SIAM J. IMAGING SCIENCES
Vol. 9, No. 2, pp. 495–519

A Wavelet Frame Method with Shape Prior for Ultrasound Video Segmentation*

Jiulong Liu†, Xiaoqun Zhang‡, Bin Dong§, Zuowei Shen¶, and Lixu Gu‖

Abstract. Ultrasound video segmentation is a challenging task due to low contrast, shadow effects, complex noise statistics, and the need for high precision and efficiency in real time applications such as operation navigation and therapy planning. In this paper, we propose a wavelet frame based video segmentation framework incorporating different noise statistics and sequential distance shape priors. The proposed individual frame nonconvex segmentation model is solved by a proximal alternating minimization algorithm, and the convergence of the scheme is established based on the recently proposed Kurdyka–Łojasiewicz property. The performance of the overall method is demonstrated through numerical results on two real ultrasound video data sets. The proposed method is shown to achieve better results compared to the related level sets models and edge indicator shape priors, in terms of both segmentation quality and computational time.

Key words. variational model, wavelet frames, shape prior, alternating proximal methods, Gaussian and Poisson noise, real time tracking

AMS subject classifications. 65K10, 68U10, 42C40, 47A52

DOI. 10.1137/15M1033344

1. Introduction.

Image segmentation is a classical and important task in medical image analysis. It is the process of dividing the image domain into subdomains, typically used to locate regions of interest (ROI) and extract boundaries (lines, curves, etc.), which allows for a more meaningful and simplified representation of images. In this paper, we are interested in ultrasound image and video segmentation within a reasonable computation time. There are many potential applications for ultrasound video segmentation in clinical diagnosis, monitoring, and treatment, and operation navigation. However, due to low contrast, complex noise statistics, and the need for efficient implementation in real time applications, it is still challenging to develop a fast and stable segmentation algorithm for ultrasound videos. We refer the reader to [30] and the references therein for a survey of methods for ultrasound image segmentation and its applications. Many methods from different perspectives of image segmentation exist in the literature. Here, we mainly focus on variational methods for their

*Received by the editors July 31, 2015; accepted for publication (in revised form) February 2, 2016; published electronically April 7, 2016. The work of the first and second authors was partially supported by NSFC (91330102, GZ1025) and 973 program (2015CB856004).
http://www.siam.org/journals/siims/9-2/M103334.html
†School of Mathematics, Shanghai Jiao Tong University, 200240 Shanghai, China (jiulong.liu@gmail.com).
‡Institute of Natural Sciences and School of Mathematics, MOE-LSC, Shanghai Jiao Tong University, 200240 Shanghai, China (xqzhang@sjtu.edu.cn).
§Beijing International Center for Mathematical Research, Peking University, 100871 Beijing, People's Republic of China (dongbin@math.pku.edu.cn). The work of this author was supported in part by the Thousand Talents Plan of China.
¶Department of Mathematics, National University of Singapore, Singapore 119076, Singapore (matzuows@nus.edu.sg).
‖School of Biomedical Engineering, Shanghai Jiao Tong University, 200240 Shanghai, China (gulixu@stju.edu.cn).

modeling flexibility, along the line of the seminal work [29, 14, 40]. Variational models, such as active contour models [14, 40], have been widely used for edge detection and motion tracking. In addition, shape priors with geometry information are extensively studied in the literature; see, e.g., [15, 6, 38, 32]. For example, a set of hand drawn shapes is incorporated into an active contour level set model in [15] for cardiac ultrasound video segmentation. In [36], topological information is also incorporated into the variational model. In [16], Cremers proposed a level set based image sequence segmentation model incorporating a statistical shape prior using an autoregressive model for the dynamic deformation and transformation of the shapes. From a computational perspective, many variants of convex relaxation and numerical algorithms have been developed and analyzed (cf., e.g., [13, 7, 42, 27, 33, 12, 3]). Some recent work, such as [36, 39], attempted to investigate the impact of different noise models for ultrasound segmentation under the variational framework. Some related variational image segmentation models will be reviewed in section 2.

1.1. Motivations and contributions. Despite the fact that variational image segmentation models have been successfully applied to solve various image segmentation problems, it is still necessary to investigate the application of variational methods to video segmentation efficiently, especially when the videos are of low quality, such as in ultrasound video. Recently, tight frame based segmentation models have been proposed in [37, 19], where promising results are shown for segmenting complex tubular structures in low contrast medical images with fine structures. The wavelet frame, constructed from multiresolution analysis (MRA), can be generated by the unitary extension principle (UEP) of [35] because of the wavelet frame's multiresolution property and redundancy, which are beneficial for both algorithm efficiency and sparse representation of images [17, 21]. The wavelet frame based models [37, 19] and the variational models of [14, 40, 3] are closely related, as rigorously studied in the series of recent work [20, 8, 9]. It is shown that wavelet frame transforms are discretizations of differential operators in variational, PDE, and piecewise smooth Mumford–Shah frameworks. Numerous experiments in image restoration and image segmentation show that the discretization by wavelet frame transforms has advantages over standard finite difference discretization. The wavelet frame based method can extract more details than existing variational methods, especially when the image contains features of different scales [37]. Meanwhile, there have been recent developments on efficient convex minimization algorithms for sparsity related problems, which also motivates us to consider a wavelet frame based approach.

Our proposed model aims to segment an ultrasound video collectively and automatically by exploring the temporal continuity of the video. Our main contribution is to propose a wavelet frame based variational video segmentation model on incorporating shape priors for single image segmentation, and sequential shape priors are computed automatically for the following frame segmentation. Such a strategy yields an efficient implementation, as required in most real time operation navigation scenarios. In addition, as studied in [36], different noise statistics, in particular Gaussian and Poisson noise, are investigated under this framework. It should be pointed out that the proposed model is different from the previous shape prior based models. For example, compared to [16], the current model does not train a set of hand drawn shapes for fitting the shape regression models, and there is no specific deformation model for the shape transition between frames. The shape priors are calculated

directly from previous segmented shapes and used as an edge factor for the current image segmentation. Furthermore, a region based labeling function is used instead of level set functions. Together with the recently developed efficient first-order optimization algorithm, this leads to an efficient implementation for the image sequences segmentation. The proposed individual image segmentation nonconvex model, under the two types of noise assumption, is solved by a proximal alternating minimization (PAM) algorithm [5], and the convergence is established based on the Kurdyka–Łojasiewicz (KL) property. Numerical results show that the use of distance function based shape priors for wavelet frame video segmentation is beneficial for accurately and efficiently extracting edges on real ultrasound data.

1.2. Outline of the paper. The rest of the paper is organized as follows. In section 2, some related variational segmentation models, mainly based on bounded variation, are presented. In section 3, we present the distance function based shape prior wavelet frame video segmentation model. In section 4, the algorithm for solving the proposed single image segmentation model is presented in detail, and in section 5, a convergence analysis for this algorithm is provided. In section 6, numerical experiments are performed on two real ultrasound video data sets, and the paper is concluded in section 7.

2. Variational segmentation models. Let $\Omega \subset \mathbb{R}^2$ be an open and bounded image domain, and let f be the given image defined on Ω. The segmentation problem consists of finding a decomposition of the region $\Omega = (\bigcup_{i=1,\dots,K} \Omega_i) \bigcup \Gamma$, where Γ is the closed edge set and Ω_i are disjoint open sets representing different ROIs. The celebrated Mumford–Shah segmentation model [29] is to minimize the functional

$$E_{\mathrm{MS}}(u,\Gamma) := \mathcal{H}^{d-1}(\Gamma) + \lambda \int_{\Omega\setminus\Gamma} |\nabla u|^2 dx + \mu \int_{\Omega} (u-f)^2 dx, \tag{2.1}$$

where $\mathcal{H}^{d-1}(\Gamma)$ is the Hausdorff measure in Ω, u is the piecewise smooth approximation of f, and the positive constants λ and μ balance the terms. The functional in (2.1) is to establish an optimality criterion for segmenting an image into subregions by forcing the approximation u to be piecewise smooth, except along the edges Γ, and the total length of the edges to be as short as possible. For more details on the theoretical aspects of the Mumford–Shah model and its approximation, we refer the reader to [29, 40]. The Mumford–Shah model is not easy to solve, and a convenient simplification is the Chan–Vese active contour model [14, 40], which takes u as a piecewise constant function for each open region Ω_i. In particular, for the two regions case of extracting an open ROI $\Sigma \subset \Omega$ and $\Gamma = \partial\Sigma$, we minimize

$$E_{\mathrm{CV}}(c_1,c_2,\Sigma) = \mathcal{H}^{d-1}(\partial\Sigma) + \mu \int_{\Sigma} (c_1-f)^2 dx + \mu \int_{\Omega\setminus\Sigma} (c_2-f)^2 dx, \tag{2.2}$$

where c_1 and c_2 represent the constant estimation of Σ (foreground) and $\Omega\setminus\Sigma$ (background). It is easy to see that for a fixed Σ, the constants c_1 and c_2 can be obtained as the mean intensity of each region. By representing the edge $\partial\Sigma$ as a Lipschitz level set function [31] $\phi : \Omega \to \mathbb{R}$ such that

$$\begin{cases} \partial\Sigma = \{x \in \Omega : \phi(x) = 0\}, \\ \Sigma = \{x \in \Omega : \phi(x) > 0\}, \\ \Omega\setminus\Sigma = \{x \in \Omega : \phi(x) < 0\}, \end{cases} \tag{2.3}$$

the length term in the energy functional (2.1) can be expressed as

$$\begin{aligned}\mathcal{H}^{d-1}(\partial\Sigma) &= \int_\Omega |\nabla H(\phi)|dx \\ &= \int_\Omega \delta_0(\phi)|\nabla\phi|dx,\end{aligned} \tag{2.4}$$

where $H(z)$ is the Heaviside function

$$H(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0, \end{cases} \tag{2.5}$$

and the one-dimensional Dirac measure $\delta_0(z) = \frac{d}{dz}H(z)$. The overall Chan–Vese functional (2.2) can be written as

$$\begin{aligned}&E_{\text{CV-LS}}(c_1, c_2, \phi) \\ &= \int_\Omega \delta_0(\phi)|\nabla\phi|dx + \frac{\mu}{2}\int_\Omega (c_1 - f)^2 H(\phi)dx + \frac{\mu}{2}\int_\Omega (c_2 - f)^2(1 - H(\phi))dx.\end{aligned} \tag{2.6}$$

Numerically, we need to replace (2.6) with a regularized functional

$$\begin{aligned}&E_{\text{CV- LS}_\epsilon}(c_1, c_2, \phi) \\ &= \int_\Omega \delta_\epsilon(\phi)|\nabla\phi|dx + \frac{\mu}{2}\int_\Omega (c_1 - f)^2 H_\epsilon(\phi)dx + \frac{\mu}{2}\int_\Omega (c_2 - f)^2(1 - H_\epsilon(\phi))dx,\end{aligned} \tag{2.7}$$

where

$$H_\epsilon(z) = \begin{cases} 1, & z > \epsilon, \\ 0, & z < -\epsilon, \\ \frac{1}{2}\{1 + \frac{z}{\epsilon} + \frac{1}{\pi}\sin(\frac{\pi z}{\epsilon})\}, & |z| \leq \epsilon. \end{cases} \quad \delta_\epsilon(z) = \begin{cases} 0, & |z| > \epsilon, \\ \frac{1}{2\epsilon}\{1 + \cos(\frac{\pi z}{\epsilon})\}, & |z| \leq \epsilon. \end{cases} \tag{2.8}$$

For fixed c_1 and c_2, the minimization of $E_{\text{CV-LS}_\epsilon}$ with respect to ϕ can be obtained by solving the following evolution equation:

$$\begin{cases} \frac{\partial\phi}{\partial t} = \delta_\epsilon(\phi)[\mathbf{div}(\frac{\nabla\phi}{|\nabla\phi|}) - \mu(c_1 - f)^2 + \mu(c_2 - f)^2] \ \text{ in } \ (0, \infty) \times \Omega, \\ \phi(0, x, y) = \phi_0(x) \ \text{ in } \ \Omega, \\ \frac{\delta_\epsilon(\phi)}{|\nabla\phi|}\frac{\partial\phi}{\partial\vec{n}} = 0 \ \text{ on } \ \partial\Omega. \end{cases} \tag{2.9}$$

The Chan–Vese model is also relaxed as a continuous field binary labeling problem in [34, 13],

$$E_{\text{CV}}(c_1, c_2, u) = \int_\Omega |Du|dx + \frac{\mu}{2}\int_\Omega u(c_1 - f)^2 dx + \frac{\mu}{2}\int_\Omega (1 - u)(c_2 - f)^2 dx, \tag{2.10}$$

where

$$u(x) = \begin{cases} 1 & \text{if } \ x \in \Sigma, \\ 0 & \text{if } \ x \in \Omega\backslash\Sigma, \end{cases} \tag{2.11}$$

and the first formula is due to the co-area formula [1, 39]

$$\mathcal{H}^{d-1}(\partial\Sigma) = \int_{\Omega} |D\chi_{\Sigma}(x)|dx. \tag{2.12}$$

Many convex relaxations of the model (2.10) and its relation to the global solutions are studied in [13, 7, 42, 27, 33, 12, 3]. In addition, an edge indicator function also can be used as a shape prior as proposed in [15, 6, 38, 32]:

$$E_{\mathrm{CV}}(c_1, c_2, u) := \int_{\Omega} |\varphi \cdot \nabla u|dx + \mu \int_{\Omega} u(c_1 - f)^2 dx + \mu \int_{\Omega} (1-u)(c_2 - f)^2 dx. \tag{2.13}$$

The function φ is a positive and regular function depending on the edges of f satisfying $\lim_{t\to\infty} \varphi(t) = 0$. The use of shape priors allows us to find boundaries that are close to the given shape priors, which are extremely useful for extracting ROIs and moving boundaries. For instance, in [11, 6] the edge indicator function

$$\varphi(x) = g(|\nabla f(x)|) = \frac{1}{1 + |\nabla \tilde{f}(x)|^2}, \tag{2.14}$$

where $\tilde{f}(x) = (G_\sigma * f)(x)$ is a smoothed version of f and G_σ is a Gaussian kernel with variance σ, is used to drive the partition to embed into the predefined edges.

3. Wavelet frame video segmentation model with shape prior. In this section, we consider a wavelet frame segmentation model in a discrete setting. Given a video sequences $\{f_t\}$ for $t \in 1, \ldots, T$, where each frame f_t is defined on a discrete grid domain $\Omega \subset \mathbb{R}^2$, our objective is to find a meaningful partition of the pixel index set Ω. We start by introducing a segmentation model for each individual frame f_t. For simplicity, we will drop the index t for the presentation of single image segmentation $f(x)$ for $x \in \Omega$ and will only consider a two-region approximation of f as

$$z(x) = z_1(x)u(x) + z_2(x)(1 - u(x)) \quad \text{for } x \in \Omega,$$

where

$$u(x) = \begin{cases} 1 & \text{if } x \in \Sigma, \\ 0 & \text{if } x \in \Omega\backslash\Sigma, \end{cases} \tag{3.1}$$

and Σ denotes the foreground ROI, $\Omega\backslash\Sigma$ denotes the background, and $z_1(x)$ and $z_2(x)$ denote the approximation of the original noise free image on Σ and $\Omega\backslash\Sigma$ with appropriate extension, respectively.

The fast tight frame transform, or decomposition operator, is denoted as $\mathcal{W}$, where the operator $\mathcal{W}$ consists of $J+1$ subfiltering operators $\mathcal{W}_0, \mathcal{W}_1, \ldots, \mathcal{W}_J$, and $\mathcal{W}^T\mathcal{W} = I$, i.e., $\mathcal{W}^T\mathcal{W}u = u$ for any image u. For more details on the wavelet frame transform and its application in imaging, we refer the reader to [21]. The proposed wavelet frame based segmentation model takes the general form

$$\min_{z_1, z_2, u \in \{0,1\}} E(z_1, z_2, u) := \mu D(z_1, z_2, u) + \sum_{i=1}^{2} R(z_i) + |\varphi \cdot \mathcal{W}u|_1, \tag{3.2}$$

where $D(z_1, z_2, u) := -\log p(f|z)$ denotes the negative log likelihood function related to data misfitting, $R(z_i)$ denotes a regularization term related to approximation z_i, and the last term is the weighted sparse regularization on the characteristic function u with predefined $\varphi(x)$. We note that the weight function is applied to all of the subbands, and this term is related to the regularity of the edge set Σ. The weight function φ allows us to incorporate shape priors, which we will discuss later. The idea of recovering a piecewise smooth function with a general smoothness regularization and general singularity set, including high-order jump discontinuities, has been studied in a recent work [9], and it is rigorously proven that the wavelet frame approach converges to a variational approach on the space of piecewise smooth functions as image resolution goes to infinity. Instead of recovering edge sets $\partial\Sigma$ directly, we propose recovering the region characteristic function, which allows for a stable and fast algorithm for a convex relaxed function. Meanwhile, this energy also can be interpreted from the perspective of maximum a posteriori (MAP) estimation, as proposed in [36].

Under the assumption of independence of noise at each pixel x, we have

$$p(f|z) \propto \prod_{x\in\Omega} p(f(x)|z(x)) = \prod_{x\in\Sigma} p(f(x)|z_1(x)) \cdot \prod_{x\in\Omega\setminus\Sigma} p(f(x)|z_2(x)),$$

and $D(z_1, z_2, u)$ reduces to

$$D(z_1, z_2, u) = -\sum_{x\in\Sigma} \log(p(f(x)|z_1(x)) - \sum_{x\in\Omega\setminus\Sigma} \log(p(f(x)|z_2(x)) \tag{3.3}$$

$$= \langle u,\ -\log p(f|z_1)\rangle + \langle \mathbf{1} - u,\ -\log p(f|z_2)\rangle. \tag{3.4}$$

Here $\langle , \rangle$ denotes the usual inner product of two vectors, and $\mathbf{1}$ denotes the all-ones vector.

In the following, we further simplify the models under the assumption of piecewise constant approximation of f, i.e., $z_i(x) = c_i$. For the noise statistics, we discuss Gaussian and Poisson noise present in the data f, for such a simplification is enough for ultrasound image segmentation.

- When the noise in data is assumed to be Gaussian, which is the usual choice in many segmentation formulations, we have

  $$p(f(x)|z_i(x)) \propto \exp^{-\frac{1}{2\sigma^2}(c_i - f(x))^2}, \tag{3.5}$$

  and the data term reduces to

  $$D(c_1, c_2, u) = \langle u,\ (f - c_1)^2\rangle + \langle \mathbf{1} - u,\ (f - c_2)^2\rangle. \tag{3.6}$$

- If f is corrupted by Poison noise, the distribution of f is in the form of

  $$p(f|c_i) = \frac{c_i^f}{f!} e^{-c_i} \quad \text{for} \quad i \in \{1, 2\}, \tag{3.7}$$

  and the data term becomes

  $$D(c_1, c_2, u) = \langle u, c_1 - f\log c_i\rangle + \langle \mathbf{1} - u, c_2 - f\log c_2\rangle + \langle \mathbf{1}, \log(f!)\rangle. \tag{3.8}$$

Furthermore, for a convenient convergent analysis of the proposed algorithm in the following section, we consider $R(c_i) = \chi_{[c_l,\ c_u]}(c_i)$, where χ is the set indicator function; i.e., we consider a constraint for $\vec{c} = \binom{c_1}{c_2} \in [c_l, c_u]$, where c_l and c_u are the lower and the upper bound for the intensity. Without loss of generality, we assume $0 < c_l < c_u$, and then we propose the model with binary constraint as

$$\min_{u\in\{0,1\},c_l\le\vec{c}\le c_u} |\varphi\cdot\mathcal{W}u|_1 + \mu D(\vec{c},u). \tag{3.9}$$

It is well known that it is hard to solve the problem (3.9) with binary constraint $u \in \{0,1\}$. We relax this constraint to $u \in [0,1]$ as in some previous work on variational image segmentation models. Finally, we propose the model

$$\min_{0\le u\le 1,c_l\le\vec{c}\le c_u} |\varphi\cdot\mathcal{W}u|_1 + \mu D(\vec{c},u) \tag{3.10}$$

for single image segmentation.

If the global minimum of (3.10) is binary, then it is naturally a global minimum of the nonconvex model (3.9). In the following, we also analyze the equivalence of the binary solution and the relaxed solution for a fixed $\vec{c}$. We rewrite the Fenchel conjugate form of (3.10) as

$$\max_{|p(x)|\le\varphi(x)} \min_{0\le u\le 1} E(u,p) := \langle \mathcal{W}^T p + I_1 - I_2,\ u\rangle, \tag{3.11}$$

where p is a dual variable with the same dimension as $\mathcal{W}u$, and

$$I_i = \begin{cases} (c_i - f)^2 & \text{Gaussian noise,} \\ c_i - f\log c_i & \text{Poisson noise} \end{cases}$$

for $i = 1, 2$. Let (u^*, p^*) be the primal dual solution pair of (3.11); then p^* is the optimal solution of the dual problem $\max_{|p(x)|\le\phi(x)} E_D(p) := \min_{0\le u\le 1} E(u,p)$. The strong duality theory leads to

$$u^* = \arg\min_{0\le u\le 1} E(u,p^*).$$

This implies that the optimal solution satisfies

$$u^*(x) = \begin{cases} 1 & \text{if } [\mathcal{W}^T p^* + I_1](x) < I_2(x), \\ 0 & \text{if } [\mathcal{W}^T p^* + I_1](x) > I_2(x), \\ [0,1], & \text{otherwise.} \end{cases}$$

It is easy to see that even though an optimal solution u^* is not binary everywhere, one can transform those nonbinary points into 0 or 1, and the new solution still satisfies the optimality of the convex relaxed problem. This binary solution is thus a global solution for the binary nonconvex model. In practice, the convex problem (3.11) is solved by a primal dual algorithm, which alternatively gives the approximation of p^* and u^*.

In the following, we will present the algorithm for solving c and u alternatively. First, combined with the Gaussian and Poisson noise distribution present in (3.6) and (3.8), the above model (3.10) can be rewritten as

$$\min_{0\le u\le 1,c_l\le\vec{c}\le c_u} |\varphi\cdot\mathcal{W}u|_1 + \mu\langle u,\xi(\vec{c})\rangle + \mu\langle \mathbf{1},\psi(\vec{c})\rangle, \tag{3.12}$$

where

- Gaussian noise:

$$\begin{cases} \xi(\vec{c}) = (c_1 - f)^2 - (c_2 - f)^2, \\ \psi(\vec{c}) = (c_2 - f)^2; \end{cases} \tag{3.13}$$

- Poisson noise:

$$\begin{cases} \xi(\vec{c}) = (c_1 - c_2) - f(\log c_1 - \log c_2), \\ \psi(\vec{c}) = c_2 - f \log c_2. \end{cases} \tag{3.14}$$

For the weighted sparse regularization term $|\varphi \cdot \mathcal{W}u|$ in (3.12), a usual choice for φ is the edge indicator of f in (2.14), as proposed in many references [6, 3, 21]. However, this is not a good choice for automatically tracking the moving boundary of the ROI in sequential frames, as shown in section 6, when we consider segmentation of the video sequence f_t for $t = 1, \ldots, T$. In [15], manually drawn contours for each frame of a video are also proposed to be used as shape priors. However, it is not practical to trace all the frames, especially when online application is required. In order to extract the moving boundary Σ_t automatically from f_t, we adopt a shape prior function φ_t defined directly from previous segmentation results. In fact, it is reasonable to assume that the relative changes of the boundaries in the neighboring frames are minor. The weight term $|\varphi_t \cdot \mathcal{W}u_t|_1$ should be able to force the shape of underlying segmentation u_t close to u_{t-1} or even to previous segmentation. In other words, the edge of the resulting partition is dominated partially by the approximated shape prior and the regularity of the current frame. More specifically, the weight function $\varphi_t(x)$ at time t is chosen as

$$\varphi_t(x) = \max_{1 \le j \le M} \{\omega(d_{t-j}(x))\}, \tag{3.15}$$

where $d_{t-j}(x)$ denotes the distance function to the boundary $\partial\Sigma_{t-j}$ found at the previous jth frame, and $\omega(y)$ is a monotone increasing function with respect to y generally. Such a function enforces that the current regularized shape is close to the previous shape, as bigger weights will result in zero wavelet coefficients, when one moves away from the shape prior. For example, power or logarithm functions can be utilized. In fact the choice of the function type is not truly important, while the parameter choices are more important so as to result in a similar weight curve. In this paper, a power function $w(y) = ay^b$ for some positive a and b will be used in the numerical experiments. A good shape function should have a relatively small value near the boundary (represented by the distance function) and increase quickly when one moves away from the shape. The weight can be viewed as a smoothed narrow band function. The parameters a and b of the power function φ determine the width of narrow band. Empirically, we find that the max function over time leads to a more stable result compared to the mean one through the experiments. An example of the weight function of two sequential frames and the resulting weight function defined as (3.15) are shown in Figure 1.

4. Algorithm. In general, an alternating minimization scheme (or block coordinate descent) is adopted to solve the problems (3.12), as done in the literature (e.g., [19, 14, 40]). The convergence of such a scheme cannot be guaranteed in general. Inspired by recent work on the so-called Kurdyka–Łjasiewicz (KL) property [5, 41, 18, 4], we adopt the PAM proposed

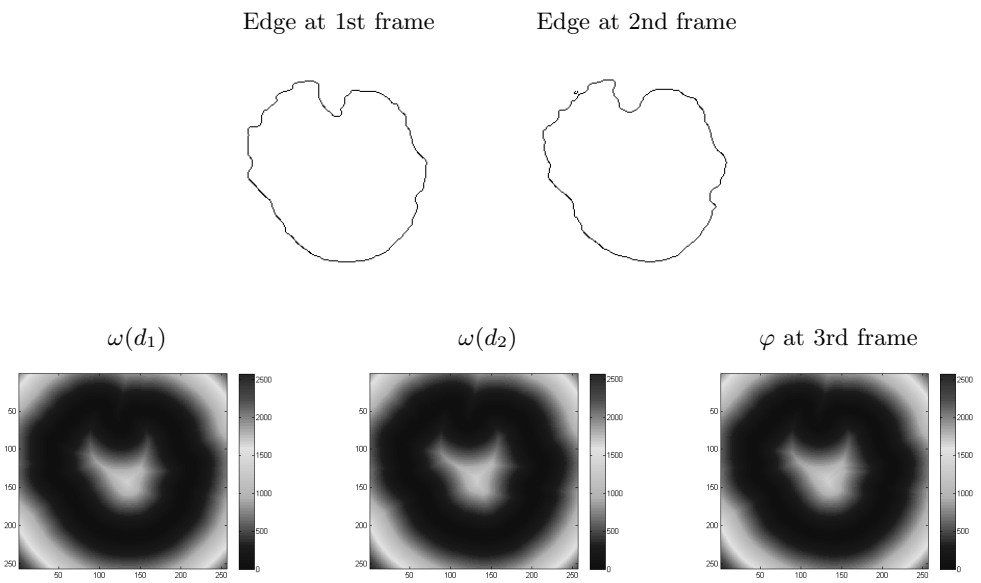

Figure 1. *An example of the shape prior, corresponding $w(d_i)$, and weight function φ.*

in [5], and we will verify the global convergence of the proposed scheme for the model (3.12) in section 5.

The PAM algorithm for solving (3.12) is given as

$$\begin{cases} u^{k+1} = \underset{0\le u\le 1}{\operatorname{argmin}} |\varphi\cdot\mathcal{W}u|_1 + \mu\langle u, \xi(\vec{c}^{\,k})\rangle + \dfrac{1}{2\sigma}\|u-u^k\|^2, \\ \vec{c}^{\,k+1} = \underset{c_l\le\vec{c}\le c_u}{\operatorname{argmin}} \mu\langle \mathbf{1}, \psi(\vec{c})\rangle + \mu\langle u^{k+1}, \xi(\vec{c})\rangle + \dfrac{1}{2\gamma}\|\vec{c}-\vec{c}^{\,k}\|^2, \end{cases} \tag{4.1}$$

where k denotes the iteration number, and σ and γ are two positive numbers.

The subproblem for solving u^{k+1} is convex and can be solved by the split Bregman method [25, 10], which is also equivalent to the well-known alternating direction of multiplier method (cf. e.g., [24, 23, 22]). This subproblem is reformulated as

$$\min_{0\le u\le 1} |\varphi\cdot\mathcal{W}u|_1 + \mu\langle u, \tilde{\xi}\rangle + \frac{1}{2\sigma}\|u-\tilde{u}\|^2, \tag{4.2}$$

where we set $\tilde{\xi} = \xi(c_1^k, c_2^k)$ and $\tilde{u} = u^k$.

By introducing the variable $v = \mathcal{W}u$, the split Bregman algorithm for solving (4.2) is

$$\begin{cases} u^{,q+1} = \underset{0\le u\le 1}{\operatorname{argmin}} \mu\langle u, \tilde{\xi}\rangle + \dfrac{\rho}{2}\|\mathcal{W}u - v^{,q} + b^{,q}\|_2^2 + \dfrac{1}{2\sigma}\|u-\tilde{u}\|^2, \\ v^{,q+1} = \underset{v}{\operatorname{argmin}} |\varphi\cdot v|_1 + \dfrac{\rho}{2}\|\mathcal{W}u^{,q+1} - v + b^{,q}\|_2^2, \\ b^{,q+1} = b^{,q+1} - (v^{,q+1} - \mathcal{W}u^{,q+1}), \end{cases} \tag{4.3}$$

where the outer iteration superscript k is omitted and superscript q denotes the inner iteration.

It is now standard to get the solution of the second subproblem in (4.3) by soft thresholding. Its formulation is

$$v^{\cdot,q+1} = \text{shrink}\left(\mathcal{W}u^{\cdot,q+1} + b^{\cdot,q}, \frac{\varphi}{\rho}\right), \tag{4.4}$$

where the pointwise soft shrinkage operator *shrink* is defined as

$$\text{shrink}(y,\nu) = \text{sgn}(y)\cdot\max\{|y|-\nu, \mathbf{0}\}. \tag{4.5}$$

Lemma 4.1. *The solution to the first subproblem of* (4.3) *is given by*

$$u^{\cdot,q+1} = \Pi_{[0,\ \ 1]}\left(\frac{\sigma\rho\mathcal{W}^T(v^{\cdot,q}-b^{\cdot,q}) - \sigma\mu\tilde{\xi} + \tilde{u}}{\sigma\rho+1}\right), \tag{4.6}$$

where Π_C *is the the projection operator onto the convex set* C.

Proof. Using the Karush–Kuhn–Tucker optimality conditions for the first subproblem in (4.3), we have

$$\begin{cases} -u \le \mathbf{0},\ u-1 \le \mathbf{0}, \\ \lambda_1 \ge \mathbf{0},\ \lambda_2 \ge \mathbf{0}, \\ \lambda_1\cdot(-u) = \mathbf{0},\ \lambda_2\cdot(u-1) = \mathbf{0}, \\ \mu\xi + \rho\mathcal{W}^T(\mathcal{W}u - v^{\cdot,q} + b^{\cdot,q}) + \frac{1}{\sigma}(u-\tilde{u}) - \lambda_1 + \lambda_2 = 0. \end{cases} \tag{4.7}$$

By using $\mathcal{W}^T\mathcal{W} = I$, the last equation yields

$$u^{\cdot,q+1} = \frac{\sigma\rho\mathcal{W}^T(v^{\cdot,q}-b^{\cdot,q}) - \sigma\mu\xi + \tilde{u}}{\sigma\rho+1} + \frac{\sigma}{\sigma\rho+1}(\lambda_1-\lambda_2) \tag{4.8}$$

$$:= \overline{u} + \frac{\sigma}{\sigma\rho+1}(\lambda_1-\lambda_2), \tag{4.9}$$

where we denote $\overline{u} = \frac{\sigma\rho\mathcal{W}^T(v^{\cdot,q}-b^{\cdot,q}) - \sigma\mu\tilde{\xi}+\tilde{u}}{\sigma\rho+1}$. Thus, for all $x\in\Omega$, we have the following pointwise relations:

- If $0 \le \bar{u}(x) \le 1$, then

$$\lambda_1(x) = 0, \quad \lambda_2(x) = 0, \quad u^{\cdot,q+1}(x) = \bar{u}(x).$$

- If $\bar{u}(x) < 0$, then

$$\lambda_1(x) = -\frac{\sigma\rho+1}{\sigma}\bar{u}(x), \quad \lambda_2(x) = 0, \quad u^{\cdot,q+1}(x) = 0.$$

- If $\bar{u}(x) > 1$, then

$$\lambda_1(x) = 0, \quad \lambda_2(x) = \frac{\sigma\rho+1}{\sigma}(\bar{u}(x)-1), \quad u^{\cdot,q+1}(x) = 1.$$

It follows that $u^{\cdot,q+1}$ is given as (4.6). ∎

In summary, the subproblem for u in (4.1) can be updated by the following scheme:

$$\begin{cases} u^{\cdot,q+1} = \mathbf{\Pi}_{[0\ 1]}\left(\frac{\sigma\rho\mathcal{W}^T(v^{\cdot,q}-b^{\cdot,q})-\sigma\mu\tilde{\xi}+\tilde{u}}{\sigma\rho+1}\right), \\ v^{\cdot,q+1} = \operatorname{shrink}(\mathcal{W}u^{\cdot,q+1}+b^{\cdot,q}, \frac{\varphi}{\rho}), \\ b^{\cdot,q+1} = b^{\cdot,q} + (\mathcal{W}u^{\cdot,q+1} - v^{\cdot,q+1}). \end{cases} \tag{4.10}$$

For the $\vec{c}^{k+1}$ subproblem, we need to solve for

$$\operatorname*{argmin}_{c_l\le\vec{c}\le c_u} \mu\langle\mathbf{1},\psi(\vec{c})\rangle + \mu\langle\tilde{u},\xi(\vec{c})\rangle + \frac{1}{2\gamma}\|\vec{c}-\tilde{\vec{c}}\|^2. \tag{4.11}$$

Here we abuse notation by setting $\tilde{u} := u^{k+1}$ and $\tilde{\vec{c}} := \vec{c}^k$.

Recall that the terms $\xi(\vec{c})$ and $\psi(\vec{c})$ are given as (3.13) and (3.14) for the case of Gaussian and Poisson noise, respectively. It is not hard to see that (4.11) is convex for both c_1 and c_2 for both cases. In fact, the closed-form solutions can be obtained for both cases.

For Poisson noise, the functions $\xi(\vec{c})$ and $\psi(\vec{c})$ are defined in (3.14). A closed-form solution for solving (4.11) is given by the following lemma.

Lemma 4.2. *For $\xi(\vec{c})$ and $\psi(\vec{c})$ defined as (3.14), the solution to (4.11) is given as*

$$\begin{cases} c_1^{k+1} = \mathbf{\Pi}_{[c_l,\ \ c_u]}\left(\frac{-\gamma\mu\langle\mathbf{1},\tilde{u}\rangle+\tilde{c}_1+\sqrt{(\gamma\mu\langle\mathbf{1},\tilde{u}\rangle-\tilde{c}_1)^2+4\gamma\mu\langle f,\tilde{u}\rangle}}{2}\right), \\ c_2^{k+1} = \mathbf{\Pi}_{[c_l,\ \ c_u]}\left(\frac{-\gamma\mu\langle\mathbf{1},\mathbf{1}-\tilde{u}\rangle+\tilde{c}_2+\sqrt{(\gamma\mu\langle\mathbf{1},\mathbf{1}-\tilde{u}\rangle-\tilde{c}_2)^2+4\gamma\mu\langle f,\mathbf{1}-\tilde{u}\rangle}}{2}\right). \end{cases} \tag{4.12}$$

Proof. Recall that $\xi(\vec{c}) = (c_1-c_2) - f(\log c_1 - \log c_2), \psi(\vec{c}) = c_2 - f\log c_2$. The Karush–Kuhn–Tucker optimality conditions of problem (4.11) are

$$\begin{cases} -\vec{c}+c_l \le \mathbf{0},\ \ \vec{c}-c_u \le \mathbf{0}, \\ \lambda_1 \ge \mathbf{0},\ \ \lambda_2 \ge \mathbf{0}, \\ \lambda_1\cdot(-\vec{c}+c_l) = 0,\ \ \lambda_2\cdot(\vec{c}-c_u) = 0, \\ \mu(\langle\tilde{u},\mathbf{1}\rangle - \frac{1}{c_1}\langle\tilde{u},f\rangle) + \frac{1}{\gamma}(c_1-\tilde{c}_1) - \lambda_{11} + \lambda_{12} = 0, \\ \mu(\langle 1-\tilde{u},\mathbf{1}\rangle - \frac{1}{c_2}\langle 1-\tilde{u},f\rangle) + \frac{1}{\gamma}(c_2-\tilde{c}_2) - \lambda_{12} + \lambda_{22} = 0, \end{cases} \tag{4.13}$$

where

$$\lambda_1 = \begin{pmatrix}\lambda_{11}\\ \lambda_{12}\end{pmatrix} \quad \text{and} \quad \lambda_2 = \begin{pmatrix}\lambda_{21}\\ \lambda_{22}\end{pmatrix}$$

are Lagrangian multipliers. Denote the solution of (4.11) as (c_1^*, c_2^*). From the fourth equation, we have

$$\frac{1}{\gamma c_1}(c_1^2 + \alpha c_1 + \beta) = \lambda_{11} - \lambda_{12},$$

where $\alpha = \gamma\mu\langle\mathbf{1},\tilde{u}\rangle - \tilde{c}_1$ and $\beta = -\gamma\mu\langle f,\tilde{u}\rangle$. For $c_1^2+\alpha c_1+\beta = 0$, the positive solution is $\bar{c_1} = \frac{-\alpha+\sqrt{\alpha^2-4\beta}}{2} = \frac{-\gamma\mu\langle\mathbf{1},\tilde{u}\rangle+\tilde{c}_1+\sqrt{(\gamma\mu\langle\mathbf{1},\tilde{u}\rangle-\tilde{c}_1)^2+4\gamma\mu\langle f,\tilde{u}\rangle}}{2}$.

In the following, we discuss the solution c_1^* for different cases of $\bar{c}_1$.

- If $0 < c_l \le \bar{c}_1 \le c_u$, we have $\frac{\bar{c}_1^2+\alpha\bar{c}_1+\beta}{\gamma\bar{c}_1} = 0$, $\lambda_{11} = 0$, and $\lambda_{12} = 0$; then the solution is $c_1^* = \bar{c}_1$.
- If $\bar{c}_1 < c_l$, we have $\frac{c_l^2+\alpha c_l+\beta}{\gamma c_l} > 0$, $\lambda_{11} > 0$, and $\lambda_{12} = 0$. Thus $c_1^* = c_l$.
- If $\bar{c}_1 > c_u > 0$, we have $\frac{c_u^2+\alpha c_u+\beta}{\gamma c_u} < 0$, $\lambda_{11} = 0$, and $\lambda_{12} > 0$. Thus $c_1^* = c_u$.

Thus, the solution $c_1^* = \Pi_{[c_l,\, c_u]}(\bar{c}_1)$. A similar argument can be applied to c_2, and it follows that the solutions are given as (4.12). ∎

For Gaussian noise with $\xi(\vec{c}) = (c_1 - f_t)^2 - (c_2 - f)^2$, $\psi(\vec{c}) = (c_2 - f)^2$, it is easy to see that the solution of (4.11) is given as

$$\begin{cases} c_1^{k+1} = \Pi_{[c_l,\, c_u]}\Big(\dfrac{2\mu\gamma\langle f, \tilde{u}\rangle + \tilde{c}_1}{2\mu\gamma\langle \mathbf{1}, \tilde{u}\rangle + 1}\Big), \\ c_2^{k+1} = \Pi_{[c_l,\, c_u]}\Big(\dfrac{2\mu\gamma\langle f, 1-\tilde{u}\rangle + \tilde{c}_2}{2\mu\gamma\langle \mathbf{1}, \mathbf{1}-\tilde{u}\rangle + 1}\Big). \end{cases} \tag{4.14}$$

The overall scheme for individual image segmentation is summarized in Algorithm 1.

Algorithm 1 Proximal alternating minimization algorithm based individual image segmentation algorithm.

Step 1. Set the initial value with $u^{0,0}$, $c_1^{0,0}, c_2^{0,0}$ and set tolerance value of the stop tolerance δ_1, δ_2;
Step 2.
while $\|u^k - u^{k-1}\|_2 > \delta_1$ or $K = 1$ **do**
  Set $u^{k+1,0} := u^k$
  while $\|u^{k+1,q} - u^{k+1,q-1}\|_2 > \delta_2$ or $k = 1$ **do**
    Update $(u^{k+1,q+1}, v^{k+1,q+1}, b^{k+1,q+1})$ with (4.10), where $\tilde{u} := u^k$;
    $q = q + 1$;
  end while
  Set $u^{k+1} := u^{k+1,q}$;
  Set $\tilde{u} := u^{k+1}$ and $\tilde{\vec{c}} := \vec{c}^k$;

  For Gaussian noise model, update $\vec{c}^{k+1}$ with (4.14).

  For Poisson noise model, update $\vec{c}^{k+1}$ with (4.12).
  $k = k + 1$;
end while
Step 3. Get the final segmentation $u = \mathbf{1}_{u^k > 0.5}$.

We now present the algorithm for processing the whole video sequence. In particular, we incorporate the weight function $\varphi(x)$, as defined in (3.15), using the distance function to a predefined shape. For the first frame f_1, we usually adopt a hand drawn initial shape and compute the distance function. After we compute the current segmentation u_t by Algorithm 1, we obtain $\Sigma_t := \{u_t(x) = 0\}$. The fast sweeping method [43] is adopted here to obtain the distance function by solving the PDE $|\nabla d(x)| = 1$ with the boundary condition $d(x) = 0, x \in$

$\partial\Sigma_t$. For completeness of presentation, we present the details in Algorithm 2. A distance function in our numerical results is illustrated in Figure 1.

Algorithm 2 Fast sweeping method for computing distance function.

Step 1. Discretize using upwind difference.

$$[(d_{i,j} - d_{xmin})^+]^2 + [(d_{i,j} - d_{ymin})^+]^2 = h^2, \tag{4.15}$$

where $d_{xmin} = \min(d_{i-1,j}, d_{i+1,j}), d_{ymin} = \min(d_{i,j-1}, d_{i,j+1})$.

Step 2. Initialize d to enforce $d(x) = 0, x \in \Gamma$.

(1) Assign exact values at grid points in or near Γ which will be fixed in later computation.

(2) Assign large positive value at all other points.

Step 3: Gauss–Seidel iteration using different sweeping order:

$$\begin{aligned} &(1)\ \ i = 1 : I, j = 1 : J; \qquad (2)\ \ i = I : 1, j = 1 : J; \\ &(3)\ \ i = I : 1, j = J : 1; \qquad (4)\ \ i = 1 : I, j = J : 1. \\ &d_{i,j}^{new} = \min(d_{i,j}^{old}, \bar{d}), \end{aligned} \tag{4.16}$$

where d is the solution to (4.15) using current $d_{i-1,j}, d_{i+1,j}, d_{i,j-1}, d_{i,j+1}$.

Finally, the whole procedure for video segmentation is present in Algorithm 3.

Algorithm 3 Video segmentation algorithm.

Step 0. Given $\{f_1, \ldots, f_T\}$ and a (hand drawn) initial shape Σ_0 for f_1.

Step 1. For $t = 1, \ldots, T$

Step 1.0. Set $u^0 = u_{t-1}$, $c_1^{0,0}, c_2^{0,0}$ as the mean intensity of initial region determined by u_{t-1}.

Step 1.1. Compute the distance function d_{t-1} and φ^k with (3.15).

Step 1.2. Apply Algorithm 1 to get final segmentation u_t and Σ_t.

5. Convergence analysis.

The convergence of the proximal alternating scheme (4.1) (Algorithm 1) is established based on the framework presented in [2].

Consider a class of nonconvex and nonsmooth problems of the form

$$L(x, y) = f(x) + Q(x, y) + g(y), \tag{5.1}$$

where

$$(H_1): \begin{cases} f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\},\ g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \text{ are proper l.s.c.;} \\ Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \text{ is a } C^1 \text{ function;} \\ \nabla Q \text{ is Lipschitz continuous on bounded subsets of } R^n \times R^m. \end{cases} \tag{5.2}$$

This class of optimization problems arises in a variety of applications, such as compressed sensing, matrix factorization, dictionary learning, sparse approximations of signals and images,

and blind decomposition. The PAM algorithm has been proposed in [5, 2] for solving this broad class of problems in the form (5.1)

$$\text{(5.3)}\qquad \begin{cases} x^{k+1} = \operatorname{argmin}\ L(x, y^k) + \frac{1}{2\theta^k}\|x - x^k\|^2, \\ y^{k+1} = \operatorname{argmin}\ L(x^{k+1}, y) + \frac{1}{2\eta^k}\|y - y^k\|^2. \end{cases}$$

A standard assumption for the algorithm is

$$\text{(5.4)}\quad (H_2): \begin{cases} \inf_{\mathbb{R}^n\times\mathbb{R}^m} L > -\infty, \\ \text{the function } L(\cdot, y_0) \text{ is } \textit{proper}, \\ \text{for some positives } l, u \text{ and } l < u, \text{ the sequences of stepsizes } \theta_k, \eta_k \text{ satisfy} \\ \quad l < \theta_k, \eta_k < u \text{ for all } k > 0. \end{cases}$$

An important tool for demonstrating the convergence of the sequence is based on the so-called Kurdyka–Łojasiewicz (KL) property. We first list some useful definitions and results. For more details on the definitions and properties, we refer the reader to the series of work [2, 5] and references therein.

Definition 5.1. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a proper l.s.c. function.*

1. *The domain of f is defined as* $\operatorname{dom}(f) := \{x \in \mathbb{R}^n : f(x) < \infty\}$.
2. *For each $x \in \operatorname{dom}(f)$, the Fréchet subdifferential of f at x is defined as follows:*

$$\text{(5.5)}\qquad \hat{\partial} f(x) = \left\{ s : \liminf_{y\to x, y\neq x} \frac{1}{\|x - y\|}(f(y) - f(x) - \langle s, y - x\rangle) \geq 0 \right\}.$$

   If $x \notin \operatorname{dom} f$, we set $\hat{\partial} f(x) = \emptyset$.
3. *The (limiting) subdifferential of f at $x \in \operatorname{dom}(f)$ is defined as follows:*

$$\text{(5.6)}\qquad \partial f(x) = \{x^* : \exists x_n \to x \text{ such that } f(x_n) \to f(x), s_n \in \hat{\partial} f(x_n) \to x^*\}.$$

   Denote the domain of ∂f as $\operatorname{dom}(\partial f) = \{x : \partial f \neq \emptyset\}$.
4. *A point x is called a critical point of f if $0 \in \partial f(x)$.*

Definition 5.2. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a proper l.s.c. function. f is said to have the KL property at $\bar{x} \in \operatorname{dom}(\partial f)$ if there exist $\eta > 0$, a neighborhood U of $\bar{x}$, and a continuous concave function $\phi : [0, \eta) \to R_+$ such that*

- $\phi(0) = 0$,
- *ϕ is C^1 on $(0, \eta)$,*
- *for all $s \in (0, \eta)$, $\phi'(s) > 0$,*
- *for all $x \in U \cap [f(\bar{x} < f(x) < f(\bar{x}) + \eta)]$, the KL inequality holds,*

$$\phi'(f(x) - f(\bar{x}) dist(0, \partial f(x)) > 1,$$

where $dist(0, \partial f(x)) := \min\{\|y\| : y \in \partial f(x)\}$.

This property was introduced in [28, 26] and is used in some recent work for establishing the global convergence of the proximal alternating scheme; see, for instance, [2, 5, 41]. However, it is not trivial to verify the conditions listed above, and thus there are some special classes of

functions which are proven to be KL functions, such as real analytic functions, semi-algebraic functions, etc.

Definition 5.3. *A smooth function $\phi(t)$ on $\mathbb{R}$ is analytic if $(\frac{\phi^{(k)}(t)}{k!})^{1/k}$ is bounded for all k and on any compact set $D \subset \mathbb{R}$. For a real function $\psi(x)$ on $\mathbb{R}^n$, one can verify the analyticity of ψ by checking the analyticity of $\phi(t) := \psi(x + ty)$ for any $x, y \in \mathbb{R}^n$.*

Definition 5.4. *A subset S of $\mathbb{R}^n$ is called the semi-algebraic set if there exists a finite number of real polynomial functions g_{ij}, h_{ij} such that*

$$S = \bigcup_j \bigcap_i \{x \in \mathbb{R}^n : g_{ij}(x) = 0, h_{ij}(x) < 0\}. \tag{5.7}$$

A function $f(x)$ is called the semi-algebraic function if its graph $\{(x, t) \in \mathbb{R}^n \times \mathbb{R}, t = f(x)\}$ is a semi-algebraic set in $\mathbb{R}^{n+1}$.

Lemma 5.5 (Lemma 5 in [2]). *Under assumptions (5.2) and (5.4), the sequences (x_k, y_k) in (5.3) are correctly defined. Moreover, the following estimation holds:*

(i) *The sequence $L(x_k, y_k)$ is nonincreasing and*

$$L(x_k, y_k) + \frac{1}{2\theta_k}\|x_k - x_{k-1}\|^2 + \frac{1}{2\mu_{k-1}}\|y_k - y_{k-1}\|^2 \le L(x_{k-1}, y_{k-1}) \quad \text{for all } k \ge 1,$$

(ii) *(square summable)* $\sum_{k=1}^{\infty}(\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2) < \infty$; *hence*

$$\lim_{k\to\infty} \|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2 \to 0.$$

Theorem 5.6 (Theorem 9 in [2]). *Assume that (5.2) and (5.4) hold. The sequence $(x^{(k)}, y^{(k)})$ generated by the PAM algorithm (5.3) converges to a critical point of (5.1) if the following conditions hold:*

(i) *$L(x, y)$ is a KL function;*

(ii) *$(x^k, y^k), k = 1, 2, \ldots,$ is a bounded sequence.*

We will demonstrate the convergence of the alternating algorithm (4.1) by verifying the conditions listed above. Recall the single image segmentation model (3.12)

$$L(u, \vec{c}) = |\varphi \cdot \mathcal{W}u|_1 + \chi_{0\le u\le 1} + \chi_{c_l\le \vec{c}\le c_u} + \langle u, \xi(c_1, c_2)\rangle + \langle \mathbf{1}, \psi(\vec{c})\rangle. \tag{5.8}$$

We rewrite it as a separable form

$$L(u, \vec{c}) = g(u) + Q(u, \vec{c}) + h(\vec{c}),$$

where

$$g(u) = |\varphi \cdot \mathcal{W}u|_1 + \chi_{0\le u\le 1}; \tag{5.9}$$

$$h(\vec{c}) = \langle \mathbf{1}, \psi(\vec{c})\rangle + \chi_{c_l\le \vec{c}\le c_u}; \tag{5.10}$$

$$Q(u, \vec{c}) = \langle u, \xi(\vec{c})\rangle. \tag{5.11}$$

We will verify the properties of the objective function $L(u,\vec{c})$ to justify the convergence of the consequence $(u^K,\vec{c}^K)$ under the above framework. We first note that

$$\mathrm{dom}(L)=\mathrm{dom}(\partial L)=[0,1]^n\times[c_l,c_u]^2$$

is bounded and compact.

Lemma 5.7. *For $Q(u,\vec{c})$ defined as in (5.11), its gradient is Lipschitz continuous on any bounded set $\subset \mathrm{dom}(L)$.*

Proof. Recall that $Q(u,\vec{c})=\langle u,\xi(c_1,c_2)\rangle$, and its gradient is given as

$$\nabla Q(u,c)=\begin{bmatrix}\xi(c_1,c_2)\\ \left\langle u,\frac{\partial\xi(c_1,c_2)}{\partial c_1}\right\rangle\\ \left\langle u,\frac{\partial\xi(c_1,c_2)}{\partial c_2}\right\rangle\end{bmatrix}, \tag{5.12}$$

where

$$\frac{\partial\xi(c_1,c_2)}{\partial c_1}=\begin{cases}2(c_1-f) & \text{Gaussian noise},\\ 1-\frac{f}{c_1} & \text{Poisson noise},\end{cases}\qquad \frac{\partial\xi(c_1,c_2)}{\partial c_2}=\begin{cases}-2(c_2-f) & \text{Gaussian noise},\\ -1+\frac{f}{c_2} & \text{Poisson noise}.\end{cases} \tag{5.13}$$

The corresponding Hessian matrices are

$$H=\begin{bmatrix}0 & 2(c_1-f) & -2(c_2-f)\\ 2(c_1-f)^\top & 2\langle u,\mathbf{1}\rangle & 0\\ -2(c_2-f)^\top & 0 & -2\langle u,\mathbf{1}\rangle\end{bmatrix} \tag{5.14}$$

for Gaussian noise and

$$H=\begin{bmatrix}0 & 1-\frac{f}{c_1} & -1+\frac{f}{c_2}\\ 1-\frac{f^\top}{c_1} & \langle u,\frac{f}{c_1^2}\rangle & 0\\ -1+\frac{f^\top}{c_2} & 0 & -\langle u,\frac{f}{c_2^2}\rangle\end{bmatrix} \tag{5.15}$$

for Poisson noise, respectively.

Since $\mathrm{dom}(L)$ is bounded and $c_l>0$, we can see that both (5.14) and (5.15) are bounded on any bounded subset. Namely, $\nabla Q(u,\vec{c})$ in (5.8) is Lipschitz continuous on any bounded set. ∎

Proposition 5.8. *Let $(u^k,\vec{c}^k)$ be the updated sequence generated by (4.1); then*

(i) *(energy nonincreasing) the sequence $L(u^k,\vec{c}^k)$, where $L(u,\vec{c})$ is defined as (5.8), is nonincreasing.*

(ii) *(square summable)* $\lim_{k\to\infty}\|x_k-x_{k-1}\|^2+\|y_k-y_{k-1}\|^2\to 0$.

Proof. By Lemma 5.5, we need only check whether assumptions (5.2) and (5.4) are satisfied. Both $g(u)$ and $h(\vec{c})$ are proper l.s.c., and $Q(u,\vec{c})$ is C^1. By Lemma 5.7, $Q(u,\vec{c})$ has a Lipschitz continuous gradient. Thus assumption (5.2) holds. Since the constraint sets for $L(u,\vec{c})$ are bounded, it is easy to see that the first two conditions in (5.4) hold. The third condition in (5.4) can be satisfied by choosing the algorithm parameters properly. ∎

Lemma 5.9. *The objective function $L(u,\vec{c})$ defined in (5.8) is a KL function on* $\mathrm{dom}(\partial L)$.

Proof. We show that all three terms in $L(u,\vec{c})$ are real analytic functions by [2]. Note that $h(\vec{c}) = \langle \mathbf{1}, \psi(\vec{c})\rangle + \mathcal{C}_{c_l \le \vec{c} \le c_u}$. The indicator function is analytic. For Gaussian noise and the Poisson case, it is easy to see that $\psi(\vec{c})$ is real analytic on $\mathrm{dom}(\partial L)$. For $Q(u,\vec{c}) = \langle u, \xi(c_1,c_2)\rangle$, we discuss the two cases. For Gaussian noise, $\xi(\vec{c}) = (c_1 - f)^2 - (c_2 - f)^2$, and it is easy to see that $Q(u,\vec{c})$ is a semi-algebraic function. For Poisson noise, $\xi(\vec{c}) = (c_1 - c_2) - f \log \frac{c_1}{c_2}$, and the function is real analytic on $\mathrm{dom}(L)$. Finally, the term $|\varphi \cdot \mathcal{W}u|_1$ is also a semi-algebraic function. Finally, we conclude that $L(u,\vec{c})$ is a KL function on $\mathrm{dom}(\partial L)$. ∎

Theorem 5.10 (sequence convergence). *The sequence $(u^k, \vec{c}^k)$ generated by (4.1) converges to a critical point of $L(u,\vec{c})$.*

Proof. By the proof of Proposition 5.8, we can see that the function $L(u,\vec{c})$ (5.8) satisfies assumptions (5.2) and (5.4). Due to the bounded constraints on u and c, the sequences are bounded. Together with Lemma 5.9 and Theorem 5.6, the theorem is proved. ∎

6. Numerical experiments.

6.1. Methods in comparison. For the reason of comparison, we also consider the level set implementation for a similar variation model for single image segmentation,

$$\min_{\phi,\ c_l \le \vec{c} \le c_u} \ |\varphi \cdot \nabla H(\phi)|_1 + \mu \langle H(\phi), \xi(\vec{c})\rangle + \mu \langle \mathbf{1}, \psi(\vec{c})\rangle, \tag{6.1}$$

where ϕ is the usual level set function for defining the boundary, and ξ, ψ are defined as in (3.13) and (3.14) for Gaussian and Poisson noise, respectively. For $\vec{c}$ fixed, a gradient flow scheme as in [14, 31] for solving (6.1) is given as

$$\begin{cases} \frac{\partial \phi}{\partial t} = & \delta_\epsilon(\phi)[\varphi \mathbf{div}(\frac{\nabla \phi}{|\nabla \phi|}) + \mu \xi(c_1,c_2)] \ \text{ in } \ (0,\infty)\times\Omega, \\ & \phi(0,x,y) = \phi_0(x) \ \text{ in } \ \Omega, \\ & \frac{\delta_\epsilon(\phi)}{|\nabla \phi|} \frac{\partial \phi}{\partial \vec{n}} = 0 \ \text{ on } \ \partial\Omega. \end{cases} \tag{6.2}$$

Once ϕ is fixed, the region Σ is obtained as $\{\phi > 0\}$, and it is easy to see that the mean intensity for each region yields the optimal solution of (6.1) with respect to $\vec{c}$ for both noise cases, as in the classical Chan–Vese model.

We thus compare the above level set formulation (6.1) with the proposed wavelet frame model (3.12). For the two models, we solve with the above gradient flow (6.2) and the proposed scheme (4.1) (Algorithm 1), respectively. Combining with different shape priors φ based on edge indicator (2.14) and distance function (3.15), and different noise statistics (Poisson (3.14) and Gaussian (3.13)), we investigate the five methods for individual image segmentation, as listed in Table 1.

6.2. Multilevel implementation. In the implementation of the wavelet frame method for single image segmentation, we use the piecewise linear B-spline wavelets with two levels of decomposition for $\mathcal{W}$. The corresponding filter masks chosen here are

$$h_0 = \left[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right], \quad h_1 = \left[-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}\right], \quad h_2 = \left[\frac{\sqrt{2}}{4}, 0, -\frac{\sqrt{2}}{4}\right].$$

Table 1
Methods in comparison.

Method	Model	Noise	Shape prior
M1	level set (6.1)	Poisson	(2.14)
M2	level set (6.1)	Poisson	(3.15)
M3	wavelet frame (3.12)	Poisson	(2.14)
M4	wavelet frame (3.12)	Gaussian	(3.15)
M5	wavelet frame (3.12)	Poisson	(3.15)

We also use the coarse-to-fine multilevel implementation to accelerate the computation speed. The procedure is to perform segmentation on a downsampled smoothed image of each frame f_t, and then the low-resolution segmentation results are upsampled and used as the initial u_0 for higher resolution image segmentation. For instance, each image of our video data set is 512×512 pixels, and we first run the segmentation for the resized image 64×64, and then resize the segmentation 128×128 as initial u_0, etc. Finally, we get the segmentation result for the fine scale image. An illustration is shown in Figure 2.

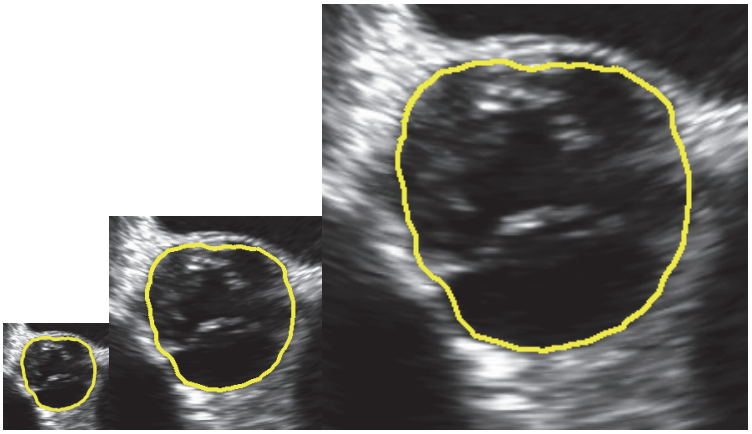

Figure 2. *The coarse-to-fine method.*

6.3. Results on real data sets. In this section, numerical experiments on real ultrasound video data are presented for the different methods listed in Table 1. The computations are performed on a PC with an Intel i5-3570 3.4GHZ CPU and 8GB memory.

The size of the first ultrasound video is $455 \times 505 \times 97$, and some frames are shown in Figure 3. For this data set, an initial shape prior Σ_0 used in Algorithm 3 for the first frame f_1 is a hand drawn contour by an expert, and then the sequential shape priors for other frames are computed according to (2.14) or (3.15).

For the distance function prior, we use the segmentation result of two previous frames and empirically set the function φ as

$$\varphi = \max\{s(d_{t-1})^{1.5}, s(d_{t-2})^{1.5}\},$$

where s is a parameter for a balance of efficiency and quality. In our tests, we determine

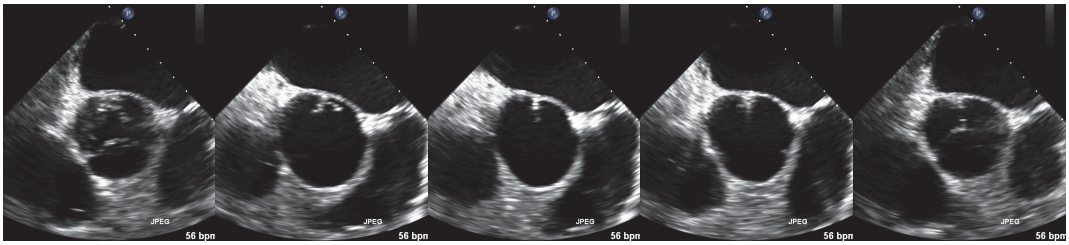

Figure 3. *Some frames of data set* 1.

whether the data set is of high quality; only one previous frame use can work sufficiently well. In fact, if more information from the previous frame is utilized, the model will be more stable, while the computation cost will increase. Thus the choice of sequential shape prior is a balance of efficiency and quality.

In Figure 4, we show the segmentation results for some selected frames, with hand drawn contours by an expert. These contours are considered the ground truths, as shown in the first row; in the second to sixth rows, we list the corresponding results with **M**1 to **M**5. Note that we have optimized the parameters for each method. We can see that both methods with edge function (2.14) as shape priors (**M**1 and **M**3) fail to track the ROI, and **M**2 and **M**4 may introduce some undesirable regions during the whole procedure of video segmentation. Once those regions appear, it will affect the sequential frames. Method **M**5, using the wavelet frame method, together with Poisson noise and distance prior, appears to be more stable for the contour tracking.

To quantitatively measure the segmentation quality, we use the Dice coefficient defined as

$$R = \frac{|\{x \in \Omega : u(x) = u_e(x)\}|}{|\Omega|}, \tag{6.3}$$

where $u_e(x)$ is the ground truth segmentation. The numerical results of this quantity for different methods are listed in Table 2, and we can see that the accuracy of **M**5 is higher for most of the frames.

To illustrate the relation between the solution of the relaxed model and the binarized solution, in Figure 5 we show the results of method **M**5 without Step 3 in Algorithm 1.

A more challenging data set of size $455 \times 505 \times 63$ is also tested and shown in Figure 6. The numerical results with different methods on this data set are shown in Figure 7. The contours tend to leak out, and it is harder to keep the topology for this data set. The comparisons for the five methods are similar to those for data set 1. We can see that the tight frame method with Poisson/Gaussian statistics with distance prior achieves the best performance, and the Poisson one achieves slightly better results for some frames. Further quantitative comparisons are presented in Table 3.

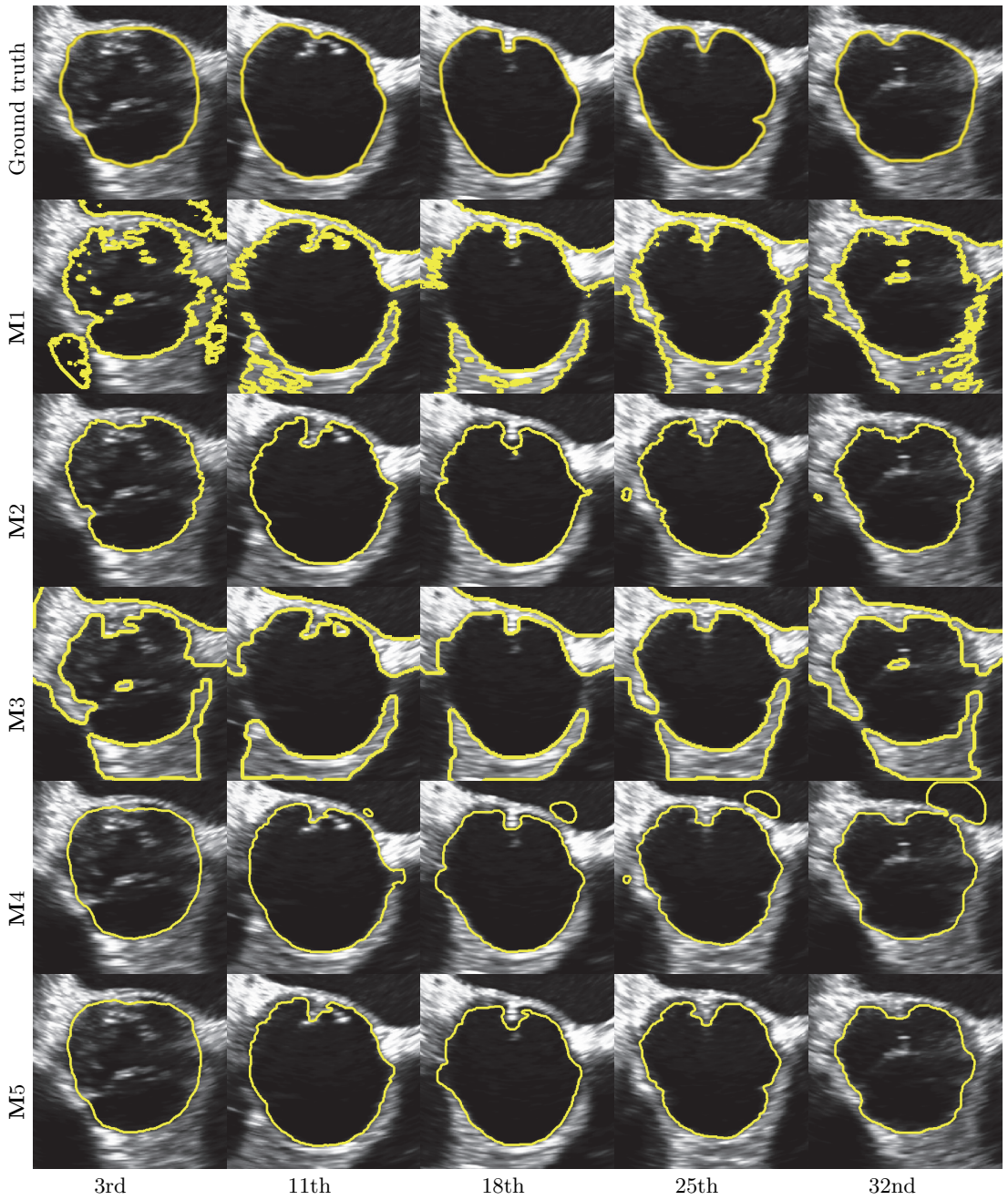

Figure 4. *First row: ground truths drawn by an expert. Second row:* $\boldsymbol{M}1$ *with* $\mu = 2.3$. *Third row:* $\boldsymbol{M}2$ *with* $s = 2.4$ *and* $\mu = 30.6$. *Fourth row:* $\boldsymbol{M}3$ *with* $\mu = 2.3$. *Fifth row:* $\boldsymbol{M}4$ *with* $s = 2.4$ *and* $\mu = 20.4$. *Sixth row:* $\boldsymbol{M}5$ *with* $s = 2.4$ *and* $\mu = 30.6$.

Table 2
Segmentation accuracy (%) of data set 1.

Method	3rd	11th	18th	25th	32nd
M1	67.70	64.49	62.07	63.67	62.44
M2	95.23	92.38	**94.48**	93.74	93.87
M3	63.47	67.44	64.81	66.10	64.40
M4	96.22	92.99	94.05	93.23	91.41
M5	**96.23**	**93.01**	94.20	**94.07**	**94.54**

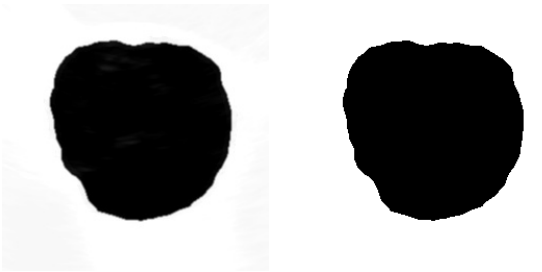

Figure 5. *The solution before and after the thresholding Step* 3 *in Algorithm* 1.

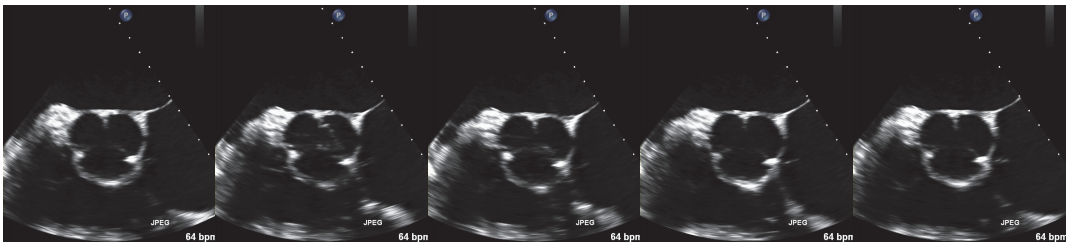

Figure 6. *Some frames of data set* 2.

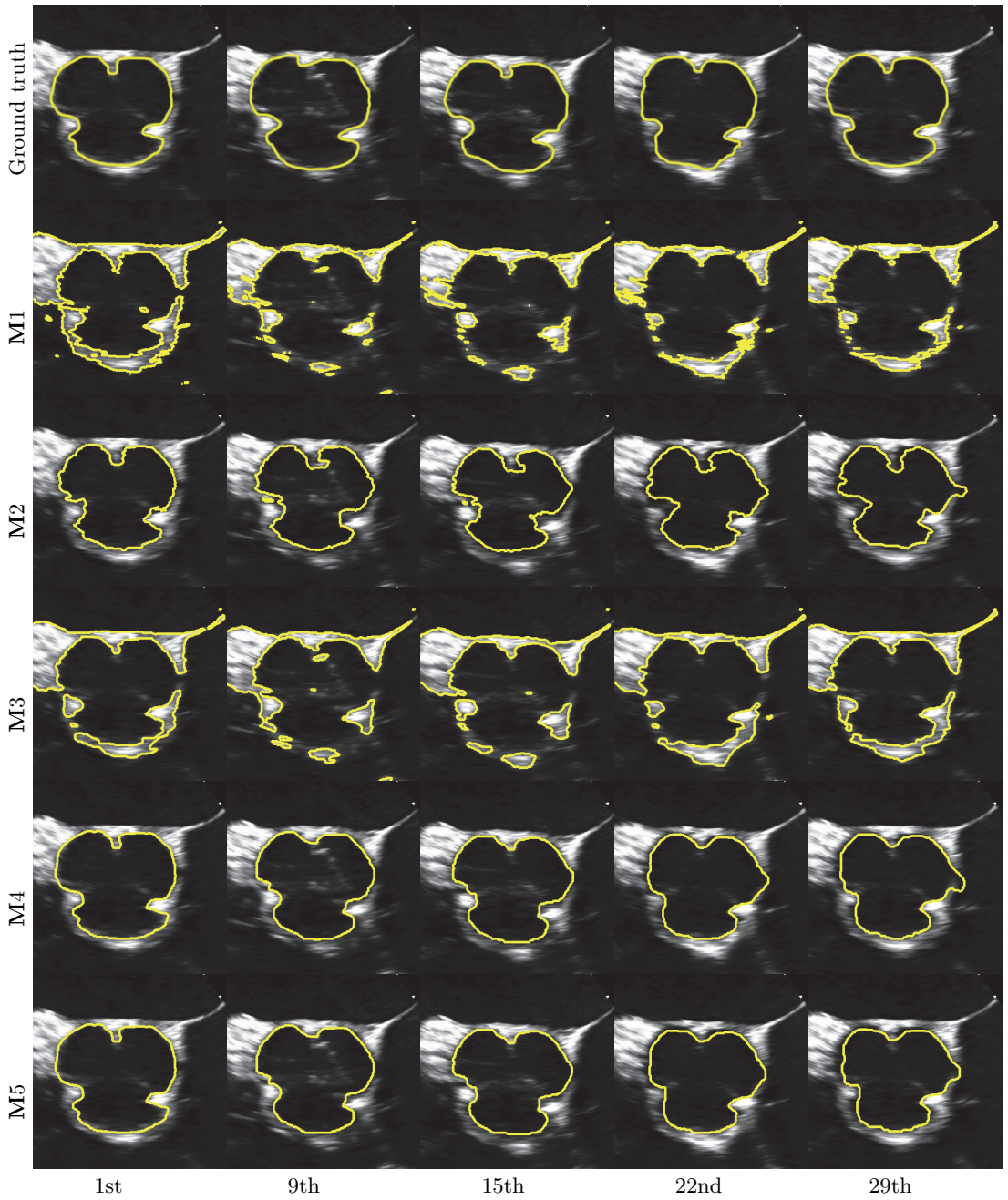

Figure 7. *First row: ground truths drawn by an expert. Second row:* $\boldsymbol{M}1$ *with* $\mu = 2.2$. *Third row:* $\boldsymbol{M}2$ *with* $s = 2.0$ *and* $\mu = 30.6$. *Fourth row:* $\boldsymbol{M}3$ *with* $\mu = 2.2$. *Fifth row:* $\boldsymbol{M}4$ *with* $s = 2.0$ *and* $\mu = 20.4$. *Sixth row:* $\boldsymbol{M}5$ *with* $s = 2.0$ *and* $\mu = 30.6$.

Table 3
Segmentation accuracy (%) of data set 2.

Method	1st	9th	15th	22nd	29th
M1	35.21	33.39	35.19	34.37	34.90
M2	96.38	95.91	95.16	95.08	93.56
M3	36.67	35.14	37.56	36.81	36.88
M4	96.36	96.58	**96.45**	**96.17**	95.03
M5	**96.41**	**96.59**	96.34	96.06	**95.19**

Table 4
Computation time for whole video segmentation (in seconds).

Method	Data set 1	Data set 2
M1	604.5	447.3
M2	511.5	365.4
M3	122.4	59.8
M4	80.7	55.0
M5	82.3	55.4

Finally, we present the computation time of the five methods for the whole video sequence in Table 4. We notice that the tight frame models **M**3, **M**4, and **M**5 reduce the computation time significantly compared to level set formulations. The multilevel implementation also reduces the computation time by a factor of 2. Thus, this model can be potentially applied in practice for real time tracking during operation navigation.

7. Conclusion. We have proposed a general wavelet frame based model for tracking the regions of interest (ROI). The proposed model allows for a general assumption for the noise present in the data and incorporates the shape prior by using weighted wavelet frame regularization. The numerical results on real ultrasound data sets demonstrate that the proposed wavelet frame model with distance prior can track the ROI effectively. Furthermore, the proposed model for single image segmentation can be solved by the proximal alternating minimization (PAM) algorithm, and the convergence of the iterative sequence is established. Compared to the level set method, the proposed scheme is significantly faster. Additionally, the use of a Poisson noise fidelity term can slightly improve the performance for ultrasound segmentation compared to the usual Gaussian noise assumption.

REFERENCES

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr., The Clarendon Press, Oxford University Press, New York, 2000.
[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.
[3] E. Bae, J. Yuan, and X.-C. Tai, *Global minimization for continuous multiphase partitioning problems using a dual approach*, Int. J. Comput. Vis., 92 (2011), pp. 112–129.
[4] C. Bao, H. Ji, and Z. Shen, *Convergence analysis for iterative data-driven tight frame construction scheme*, Appl. Comput. Harmon. Anal., 38 (2015), pp. 510–523.

[5] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program. Ser. A, 146 (2014), pp. 459–494.
[6] X. Bresson, S. Esedoḡlu, P. Vandergheynst, J.-P. Thiran, and S. Osher, *Fast global minimization of the active contour/snake model*, J. Math. Imaging Vision, 28 (2007), pp. 151–167.
[7] E. S. Brown, T. F. Chan, and X. Bresson, *Completely convex formulation of the Chan-Vese image segmentation model*, Int. J. Comput. Vis., 98 (2012), pp. 103–121.
[8] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: Total variation, wavelet frames, and beyond*, J. Amer. Math. Soc., 25 (2012), pp. 1033–1089.
[9] J.-F. Cai, B. Dong, and Z. Shen, *Image restoration: A wavelet frame based model for piecewise smooth functions and beyond*, Appl. Comput. Harmon. Anal., to appear. doi:10.1016/j.acha.2015.06.009.
[10] J.-F. Cai, S. Osher, and Z. Shen, *Split Bregman methods and frame based image restoration*, Multiscale Model. Simul., 8 (2009), pp. 337–369.
[11] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic active contours*, Int. J. Comput. Vis., 22 (1997), pp. 61–79.
[12] A. Chambolle, D. Cremers, and T. Pock, *A convex approach to minimal partitions*, SIAM J. Imaging Sci., 5 (2012), pp. 1113–1158.
[13] T. F. Chan, S. Esedoḡlu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.
[14] T. F. Chan and L. Vese, *Active contours without edges*, IEEE Trans. Image Process., 10 (2001), pp. 266–277.
[15] Y. Chen, H. D. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. A. Geiser, *Using prior shapes in geometric active contours in a variational framework*, Int. J. Comput. Vis., 50 (2002), pp. 315–328.
[16] D. Cremers, *Dynamical statistical shape priors for level set-based tracking*, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006), pp. 1262–1273.
[17] I. Daubechies, B. Han, A. Ron, and Z. Shen, *Framelets: MRA-based constructions of wavelet frames*, Appl. Comput. Harmon. Anal., 14 (2003), pp. 1–46.
[18] Q. Ding, Y. Zan, Q. Huang, and X. Zhang, *Dynamic SPECT reconstruction from few projections: A sparsity enforced matrix factorization approach*, Inverse Problems, 31 (2015), 025004 (26 pages).
[19] B. Dong, A. Chien, and Z. Shen, *Frame based segmentation for medical images*, Commun. Math. Sci., 9 (2010), pp. 551–559.
[20] B. Dong, Q. Jiang, and Z. Shen, *Image Restoration: Wavelet Frame Shrinkage, Nonlinear Evolution PDEs, and Beyond*, UCLA CAM Report 13-78, Department of Mathematics, University of California Los Angeles, 2013.
[21] B. Dong and Z. Shen, *MRA based wavelet frames and applications*, in Mathematics in Image Processing, IAS/Park City Math. Ser. 19, Amer. Math. Soc., Providence, RI, 2013, pp. 9–158.
[22] E. Esser, *Applications of Lagrangian-based Alternating Direction Methods and Connections to Split Bregman*, UCLA CAM report 09-31, Department of Mathematics, University of California Los Angeles, 2009.
[23] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.
[24] R. Glowinski and A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires*, Rev. Française Automat. Informat. Recherche Oprationnelle Sér. Rouge Anal. Numér., 9 (1975), pp. 41–76.
[25] T. Goldstein and S. Osher, *The split Bregman method for L1 regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
[26] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier (Grenoble), 48 (1998), 769–783.
[27] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr, *Convex multi-class image labeling by simplex-constrained total variation*, in Scale Space and Variational Methods in Computer Vision, Second International Conference (Voss, Norway), Lecture Notes in Comput. Sci. 5567, Springer, Berlin, 2009, pp. 150–162.
[28] S. Łojasiewicz, *Sur la géométrie semi-et sous-analytique*, Ann. Inst. Fourier (Grenoble), 43 (1993), p. 1575–1595.

[29] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Commun. Pure Appl. Math., 42 (1989), pp. 577–685.
[30] J. A. Noble and D. Boukerroui, *Ultrasound image segmentation: A survey*, in IEEE Trans. Medical Imaging, 25 (2006), pp. 987–1010.
[31] S. Osher and J. A. Sethian, *Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys., 79 (1988), pp. 12–49.
[32] N. Paragios and R. Deriche, *Geodesic active contours and level sets for the detection and tracking of moving objects*, IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), pp. 266–280.
[33] T. Pock, A. Chambolle, D. Cremers, and H. Bischof, *A convex relaxation approach for computing minimal partitions*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, 2009, pp. 810–817.
[34] R. B. Potts, *Some generalized order-disorder transformations*, Proc. Cambridge Philos. Soc., 48 (1952), pp. 106–109.
[35] A. Ron and Z. Shen, *Affine systems in $L_2(\mathbf{R}^d)$: The analysis of the analysis operator*, J. Funct. Anal., 148 (1997), pp. 408–447.
[36] A. Sawatzky, D. Tenbrinck, X. Jiang, and M. Burger, *A variational framework for region-based segmentation incorporating physical noise models*, J. Math. Imaging Vision, 47 (2013), pp. 179–209.
[37] C. Tai, X. Zhang, and Z. Shen, *Wavelet frame based multiphase image segmentation*, SIAM J. Imaging Sci., 6 (2013), pp. 2521–2546.
[38] W. Tao and X.-C. Tai, *Multiple piecewise constant with geodesic active contours (MPC-GAC) framework for interactive image segmentation using graph cut optimization*, Image Vis. Comput., 29 (2011), pp. 499–508.
[39] D. Tenbrinck, A. Sawatzky, X. Jiang, M. Burger, W. Haffner, P. Willems, M. Paul, and J. Stypmann, *Impact of physical noise modeling on image segmentation in echocardiography*, in Proceedings of the Third Eurographics Workshop on Visual Computing for Biology and Medicine, Norrköping, Sweden, T. Ropinski, A. Ynnerman, C. Botha, and J. Roerdink, eds., 2012, pp. 33–40.
[40] L. A. Vese and T. F. Chan, *A multiphase level set framework for image segmentation using the Mumford and Shah model*, Int. J. Comput. Vis., 50 (2002), pp. 271–293.
[41] Y. Xu and W. Yin, *A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion*, SIAM J. Imaging Sci., 6 (2013), pp. 1758–1789.
[42] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer, *Fast global labeling for real-time stereo using multiple plane sweeps*, in Proceedings of the 13th International Fall Workshop on Vision, Modeling, and Visualization (Konstanz, Germany), O. Deussen, D. Keim, and D. Saupe, eds., Akademische Verlagsgesellschaft, Heidelberg, 2008, pp. 243–252.
[43] H. Zhao, *A fast sweeping method for eikonal equations*, Math. Comp., 74 (2005), pp. 603–627.