976** | **0.9958** | **0.9929** | **0.9894** | - | - | - | - |

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

| GN (\( \sigma \)) | 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 | 1 | **0.9996** | 0.9992 | **0.9989** | **0.9978** |
| L1FS             | 1 | 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 | - | - | - |
| L1FS             | **0.9973** | **0.9937** | **0.9897** | **0.9854** | **0.9810** | **0.9732** | - | - |
| RVIN (%)         | 10% | 20% | 30% | 40% | 50% | 60% | - | - |
| FCM              | 0.8914 | 0.7955 | 0.7022 | 0.6116 | - | - | - | - |
| L2FS             | 0.9942 | 0.9913 | 0.9901 | **0.9862** | - | - | - | - |
| L1FS             | **0.9971** | **0.9939** | **0.9905** | 0.9793 | - | - | - | - |

5.2. Test on Fig. 1(b). The performance comparison for the multiphase synthetic piecewise constant gray image Fig. 1(b) 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 the noise level $\sigma \geq 50$. For RVIN, FCM_S2 achieves high SA since it can smooth out some noise in the segmentation process. L2FS performs much better than FCM_S2, and L1FS performs 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. 5(g)-5(j)) and FCM_S2 (Fig. 5(k)-5(l)). Most of them look very “noisy” except Fig. 5(k). 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 quality. However, for SPIN and RVIN, it is obvious that L1FS preserves the contrast much better than L2FS and has better segmentation.
Figure 6. Multiphase segmentation on the synthetic color image Fig. 1(c) with different levels of GN, SPIN, and R VIN. First row: images contaminated different types of noise with different levels. Second row to last row: results of FCM, L2FS and L1FS, respectively.

5.3. Test on Fig. 1(c). In Tab. 3 and Fig. 6, we test the multiphase synthetic piecewise constant color image Fig. 1(c) with various levels of GN, SPIN, and R VIN.

From Tab. 3, in the GN case, when the standard deviation of noise $\sigma \leq 20$, all the three methods including FCM, L2FS, 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 and L1FS still achieve very large SA values. Note that we initialize $U$ randomly for GN in this test. For SPIN, FCM has the worst performance which is far lower than that of L2FS and L1FS. It is also obvious that L1FS outperforms L2FS. For R VIN, L1FS performs slightly better than L2FS when the noise level is lower than 30%, while for 40% noise, L2FS yields a slightly higher SA than L1FS. Both
methods fail to return good results when the noise level is higher than 40%.

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 “noisy” in most cases. Both the results of L2FS (the third row) and L1FS (the last row) are very clean. However, in terms of contrast, L1FS outperforms L2FS particularly for SPIN and RVIN (Fig. 6(o)-6(r) and Fig. 6(u)-6(x)).

5.4. Test on real images. We test some real images in Fig. 7 without artificial noise. 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 better for postprocessing. Moreover, L1FS preserves slightly better contrast in the piecewise constant approximation than FCM.

In Figs. 7(d)-7(f), 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 segments of FCM and L1FS when applying on the women image Fig. 7(d), respectively. We find that the segmented regions of FCM in Fig. 8(a)-8(e) are somewhat “noisy”. In particular, the background lattice pattern is partitioned into five regions as shown in Fig. 8(a)-8(e). 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. 8(g)-8(h). We further compare the blue parts of the clothes corresponding to Fig. 8(c) and Fig. 8(i). In Fig. 8(c), some background pattern heavily affects the estimation of \( C \) and therefore the color is not blue any more. However, in Fig. 8(i), 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.

Acknowledgment. The research of F. Li was supported by the 973 Program 2011CB707104, the research of S. Osher and J. Qin was supported by ONR Grants N00014120838 and N00014140444, NSF Grants DMS-1118971 and CCF-0926127, and the Keck Foundation, the research of M. Yan was supported by NSF Grants DMS-1349855 and DMS-1317602 and ARO MURI Grant W911NF-09-1-0383. The work was done when the first author was visiting UCLA Department of Mathematics.
Figure 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.

Figure 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.
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 Esedoğlu, 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] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[8] 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.

[9] 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.

[10] Tony F Chan and Selim Esedoglu. Aspects of total variation regularized l1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[11] Tony F Chan, Selim Esedoglu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denosing models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[12] Tony F Chan and Luminita A Vese. Active contours without edges. *Image Processing, IEEE transactions on*, 10(2):266–277, 2001.

[13] 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.

[14] Yunmei Chen and Xiaojing Ye. Projection onto a simplex. *arXiv preprint arXiv:1101.6081*, 2011.

[15] 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.

[16] Roland Glowinski and A Marroco. Sur l’approximation, par élements 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(2):11–76, 1975.

[17] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[18] Wei Hong Guo, Jing Qin, and Sibel Tari. Automatic prior shape selection for image segmentation. *Research in Shape Modeling*, 2015.

[19] Xiaoxia Guo, Fang Li, and Michael K Ng. A fast l1-TV algorithm for image restoration. *SIAM Journal on Scientific Computing*, 31(3):2322–2341, 2009.

[20] Yan Yan 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.

[21] 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.

[22] 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.
[23] Yoon Mo Jung, Sung Ha Kang, and Jianhong Shen. Multiphase image segmentation via Modica-Mortola phase transition. *SIAM Journal on Applied Mathematics*, 67(3):1213–1232, 2007.

[24] P.R. Kersten. Fuzzy order statistics and their application to fuzzy clustering. *Fuzzy Systems, IEEE Transactions on*, 7(6):708–712, 1999.

[25] Stelios Krimidis and Vassilios Chatzis. A robust fuzzy local information c-means clustering algorithm. *Image Processing, IEEE Transactions on*, 19(5):1328–1337, 2010.

[26] 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.

[27] Fang Li, Chaomin Shen, and Chunning Li. Multiphase soft segmentation with total variation and H1 regularization. *Journal of Mathematical Imaging and Vision*, 37(2):98–111, 2010.

[28] 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.

[29] 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.

[30] 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.

[31] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[32] Mila Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1-2):99–120, 2004.

[33] 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.

[34] Dzung L Pham. Fuzzy clustering with spatial constraints. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 2, pages II–65. IEEE, 2002.

[35] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[36] 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.

[37] Jianhong Jackie Shen. A stochastic-variational model for soft Mumford-Shah segmentation. *International Journal of Biomedical Imaging*, 2006:92329, 2006.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.