

Article

A PDE-Free Variational Method for Multi-Phase Image Segmentation Based on Multiscale Sparse Representations [†]

Julia Dobrosotskaya ^{*,†,‡} and Weihong Guo [‡]

Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, Cleveland, OH 44106, USA; wxg49@case.edu

^{*} Correspondence: jxd365@case.edu; Tel.: +1-310-948-8301[†] This paper is an extended version of our paper published in J. Dobrosotskaya and W. Guo. ‘A PDE-free variational model for multiphase image segmentation’. In “Wavelets and Sparsity XVI”, Volume 9597, 2015.[‡] Current address: Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, 2049 Martin Luther King Jr. Drive, Cleveland, OH 44106-7058, USA.

Received: 30 December 2016; Accepted: 19 June 2017; Published: 13 July 2017

Abstract: We introduce a variational model for multi-phase image segmentation that uses a multiscale sparse representation frame (wavelets or other) in a modified diffuse interface context. The segmentation model we present differs from other state-of-the-art models in several ways. The diffusive nature of the method originates from the sparse representations and thus propagates information in a different manner comparing to any existing PDE models, allowing one to combine the advantages of non-local information processing with sharp edges in the output. The regularizing part of the model is based on the wavelet Ginzburg–Landau (WGL) functional, and the fidelity part consists of two terms: one ensures the mean square proximity of the output to the original image; the other takes care of preserving the main edge set. Multiple numerical experiments show that the model is robust to noise yet can preserve the edge information. This method outperforms the algorithms from other classes in cases of images with significant presence of noise or highly uneven illumination

Keywords: multiphase segmentation; variational method; diffuse interface; wavelets

1. Introduction

Image segmentation is a technique of partitioning an image domain into multiple regions such that each region is homogeneous with respect to some characteristic such as intensity, texture and/or color. It is often used to locate objects or to find the respective boundary. It is an active research topic with numerous practical applications. Examples include medical imaging, machine vision and video surveillance. Image segmentation can be very challenging for images with noise, low contrast and multi-scale, multi-directional details. Conducting multiphase segmentation, rather than binary “object vs. background” segmentation in these cases adds complexity to the problem.

Image segmentation has been studied for many years. Methods in the literature include simple intensity thresholding, edge detection-based and clustering-based segmentation, watershed approaches, variational and graph-based techniques. A variety of methods was proven to be efficient for a variety of images (datasets) with different properties. For instance, thresholding [1–3] is the simplest and fastest method that works well for images with clearly-defined boundaries between target regions and without significant noise present. The watershed method [4,5] tends to be more successful when the objects of interests are closely grouped together and, possibly, touch and share their boundary (e.g., microscopy imagery of cells). It however requires correct seeds for each object to

segment [6–8]. Image segmentation can be very difficult for images with noise, low contrast and/or multi-scale, multi-directional details. Variational models usually involve solving an appropriate minimization problem, and thus are, on the one hand, more time consuming, yet, on the other hand, are more robust and yield more accurate results due to the regularity and fidelity terms providing clear and numerically-implementable requirements on the output. Many variational segmentation models originate from the Potts model [9] that minimizes the sum of the boundary length and data fidelity. The Potts model is NP hard in multiphase cases. Graph cut-based models were proposed to approximate global minimizers in computationally-efficient ways [10–15]. The Mumford–Shah model [16] can be viewed as a discrete case of the Potts model that assumes piecewise smoothness of the image intensity.

The geodesic active contour model [17], Chan–Vese active contour model [18,19] and other level set methods represent another important class of techniques that were proposed to improve segmentation efficiency. In particular, the multiphase Vese–Chan model [19] assumes piecewise constant image intensity and is easier to implement than Mumford–Shah. A method involving Γ approximations of the total variation by functionals with equispaced multiple well potentials in the context of multiphase segmentation is presented in [20]. However, due to the nonconvexity of the model, the segmentation results highly rely on the selection of the initial approximation, thus requiring careful practical implementation. Recently, convex relaxation-based approaches have been studied to improve the robustness of segmentation, as well as the computational efficiency [21–27]. There is also work focused on multiphase fuzzy segmentation and applied on medical images [28]. Diffuse-interface models adapted to the graph context were shown to be efficient in binary segmentation for a variety of datasets [29,30]. Other not directly related, but effective segmentation methods listed on the Berkeley segmentation dataset and benchmark website include [31–35]. These techniques are contour detection oriented and are directly associated with the object vs. background classification.

About our method. As was remarked in [19], the multi-phase image segmentation problem is different from any binary segmentation task. Our method, just as the above mentioned level set segmentation, by construction avoids the issues of ‘vacuum’ and ‘overlap’; moreover, m binary functions are enough to encode 2^m -valued piecewise constant approximation. The case when $m = 2$ is considered especially interesting, since, formally, the partition of any piecewise smooth image near any boundary junction can be encoded by only two binary functions, per the four color theorem. Therefore, in our problem setup, we focus on the four-phase image segmentation as a typical case of a problem of approximating a given image by a piecewise constant one, attaining at most four intensity values, with connected components of the output satisfying certain regularity conditions. Those conditions are imposed by the properties of the energy functional we introduce in Section .

Our model uses such features of diffuse interface behavior as coarsening and phase separation to merge relevant image elements (coarsening) and separate others into distinct classes (phase separation). To balance these two tendencies, one can adjust the diffuse interface parameter ϵ , just as in the classical diffuse interface models that arise in material science. However, in the new spatial derivative-free setup, the interface width is no longer proportional to ϵ (due to the well-localized elements in the chosen sparse representation systems and, thus, a completely different diffusive nature of the model), allowing combining the advantages of non-local information processing with sharp edges in the output. Numerical experiments confirm the effectiveness of the proposed method.

This technique is especially suitable for images where the segmented output is expected to have piecewise smooth edges and relatively large connected components (that is determined by the diffuse interface value ϵ and the weights of the fidelity terms; see Section 1) within each segmented class (and nearly no isolated pixels).

Novelty and contributions. Preliminary findings about this approach appeared in the SPIE Conference Proceedings [36]. In this paper, we do an in-depth investigation of this segmentation model. We provide a detailed description of the numerical scheme and conditions that make it gradient stable; we include the formal proof of the minimizer existence, a detailed description of the fidelity

term choice and parameter choice; we also describe generalizations and adaptations of this method for particular applications: using it with post-processing for specific needs of medical imaging (blood vessel detection) and generalization of the method to process color (vector-valued) images, both illustrated with examples of numerical simulations. Moreover, following up on the requests of this journal's reviewers, we include the quantitative area-segmentation comparisons to other methods (using benchmarks described in [31] and other works associated with the Berkeley segmentation database(s); please, see Section 4.1 and Appendix B.1.).

In Sections 1.1 and 1.2, we provide a more detailed explanation of our model design, the reasons for using wavelets or similar multiscale systems in the variational context and briefly describe the properties of the wavelet Ginzburg–Landau energy. Sections 2 and 3 introduce the variational model and its discrete implementation, respectively. Section 4 contains the results of numerical simulations and a discussion of the effectiveness of our method.

1.1. PDE-Free Variational Models Based on Sparse Representations: Motivation

Similarities between image topology features and phase transitions behavior in material science and fluid mechanics motivated the integration of diffuse interface models into image processing [37,38]. These techniques are sometimes based on the phase-field models, such as the one of Modica and Mortola [39], and allow for topology transition between two steady states of a certain physical or artificial system. Such a construction seems naturally consonant with image processing applications that consider the image intensity bitwise and treat (recover) respective binary values as two equilibria of a model system [40,41].

In general, non-linear diffusion models allow the inclusion of a priori knowledge to ensure regularity while preserving important features including edges [42]. Anisotropic diffusion provides the advantage of combining image regularization with the possibility of edge enhancement [43,44].

However, in signal and especially image processing, we deal with discrete signals and spatially localized features, which do not directly fit into the context of differentiation, even in the weak sense, and the Fourier basis, containing the eigen-functions of certain differential operators, does not provide spatial localization. Moreover, Fourier coefficients of a typical image are not sparse.

A logical and elegant resolution of these difficulties lies in constructing a class of operators based on multiscale sparse representation systems, such as multiscale frames (wavelet or other $\ast$ -lets). These operators can be designed to act on the chosen frame in the same scale proportional manner as the differential operators act on the Fourier basis. It turns out that with a suitable choice of the multiscale sparse representation system, the new operators retain the desired properties of their differential prototypes, while bringing in new advantages, such as being inherently multiscale, well-localized in space and frequency and, for some multiscale sparse representation systems, allowing some control of the directional features. However, even using the most basic systems like 2D separable wavelets yields non-trivial improvement in the diffuse interface imaging techniques.

1.2. Wavelets in Variational Models for Image Analysis and Recovery, Wavelet Ginzburg–Landau Energy

Wavelets are a well-known powerful tool of image processing, primarily for their ability to capture important signal features within a few terms of the wavelet decomposition; significant attention has been devoted to investigating the properties of the sparse representation systems as effective tools of harmonic analysis within new models that better reflect the modern signal processing needs. In this paper, we exploit the intrinsic connections between the classical differential operators and pseudo-differential ones based on sparse representations (already studied in [45,46]) in the new context of the segmentation problem where two functions and possibly some additional parameters need to be recovered (via solving coupled systems of evolutionary differential equations).

Wavelets were used for image enhancement and recovery in a multitude of variational settings; see [47,48] and more. In particular, [47] discusses the total variation (TV) minimization in the wavelet

domain, that successfully reproduces lost coefficients. The relationship between wavelet-based image processing algorithms and variational problems in general is analyzed in [49,50].

The wavelet Ginzburg–Landau energy was constructed from the idea of designing new types of pseudo-differential energy functionals that inherit important properties of the ones involving derivatives, introduce the advantages of non-local analysis of the image features and leave out the computational drawbacks associated with the discrete differentiation.

The key idea in the design of wavelet Ginzburg–Landau (WGL) [45,46] combines the basic geometric framework of diffuse interface methods with advantages of the well-localized and inherently multiscale wavelet operators. It originally appeared in a variational model that was adapted to a wavelet setting with the goal of using it for image inpainting. However, it is an efficient regularizing term, and it was shown to be successfully used in multiple image analysis and recover settings with properly chosen forcing terms.

The analytical properties of the wavelet Ginzburg–Landau (GL) energy and its imaging applications are described in detail in [45,46]. We present a very brief summary below.

Here and further in the text, whenever the domain of integration is omitted, we assume integrating over the entire domain of the respective functions.

The total variation (TV) seminorm was proven to be a natural and efficient measure of image regularity [51,52]. However, in many imaging applications, one would like to retain its advantages while making the processing non-local and reduce the computational load related to the curvature calculations. The problem can be re-formulated using the phase-field method providing approximation to the TV functional (in the Γ sense). The Ginzburg–Landau (GL) functional,

$$GL(u) = \frac{\epsilon}{2} \int |\nabla u(x)|^2 dx + \frac{1}{4\epsilon} \int W(u) dx, \quad W(u) = (u^2 - 1)^2, \quad (1)$$

is a diffuse interface approximation to the total variational functional $\int |\nabla u(x)| dx$ in the case of binary images [39,53].

GL energy is used in modeling of a vast variety of phenomena including the second-order phase transitions. However, if used in signal processing applications, diffuse interface models tend to produce results that are oversmoothed comparing to the optimal output. In the new model, the H^1 seminorm $\int |\nabla u(x)|^2 dx$ is replaced with a Besov (1-2-2) seminorm. This allows one to construct a method with properties similar to those of the PDE-based methods, but without as much diffuse interface scale blur.

The “wavelet Laplace operator” was defined by having the wavelet basis functions as eigenfunctions and acting on those in the same “scale - proportional” manner as the Laplace operator acts on the Fourier basis. Given an orthonormal wavelet ψ , the “wavelet Laplacian” of any $u \in L^2(R)$ is formally defined as:

$$\Delta_w u = - \sum_{j=0}^{+\infty} 2^{2j} \int \langle f, \psi_{j,\kappa} \rangle \psi_{j,\kappa} d\kappa, \quad \psi_{j,\kappa} = 2^{j/2} \psi(2^j x - \kappa). \quad (2)$$

Then, the “wavelet Allen–Cahn” equation $u_t = \epsilon \Delta_w u - \frac{1}{\epsilon} W(u)$ describes the gradient descent in the problem of minimizing the wavelet Ginzburg–Landau (WGL) energy:

$$WGL(u) := \frac{\epsilon}{2} |u|_B^2 + \frac{1}{4\epsilon} \int W(u) dx, \quad \text{where } |u|_B^2 = \sum_{j=0}^{+\infty} 2^{2j} \int |\langle f, \psi_{j,\kappa} \rangle|^2 d\kappa \quad (3)$$

is the square of the Besov 1-2-2 (translation-invariant) semi-norm if the wavelet ψ is r -regular, $r \geq 2$. The same exact energy construction (3) works just as well in the 2D case (in the case of the separable 2D wavelet basis, one needs to include summation over all types of coefficients: ‘horizontal’, ‘vertical’, ‘diagonal’, in the above formulas).

WGL functionals are inherently multiscale and take advantage of simultaneous space and frequency localization (due to the wavelet properties), thus allowing much sharper minimizer transitions for the same values of the interface parameters compared to the classical GL energy.

WGL energy does not have any local maxima, but a local minimum can be found via the gradient descent method. The latter problem is equivalent to solving:

$$\begin{cases} u_t = \epsilon \Delta_w u - \frac{1}{\epsilon} u^3 + \frac{1}{\epsilon} u \\ u(x, 0) = u_0(x) \in H^1([0, 1]^2). \end{cases} \quad (4)$$

The above problem is well-posed: it has a unique solution that exists globally in time and converges to a steady state as $t \rightarrow \infty$. The steady state solution is infinitely smooth provided that wavelet ψ used in the construction of the energy has sufficient regularity.

The WGL energy approximates a weighted TV functional in the variational (Γ -convergence) sense.

The WGL was initially designed for the recovery of binary signals; however, it is easily adaptable to a wide class of imaging applications via bit-wise processing ([40] and more).

Variational techniques with WGL as the regularizer were developed for such problems as inpainting, denoising, superresolution, segmentation and more [45]. In all of those cases, the minimized energy was respectively modified to include one or more fidelity terms suitable for each type of problem. When used for the purposes of (binary) segmentation, the input of the method was assumed to include “partial classification” of the image into two classes, and the variational model “filled in the gaps”: the diffusive nature of the model allowed propagating the information, and the edge-preserving term enforced the boundaries of the objects.

The model we introduce in this paper is different from other WGL-based ones in several ways. In the new model, we recover a more complicated output: the binary functions that define the separation of image pixels into classes and the values of constants of the piecewise constant approximation within one unified procedure. It is achieved by solving the system of ordinary differential equations associated with the gradient descent minimization of the proposed energy functional.

2. The Proposed Model for the Four-Phase Image Segmentation

2.1. The Problem Setup

Let us denote the original image f and assume it is a function from $H^1([0, 1]^2)$. We are looking for the four-phase segmented version of the given image, i.e., its nearly piecewise constant approximation in the following form:

$$u(x) = c_1 \phi_1(x) \phi_2(x) + c_2 \phi_1(1 - \phi_2(x)) + c_3(1 - \phi_1(x)) \phi_2(x) + c_4(1 - \phi_1(x))(1 - \phi_2(x)) \quad (5)$$

Function u is, by design, a piecewise constant function attaining four possible values c_1, c_2, c_3, c_4 , and functions ϕ_1 and ϕ_2 are binary (take on values of zero and one). For simplicity, let us assume that $c_1 \leq c_2 \leq c_3 \leq c_4$. Since our method is a diffuse interface, the recovered functions ϕ_i will be nearly binary, such that the width of the transition between black and white is comparable to the size of one pixel.

We assume that the desired function u can be obtained by minimizing the following energy:

$$E(\phi_1, \phi_2, \vec{c}) = WGL(\phi_1) + WGL(\phi_2) + \mu \|u - f\|_{L^2}^2 + \lambda |Pr_\Lambda(u - f)|_B^2 \quad (6)$$

with respect to functions ϕ_1 and ϕ_2 , as well as $\vec{c} = (c_1, c_2, c_3, c_4)$. The minimizing functions ϕ_1, ϕ_2 are expected to be nearly-binary functions from $H^1([0, 1]^2)$. The fidelity part of the energy consists of two terms. The term $\mu \|u - f\|_{L^2}^2$ ensures the mean-square proximity of the resulting piecewise constant output u to the original function f . The edge-preserving term $\lambda |Pr_\Lambda(u - f)|_B^2$ takes care of retaining the most important edge information: here Λ is the subset of the wavelet modes in the wavelet decomposition that are significant enough to be preserved in the segmented version of the

image; one of the possible ways of choosing Λ is described in Section A4. The values μ and λ are weighting parameters that are chosen depending on the image quality, detail resolution and amount of reliable edge information that needs to be preserved. The seminorm $|\cdot|_B$ denotes the Besov 1-2-2 seminorm (just as in (3)).

The existence of minimizers for E can be proven using a standard compactness argument (see Appendix A.1). The minimizers are not unique (the easiest counterexample is a constant input function). The output of our numerical simulations depends on the initial guess, which allows us to steer the solution in the desired direction. We provide details about the initialization procedure in Section 3.2.

2.2. Motivation for the Energy Design

The main objective of any image processing method is to successfully incorporate all of the known data properties and requirements on the output into a tunable mathematical model. In variational models, the terms of the minimized energy often are categorized as either regularizing or fidelity (forcing) terms. Let us explain the features of the above energy in relation to our segmentation goal.

In our case, the pixels in the output image will be inadvertently grouped into colored connected components. Each of such components should have a reasonably regular geometric structure (due to the influence of the regularizing WGL terms) unless otherwise imposed by the edges of the original image (due to the influence of the edge-preserving and/or spatial fidelity terms). The values of the functions ϕ_1 and ϕ_2 are also expected to be nearly binary.

The terms $WGL(\phi_1)$ and $WGL(\phi_2)$ represent the regularizing part of the energy. They bring in the properties of the diffusive model inherited from the classical second order phase transition ‘coarsening’ and ‘phase separation’, which facilitate the needed evolution within the image: merging relevant image elements (coarsening) and separating others into distinct classes (phase separation). To balance these two tendencies, one can adjust the diffuse interface parameter ϵ , just as in the classical diffuse interface models that arise in material science. However, in the new spatial derivative-free setup, the interface width is no longer proportional to ϵ (due to the well-localized elements in the chosen sparse representation systems and, thus, a completely different diffusive nature of the model), allowing combining the advantages of non-local information processing with sharp edges in the output.

For each of the functions ϕ_i , we aim to minimize WGL, thus making their supports somewhat regularly-shaped (due to the coarsening tendency) and their values nearly binary (due to phase-separation).

The fidelity components of the energy are $\mu\|u - f\|_{L^2}^2 + \lambda|Pr_\Lambda(u - f)|_B^2$. These terms impose consistency with the original image f in the L^2 sense and preserve significant edges within the original image.

If the piecewise constant approximation values are known in advance, the simplified version of this algorithm can recover the corresponding regions by finding ϕ_1 and ϕ_2 . However, in general, the coefficients c_i are not defined in advance and cannot be easily recovered from the histogram analysis of the image values, as shown in Appendix A.2. In the rest of the paper, we assume that $\vec{c}$ is unknown, and in all of our numerical experiments included in this paper, we recovered ϕ_1 , ϕ_2 and $\vec{c}$ simultaneously.

In the rest of the paper, we will show that our method provides a powerful tool of using diffusive (in the alternative sense of gradually suppressing the high scale detail in the context of a multiscale multidirectional representation system such as wavelets) propagation of information without the associated blurring. The trade-off between preserving the fine-scale features and connecting all relevant components is resolved by using the forcing terms, including the edge-preserving term, obtained by the specialized scale-dependent thresholding.

2.3. Solving the Minimization Problem

We look for ϕ_1, ϕ_2 , the minimizers (local) of the energy (4), as well as for the coefficients c_i , if those are not fixed in advance.

Let $c_i = p_i^2$ for convenience in writing the gradient descent equations. We can do so because we are assuming that the range of values of the image is non-negative; thus, the optimal approximation values (even in the case of noisy images) are non-negative.

Consider a variation of this energy at a point $(\phi_1, \phi_2, \vec{p})$ in the direction $(\psi_1, \psi_2, \vec{s})$:

$$\begin{aligned} \left. \frac{d}{dt} E(\phi_1 + t\psi_1, \phi_2 + t\psi_2, \vec{p} + t\vec{s}) \right|_{t=0} = \\ \int_{[0,1]^2} [-\Delta_w \phi_1 + W'(\phi_1) + 2 \frac{\partial u}{\partial \phi_1} (\mu(u-f) - \lambda \Delta_w Pr_\Lambda(u-f))] \psi_1 d\vec{x} + \\ \int_{[0,1]^2} [-\Delta_w \phi_2 + W'(\phi_2) + 2 \frac{\partial u}{\partial \phi_2} (\mu(u-f) - \lambda \Delta_w Pr_\Lambda(u-f))] \psi_2 d\vec{x} + \\ \int_{[0,1]^2} 4(\mu(u-f) + \lambda \Delta_w Pr_\Lambda(u-f)) \Phi(x) \vec{p} d\vec{x} \cdot \vec{s} \end{aligned} \quad (7)$$

Here, $\frac{\partial u}{\partial \phi_1} = (c_1 - c_2 - c_3 + c_4)\phi_2 + (c_2 - c_4)$, $\frac{\partial u}{\partial \phi_2} = (c_1 - c_2 - c_3 + c_4)\phi_1 + (c_3 - c_4)$, and:

$$\Phi = \begin{bmatrix} \phi_1 \phi_2 & 0 & 0 & 0 \\ 0 & \phi_1(1 - \phi_2) & 0 & 0 \\ 0 & 0 & (1 - \phi_1)\phi_2 & 0 \\ 0 & 0 & 0 & (1 - \phi_1)(\phi_2) \end{bmatrix} \quad (8)$$

We look for the minimizers of the chosen energy functional as the steady state solutions (in the weak sense) to the gradient descent system:

$$\begin{cases} \partial_t \phi_1 = \Delta_w \phi_1 - W'(\phi_1) - 2 \frac{\partial u}{\partial \phi_1} (\mu(u-f) + \lambda \Delta_w Pr_\Lambda(u-f)) \\ \partial_t \phi_2 = \Delta_w \phi_2 - W'(\phi_2) - 2 \frac{\partial u}{\partial \phi_2} (\mu(u-f) + \lambda \Delta_w Pr_\Lambda(u-f)) \\ \partial_t \vec{p} = - \int_{[0,1]^2} 4(\mu(u-f) + \lambda \Delta_w Pr_\Lambda(u-f)) \Phi(x) \vec{p} d\vec{x}. \end{cases} \quad (9)$$

Here, $\phi_1(x, t), \phi_2(x, t) \in H^1([0, 1]^2 \times [0, T])$ and $\vec{p}(t) \in H^1([0, T])$ for any $T > 0$.

Our method has a non-local nature as the equations defining the minimizers (ϕ_i) contain a non-local operator (Δ_w) , thus allowing creating better connections between components of the segmented image. Interaction of the diffusive term that facilitates merging of similar-colored regions and the term minimizing the double-well potential, i.e., responsible for keeping functions ϕ_i nearly binary, closely resembles the “coarsening” and “phase separation” attributes of the classical Allen–Cahn equation. However, due to the wavelet-based operator we use instead of the Laplacian, we can achieve better phase separation and, thus, narrower black to white transitions even while requiring larger connected components in the output. In other words, larger values of parameter ϵ indeed facilitate the diffusion process that merges more pixels into one connected component, yet the loss of contrast (i.e., increasing the width of the black to white transition) is most certainly less noticeable than in diffuse-interface methods based on the classical models involving diffusive differential operators. These features were investigated in detail in [45].

The edge-enforcing component enhances the recovery of more precise boundaries of the segmented regions. The spatial consistency term naturally enforces the L^2 proximity between the input image and its segmented version.

Since our model is variational and any minimization process involves an initial guess, we describe one of the possible ways to choose the initial values in Section 3.2.

3. Numerical Implementation

3.1. Numerical Scheme

To compute the minimizers of the proposed variational problem, we implement the gradient descent method described in Section 2.3 via a semi-implicit numerical scheme. To guarantee gradient-stable convergence to a steady state (for functions ϕ_1, ϕ_2), we use the convexity splitting scheme from [54]: $E(\phi_1, \phi_2, \vec{c}) = (E_{11} - E_{12}) + (E_{21} - E_{22})$, where:

$$E_{11} = \frac{\epsilon}{2} |\phi_1|_B^2 + \frac{\epsilon}{2} |\phi_2|_B^2 + A_1 (\|\phi_1\|_{L^2}^2 + \|\phi_2\|_{L^2}^2) \quad (10)$$

$$E_{12} = -\frac{1}{4\epsilon} \int W(\phi_1) - \frac{1}{4\epsilon} \int W(\phi_2) + A_1 (\|\phi_1\|_{L^2}^2 + \|\phi_2\|_{L^2}^2) \quad (11)$$

$$E_{21} = A_2 (\|\phi_1\|_{L^2}^2 + \|\phi_2\|_{L^2}^2), \quad E_{22} = A_2 (\|\phi_1\|_{L^2}^2 + \|\phi_2\|_{L^2}^2) - \mu \|u - f\|_{L^2}^2 - \lambda |Pr_\Lambda(u - f)|_B^2 \quad (12)$$

Here, the constants $A_1, A_2 > 0$ need to be chosen in such a way that all four energies are convex, making E a difference of strictly convex energy functionals. If $\lambda = 0$, it can be achieved by choosing $A_1 \gtrsim \frac{1}{\epsilon}$ and $A_2 \gtrsim \mu$. If we consider only the “continuous, non-discretized” form of the energy, the needed A_2 , generally speaking, does not exist. However, in the discretized form, all functions involved in the gradient descent are represented by matrices; therefore, only finitely many levels of wavelet decomposition are present in the Besov seminorm in the fidelity term; thus, we may choose $A_2 \gtrsim \mu + \lambda 2^{2J_{max}}$, where J_{max} is the maximum level of wavelet decomposition possible for matrices of the chosen size.

Let us describe the discretized version of the gradient descent system. The gradient descent is performed with respect to an artificial time parameter t . Let δt denote the time step size, then $t = n\delta t$ after n steps of the gradient descent algorithm. Let the spatial domain $[0, 1] \times [0, 1]$ be represented by the points on the uniform grid of size $N \times N$ with step h : (x_j, y_k) , $x_j = jh$, $y_k = kh$, $0 \leq j, k \leq N$, $h = \frac{1}{N-1}$. Due to the use of the discrete wavelet transform, it is convenient to pick $N = 2^l$, $l \in \mathbb{N}$. Let ϕ_i^n denote a matrix $N \times N$ that approximates the values of the ϕ_i at time $n\delta t$, and u^n is defined accordingly, so $u_{j,k}^n \approx u(jh, kh, \delta t)$. Naturally, we assume that the given image f is a matrix $N \times N$ with entries $f_{j,k} = f(hj, hk)$.

If the intensity values of the output are not fixed and, thus, we update $\vec{c}$ at each step as well, we let $\vec{c}^n$ denote the value of the intensity vector after n steps of the gradient descent algorithm.

For the semi-implicit numerical scheme implementing the equations of the gradient descent system, we formulated the above results in the following equations:

$$\begin{aligned} u^n &= c_1^n \phi_1^n \phi_2^n + c_2^n \phi_1^n (1 - \phi_2^n) + c_3^n (1 - \phi_1^n) \phi_2^n + c_4^n (1 - \phi_1^n) (1 - \phi_2^n), \\ \phi_1^{n+1} - \delta t \Delta_w \phi_1^{n+1} + \delta t (A_1 + A_2) \phi_1^{n+1} &= \phi_1^n + \delta t (-W'(\phi_1^n) - 2((c_1^n - c_2^n - c_3^n + c_4^n) \phi_2^n + \\ &+ (c_2^n - c_4^n))(\mu(u^n - f) + \lambda \Delta_w Pr_\Lambda(u^n - f)(A_1 + A_2) \phi_1^n), \\ \phi_2^{n+1} - \delta t \Delta_w \phi_2^{n+1} + \delta t (A_1 + A_2) \phi_2^{n+1} &= \phi_2^n + \delta t (-W'(\phi_2^n) - 2((c_1^n - c_2^n - c_3^n + c_4^n) \phi_1^n + \\ &+ (c_3^n - c_4^n))(\mu(u^n - f) + \lambda \Delta_w Pr_\Lambda(u^n - f) + (A_1 + A_2) \phi_2^n). \end{aligned} \quad (13)$$

If the expected constant intensity values of the output image are not known in advance, they can be computed simultaneously with the segmented region defining functions ϕ_1, ϕ_2 :

$$\begin{aligned} c_i^n &= (p_i^n)^2, \\ p_1^{n+1} &= p_1^n (1 - \delta t) h^2 \sum_{j,k} 4(\mu(u_{j,k}^n - f_{j,k}) + \lambda \Delta_w Pr_\Lambda(u_{j,k}^n - f_{j,k})) (\phi_1^n)_{j,k} (\phi_2^n)_{j,k} \\ p_2^{n+1} &= p_2^n (1 - \delta t) h^2 \sum_{j,k} 4(\mu(u_{j,k}^n - f_{j,k}) + \lambda \Delta_w Pr_\Lambda(u_{j,k}^n - f_{j,k})) (\phi_1^n)_{j,k} (1 - (\phi_2^n)_{j,k}) \\ p_3^{n+1} &= p_3^n (1 - \delta t) h^2 \sum_{j,k} 4(\mu(u_{j,k}^n - f_{j,k}) + \lambda \Delta_w Pr_\Lambda(u_{j,k}^n - f_{j,k})) (1 - (\phi_1^n)_{j,k}) (\phi_2^n)_{j,k} \\ p_4^{n+1} &= p_4^n (1 - \delta t) h^2 \sum_{j,k} 4(\mu(u_{j,k}^n - f_{j,k}) + \lambda \Delta_w Pr_\Lambda(u_{j,k}^n - f_{j,k})) (1 - (\phi_1^n)_{j,k}) (1 - (\phi_2^n)_{j,k}). \end{aligned} \quad (14)$$

To update functions ϕ_1 and ϕ_2 , the operators on the left-hand side of the respective equations are inverted in the wavelet domain. In the actual numerical simulations, we use the stationary wavelet transform (redundant, undecimated wavelet transform) due to its translation invariant properties that are pivotal in image processing (especially when dealing with natural images having no intrinsic dyadic structure).

3.2. Initialization

We rescale the range of the intensity values for each processed image to the interval $[0, 1]$ for convenience. In particular, it simplifies the description of feasible ranges for the fidelity weights: λ should be of the same order as ϵ^{-1} , and μ should be comparable to $\epsilon^{-1}2^{-2j_{max}}$, where j_{max} is the maximum level of wavelet decomposition used in the edge-preserving fidelity term (see more details in the next subsection). In cases when the input image is known to have a significant level of noise present, we rescaled the values and discard the furthest outliers at the same time, namely cut off the lowest and highest 0.05% of the value range.

The initial guess for the gradient descent system is provided by a simple procedure involving preliminary intensity value distribution analysis of the processed image. A faster convergence can be obtained if the initial guess is provided by the k-means algorithm, which, however, is more computationally expensive, so it is not the best choice of the initial guess. Nevertheless, it is worth mentioning that our algorithm, among other things, can be efficiently used as a post-processing routine following k-means segmentation with the purpose of reducing the number of connected components in the output and regularizing their boundary. We discuss this in the context of color image classification in Section 4.3. Appendix A.2 provides more comments on the choice of the initial guess for $\tilde{c}$.

Here and further in the text, we will use the percentile function:

$$pctl(k) = \text{such value } v \text{ that } k\% \text{ of the image pixels have intensity values less than } v \quad (15)$$

The initial values of the coefficients c_i can be assigned somewhat uniformly from the range of the given image intensity values; we typically took: $c_1 = pctl(5)$, $c_2 = pctl(35)$, $c_3 = pctl(65)$, $c_4 = pctl(95)$ (just as earlier, we assume that f is the given image represented by a matrix $N \times N$ with entries $f_{i,j}$). Then, the preliminary segmented domains can be determined by their indicator functions (matrices) as follows: for $m \in \{1, 2, 3\}$ and any pixel position (i, j) :

$$I_m(i, j) = \begin{cases} 1, & \text{whenever } \operatorname{argmin}_{k \in \{1, 2, 3, 4\}} |f_{i,j} - c_k| = m \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

then:

$$\phi_1 = \max\{I_1, I_2\}, \quad \phi_2 = \max\{I_1, I_3\} \quad (17)$$

The initial value of the approximating piece-wise constant function u then is $u = c_1\phi_1\phi_2 + c_2\phi_1(1 - \phi_2) + c_3(1 - \phi_1)\phi_2 + c_4(1 - \phi_1)(1 - \phi_2)$. Figure 2b,c shows an example of such initialization.

3.3. Choosing the Edge-Forcing Term

The fidelity term $\lambda |Pr_\Lambda(u - f)|_B^2$ is responsible for preserving the main set of edges associated with a subset of wavelet modes that we denoted Λ . The wavelet modes that are “significant” in the segmentation of a given image are found in the same adaptive thresholding algorithm that was previously described in [45]. Even though the thresholding approach was originally designed for binary images, it works well in detecting the edges that are ‘locally binary’, i.e., rapid jumps from one nearly constant intensity region to another. Here is a summary of the thresholding procedure we use to determine the edges that correspond to significant intensity jumps and, therefore, to define Λ .

The redundant (i.e., stationary or undecimated [55]) wavelet transform of an image f ($N \times N$) produces four matrices of coefficients: approximation, horizontal, vertical and diagonal at each level of

decomposition, each matrix of the same size as the original image. Let $C_{j,\vec{k}}^\tau$ formally denote coefficients of the wavelet decomposition of f , indexed by the scale (i.e., dilation parameter) j , translation $\vec{k}$ and the ‘type’ of the coefficient τ . For any index $(j, \vec{k}, \tau)$, $j = 0, \dots, J_{\max} - 1$, $\vec{k} = (k_1, k_2)$ with $k_i = 0, \dots, N - 1$ and τ is ‘a’, ‘h’, ‘v’ or ‘d’, and $J_{\max}$ is the maximum depth of decomposition (number of scales in the decomposition).

We define the subset Λ of significant modes as follows. For any image, depending on its size and quality, we may decide to preserve only significant coefficients of scales within a certain range: $j_{\min} \leq j \leq j_{\max}$ (notice that in our notation, the scales increase from coarsest to finest). If the image contains significant amount of noise, it makes sense to choose $j_{\max} \leq J_{\max} - 2$. It does not make sense to preserve the approximation coefficients of the coarsest level; sometimes, it is convenient to disregard one or two coarsest scales of decomposition completely. Within the chosen range of decomposition scales, we pick the significant modes as follows:

$$(j, \vec{k}, \tau) \in \Lambda \text{ if and only if } |C_{j,\vec{k}}^\tau| \geq \gamma 2^j SD(C_{j,\cdot}^\tau),$$

where SD denotes the standard deviation computed over the set of all coefficients of scale j , within the ‘type’ τ subset, if $\tau = \text{‘h’}, \text{‘v’}, \text{‘d’}$.

If $\tau = \text{‘a’}$,

$$(j, \vec{k}, a) \in \Lambda \text{ if and only if } |C_{j,\vec{k}}^a| \geq \gamma 2^{j-1} SD(C_{j,\cdot}^a).$$

The constant coefficient γ used in this scale- and data-dependent thresholding technique helps produce the best results when chosen individually for each image, but most typically has order $O(1)$.

We illustrate the performance of the method as well as technical nuances leading to the best results in the next section. Figures 1a and more contain examples of the images processed using our method. Figure 2a shows an example of the edges chosen to define the edge-forcing term in one of our segmentation experiments; the set of respective coefficients was obtained by the procedure described above. The images representing the edge information used in our model in the processing of other images can be found in Appendix B.2.

The design of the model and numerical experiments indicate that one should not use the edge fidelity term at all or at least not as a dominating fidelity term, in the cases of smooth images with no visible edges or very noisy images where noise can interfere with the edge detection. Whenever the edges are present, while the noise is relatively mild, it makes sense to use the edge fidelity term, as well as the spatial term, with equal or comparable weighting coefficients.

4. Results

4.1. Four-Phase Segmentation of Grayscale Images

Our numerical experiments involve images from several commonly-considered classes: natural, artificial (cartoon-like), medical (MRI) and images from those categories corrupted by noise and blur. We show that our variational multi-phase segmentation approach can efficiently segment various images, including the tough noisy and blurry images with edges of various scales and directions and even harder ones with low contrast.

Highlighting the previously-mentioned feature of the model, a more relaxed dependence on ϵ as the diffuse interface parameter comparing to the classical PDE-based models, let us remark that in our numerical examples, a typical value of ϵ is of order $N^{-1/2}$ and is always greater than $\frac{4}{N}$ (assuming $N \times N$ is the size of the segmented image). It facilitates the wavelet-based diffusion that forms components of reasonable scale with a piecewise smooth boundary, yet the actual black to white transitions are far from blurry and have a width that is small enough to leave no artifacts after thresholding (if necessary).

To further illustrate the method's effectiveness, we include the outputs of three other related segmentation techniques: the Vese–Chan multi-phase method [19], a graph cut method [14] and a fuzzy segmentation method [28] for comparison.

Figures 1 to 3 illustrate segmentation of a noisy MRI image (Figure 1b) that contains many fine details. Specifically, Figure 2 shows additional details concerning the initialization of the gradient descent minimization and the choice of the edge fidelity term. The MRI image has a size of 256 by 256; the parameters used in simulations are $\epsilon = 1/32$, $\mu = 2.7$, $\lambda = 0.75\epsilon$.

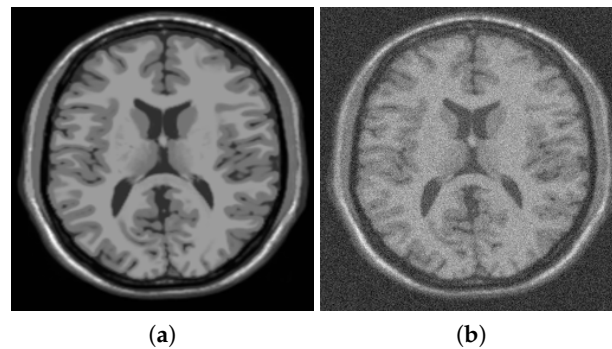


Figure 1. (a) Original ('clean') MRI image, (b) noisy MRI image (noise: 15 dB, Gaussian).

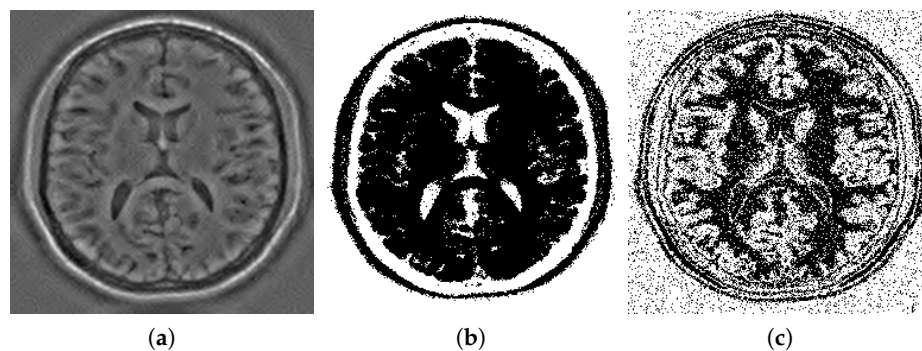


Figure 2. More details about the above MRI image segmentation and the gradient descent setup: (a) the edges to be preserved, (b) initial guess for ϕ_1 , (c) initial guess for ϕ_2 (percentiles used for computing the initial guess: 15%, 38%, 61%, 85%).

Figure 4 shows the output of some well-established segmentation methods applied to the same noisy MRI image. Visual examination of the segmented images demonstrates that the fuzzy segmentation result still has noise in its output, and the Vese–Chan method mostly captures the big scale regions. Only the proposed method and the graph cut methods can provide accurate separation of white matter (shown in white), gray matter (shown in gray) and the rest, which contains either cerebrospinal fluid or the background (shown in darker shades). The results of the graph cut and the proposed method are very similar. However, a zoom-in shows that the proposed method recovers more details in the white matter region. It is especially noticeable in the lower parts of the image.

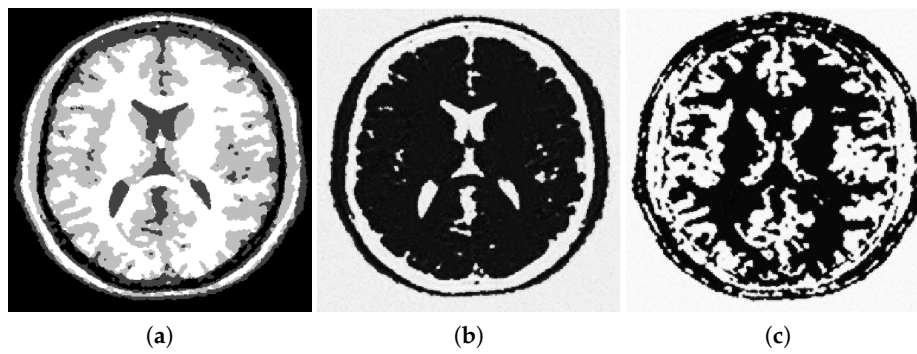


Figure 3. (a) the overall segmentation output of the proposed method, (b) output ϕ_1 , (c) output ϕ_2 .

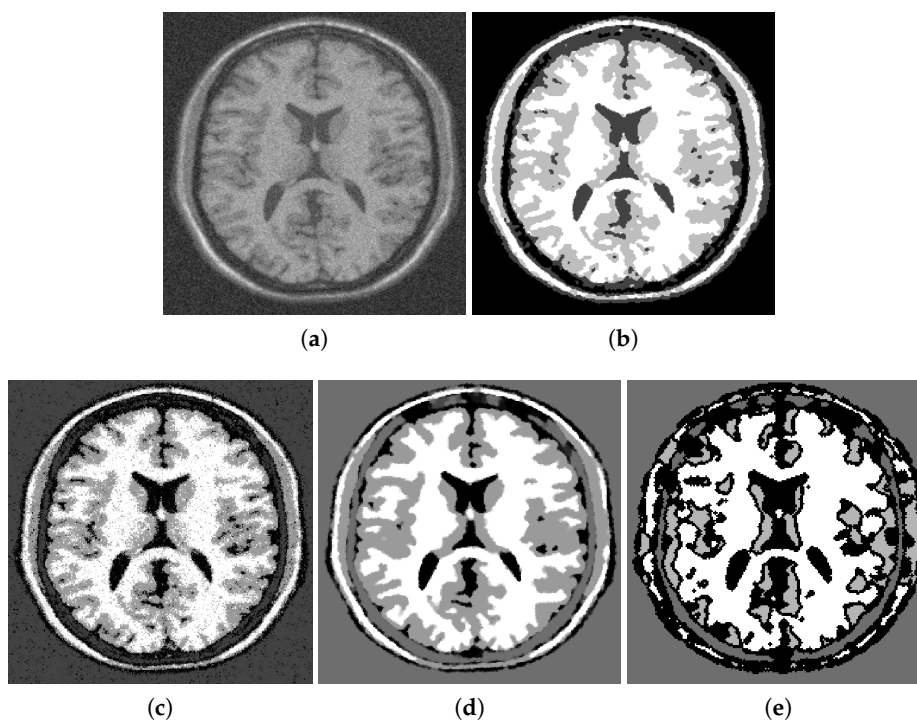


Figure 4. (a) Input image, (b) output of our segmentation method, (c) fuzzy segmentation [28], (d) graph cuts [14], (e) Vese–Chan [19].

Figure 5 illustrates the fact that the proposed method was the only one that was able to segment the veins of a leaf compared to others. The graph cut method was only able to segment the primary vein while leaving smaller ones out. The four-phase Vese–Chan method was stuck at a local minimum and failed to produce a reasonable segmentation. Somewhat similarly to the graph-cut approach, the fuzzy segmentation algorithm did well in segmenting the big scale veins, but failed to detect the smaller ones. The ‘Leaf’ image has a size of 128 by 128; the parameters used in simulations are $\epsilon = 1/22$, $\mu = 2$, $\lambda = 0.5\epsilon$.

The noisy ‘Peppers’ image in Figure 6 is a tricky one to segment due to inhomogeneous illumination. The two large peppers in the middle have very bright spots. The segmentation result of the proposed method is more consistent with human perception. The graph cut and the fuzzy segmentation results have leftover artificial noise, while the Vese–Chan method fails to segment the image into four classes. The parameters of simulation were similar to the ones used for the MRI image except for the edge fidelity that had to be decreased due to the noise present.

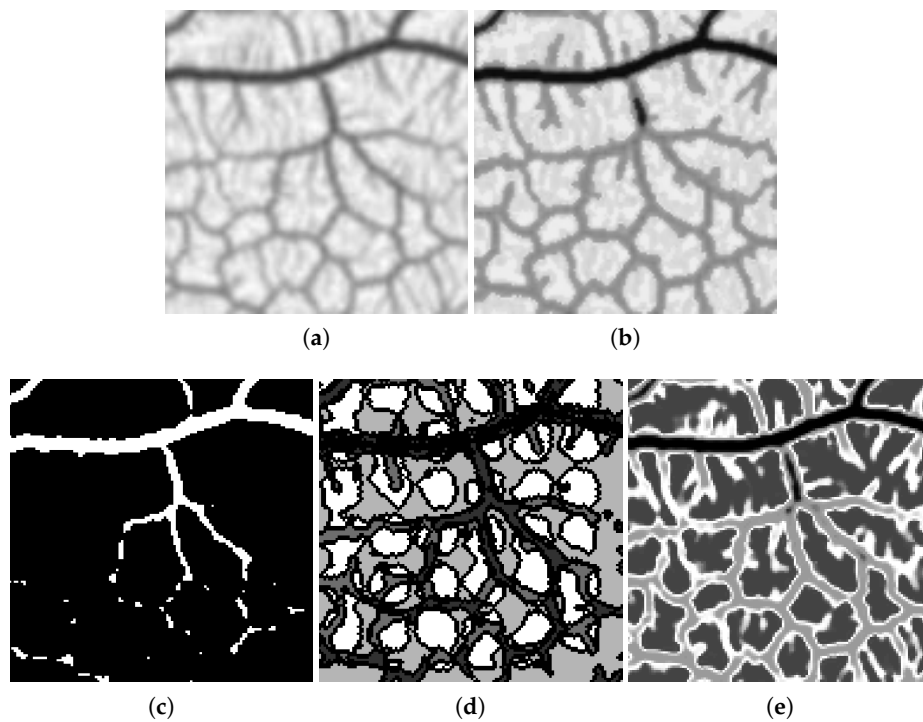


Figure 5. Segmentation of the 'Leaf' image: (a) the original image; the segmented outputs of (b) the proposed method, (c) graph cut method, (d) Vese–Chan and (e) fuzzy segmentation methods.

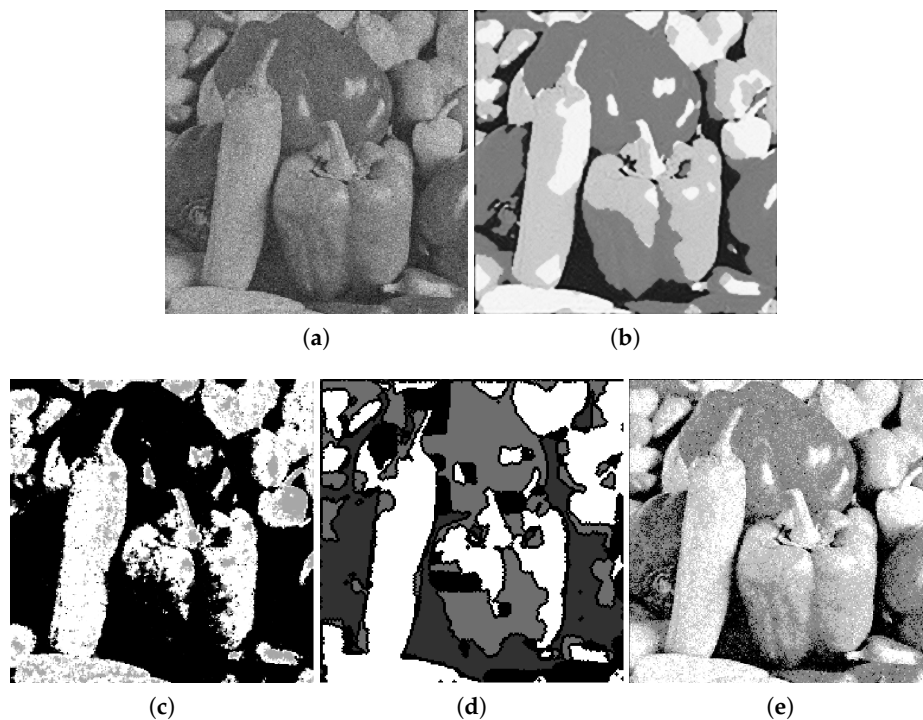


Figure 6. Compare the segmentation results of (a) the peppers image with added Gaussian noise (SNR = 15 dB) using (b) the proposed method, (c) graph cut, (d) Vese–Chan and (e) fuzzy segmentation methods.

Our next numerical segmentation example is segmenting a four-color grayscale image after adding various levels of noise to the original image ‘Sectors’ is shown in Figure 7. Without any noise present, the segmentation precision is more than 99.9% (of pixels are classified correctly), and the vector of constants $\vec{c}$ is recovered within <0.05% of the actual values.

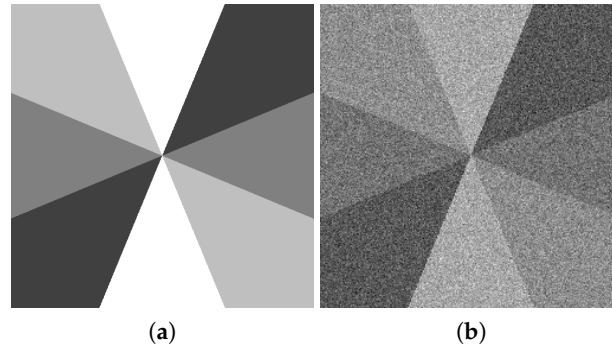


Figure 7. (a) The original four-valued image (b) with added noise.

The robustness of our segmentation method to noise is illustrated in Figure 8: the three images in each row show the noisy input image, direct output of the minimization routine and the result of rounding it to the nearest of the four recovered values c_i . The rows correspond to the increasing levels of Gaussian noise added to the original image: 15 dB, 10 dB and 5 dB. The figure caption also contains information about the percentage of correctly-classified pixels.

As the noise level increased, it was necessary to decrease the weight of the edge fidelity term and increase the spatial weighting term. For the ‘Sectors’ image with added noise of SNR = 10 dB, the simulation parameters were $\epsilon = \frac{1}{30}$, $\mu = 5$, $\lambda = 0.02\epsilon$. As was mentioned in Section 1.2, the WGL energy is not anisotropic. However, for practical purposes, this anisotropy is not affecting the minimization results in any significant way. It is clearly not an issue when we process natural images, where one cannot visually discern straightforward directional patterns. The performance of the method on test images, such as the ‘Sectors’ example, which contains radial edges at a variety of angles, shows the absence of the impact of the WGL anisotropy on the recovery quality. Since the ground truth segmentation for this image was available, our method’s performance was also evaluated with respect to such benchmarks as the Rand index, variation of information and segmentation cover (described in Section 2.3 in [31] and characterized as appropriate methods for evaluating region-oriented, rather than contour-oriented benchmarks); see Table 1.

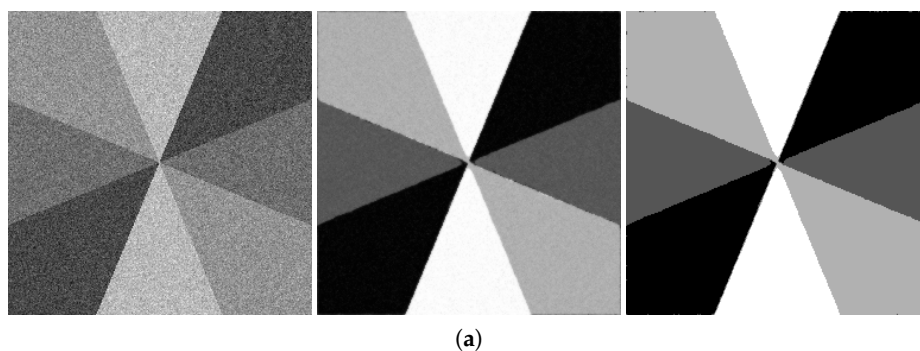


Figure 8. Cont.

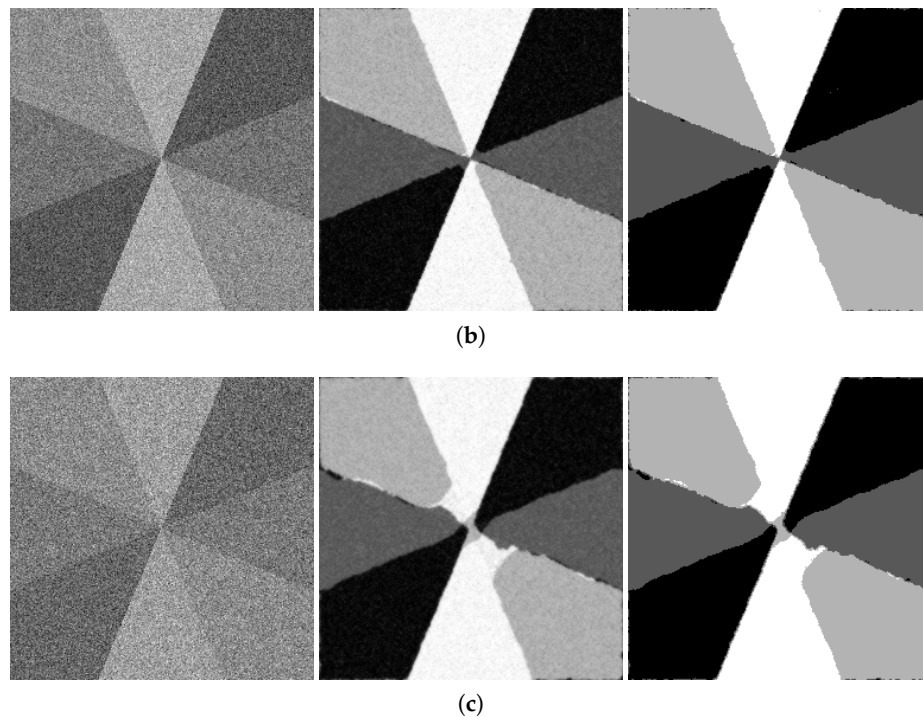


Figure 8. Each row ((a) to (c)) shows the input image, the minimization output and the thresholded (rounded) output (from left to right) for different levels of additive noise: (a) noise = 15 dB, >99% classified correctly, (b) noise = 10 dB, >98% classified correctly, (c) noise = 5 dB, >95% classified correctly.

Table 1. Evaluation of the performance for various levels of SNR (comparison to ground truth).

SNR/Benchmark	Rand Index	Variation of Information	Segmentation Covering
15 db	0.9881	0.1903	0.9763
10 db	0.9846	0.2382	0.9716
7 db	0.9807	0.2991	0.9618
5 db	0.8901	0.5099	0.9533
3 db	0.9377	0.7165	0.8759

Figure 9 shows the comparison of our method with others when applied to the ‘Sectors’ image with a noise level of 10 dB. We see that the graph cut method could not fight the relatively high level of noise, and the result is still polluted with it. The k-means output is also extremely fragmented. The four-phase Vese–Chan method ended up at a local minimum far from the desired output. The fuzzy segmentation method output looks fine except for the noise still noticeable all over the output image. The same holds for the graph-cut technique.

The quantitative comparison is provided in Table 2 below.

Table 2. Segmentation evaluation for several methods applied to the ‘Sectors’ image with 10 dB noise.

Method	Rand Index	Variation of Information	Segmentation Covering
Proposed method	0.9846	0.2382	0.9716
Fuzzy	0.7253	2.3699	0.4622
Vese–Chan	0.8901	1.1218	0.774
Graph Cuts	0.7523	1.1897	0.2711
k-means	0.7081	2.6804	0.4137

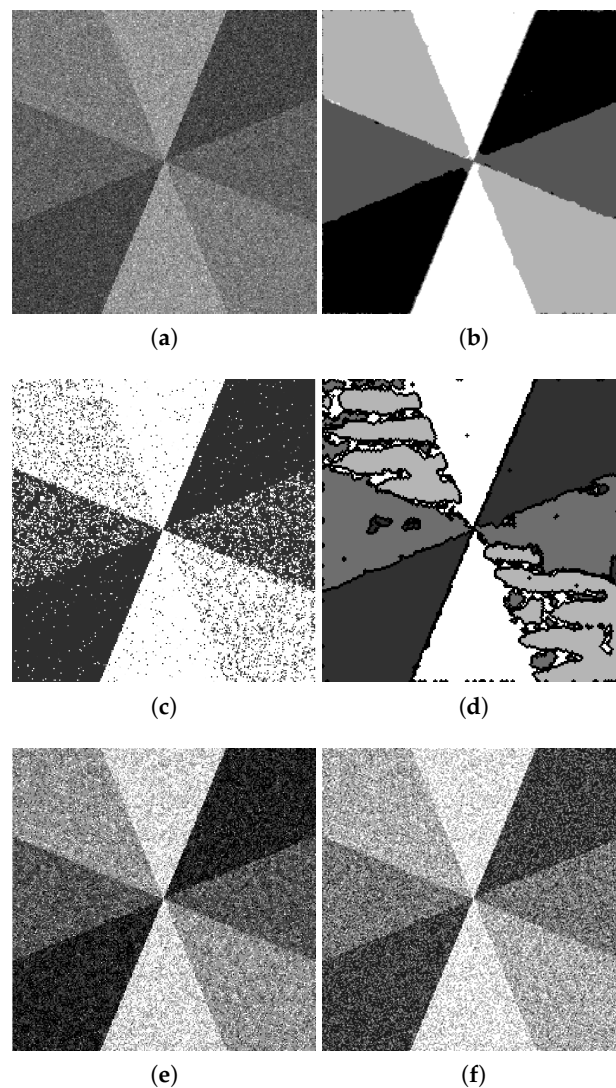


Figure 9. Comparison to other methods: (a) the input image (noisy ‘Sectors’ images); segmentation output of (b) the proposed method, (c) graph cut method, (d) Vese–Chan method, (e) fuzzy segmentation method, (f) k-means.

We include the segmentation results of some images from the Berkeley segmentation database ([60]); see Figure 10. However, it must be noted that our method is not designed to address the problem of object vs. background detection. It is aiming at classifying the pixels of an image into classes while balancing out the requirements on the output such as the mean-square color similarity within one class, preserving the chosen edges from the original image and the “driving force” of the method: the phase separation and coarsening that lead to grouping pixels into major connected components with a piecewise smooth boundary.

We performed tests on images of sizes $2^n \times 2^n$ (for the convenience of applying the wavelet transform), for $n = 7, 8, 9$. Depending on the image and the choice of parameters, the simulation (which aims at achieving an energy minimizing steady state, i.e., finding a local minimizer) might require between 100 and several hundreds of iterations. The simulations were performed using a PC (Win 7) with Intel i5-3570K CPU 3.40 GHz, 8 Gb RAM, operating system Windows 7, in MATLAB with no parallel computing optimization in our code. It takes about 30 s per 100 iterations for a 256 by 256 image with the maximum level of wavelet decomposition (five scales) w.r.t. the Daubechies-four wavelet (‘db4’). The processing can be significantly sped up by preprocessing, such as preliminary

denoising, but we are not addressing this issue here; all simulations were initialized automatically as described in Section 3.2.

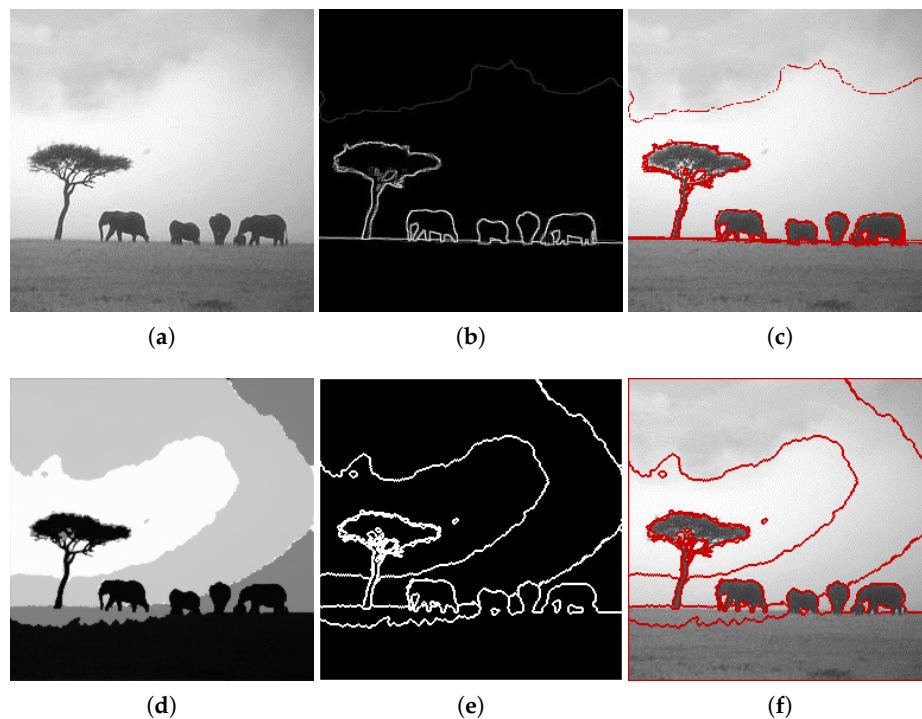


Figure 10. (a) Original image (from BSDS500), (b) ground truth boundaries obtained by manual segmentation (from BSDS500), (c) ground truth boundaries shown on the original, (d) the result of the proposed four-class segmentation method, (e) boundaries of the regions shown in (d), (f) boundaries of the segmented regions shown on the original image.

4.2. Blood Vessel Detection in Medical Images

The detection of blood vessels has been an important problem for a while. Most of the successful algorithms are based on machine learning and require a training set of images of comparable quality for successful processing ([56], and many other publications).

The output of our segmentation algorithm applied to an image with clear detalization (see Figure 5a,b) shows the advantage of the multi-phase segmentation applicable to the blood vessel detection, since the vessels can actually belong to two classes out of four classes due to highly variable intensity over the entire image. In a way, the detection becomes about the contrast of the blood vessels versus the background, rather than relying purely on the image intensity values. Here, the edge preserving fidelity term plays an important role. However, the images of leaves with clearly pronounced vessels are different in quality from the bio-medical images that typically contain blood vessels. The detalization in bio-medical images is often much worse due to the difficulties of non-invasive acquisition.

A good example is provided by retinal images, where the detection of blood vessels is sometimes needed to exclude the respective pixels from consideration during automatic or semi-automatic processing (for instance, the ratio of intensities at the green and yellow frequencies provides a reasonable estimation for the density of macular pigment, useful in the diagnostics of age-related macular degeneration [57]).

We consider two examples of our segmentation algorithm applied to retinal images.

The first image we process is a cut-out from a relatively high resolution image, and the capillary structures captured there are barely visible, just because the actual size of those vessels is small relative

to the image resolution. The second image contains blood vessels of more various scales, and thus, the corresponding pixels vary more in intensity.

To ‘even out’ the background, we preprocessed the images by removing the coarsest component of the wavelet decomposition:

$$u_{new} = u - \int \langle u, \phi_{0,\kappa} \rangle \phi_{0,\kappa} d\kappa = \sum_{j=0}^{+\infty} \int \langle u, \psi_{j,\kappa} \rangle \psi_{j,\kappa} d\kappa.$$

In terms of practical implementation, we ‘zeroed out’ the coarsest approximation coefficients in the stationary wavelet decomposition of each image. Now, we can expect the blood vessels to be contained within the set of pixels of the two darkest shades of the segmented image, i.e., those with intensities c_1 and c_2 .

Due to the fact that some components of the segmented outputs are actually wider than the blood vessels, thus merging capillaries with shadows or parts of uneven background, we use the “skeleton” feature of the MATLAB morphological function for binary processing, thus extracting the central line of each component of the set with the characteristic function we detected from the gradient descent output: $\chi = \max\{\text{sign}(\phi_1), \text{sign}(\phi_2)\}$. To emphasize the center lines corresponding to the darker segments, we widened them, as shown in Figure 11d.

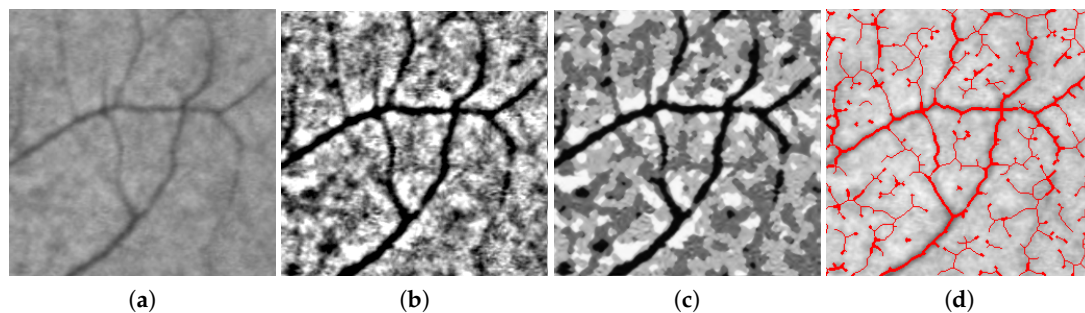


Figure 11. (a) Original image, (b) image after background removal (c) segmented output, (d) segmented output with post-processing.

4.3. Color Image Segmentation

Color image segmentation falls into a more general category of vector-valued image segmentation, which has been studied for a while. As was mentioned in the Introduction, each segmentation task has its consistency requirements and the expected output properties. We extend our technique to the color images with the purpose of obtaining an output with a certain lower bound on the size of connected components of segmented regions and the piecewise regularity of the boundary of those connected components.

In the case of color image segmentation, the initial guess becomes even more important (due to the general non-convexity of the minimized energy and the increased complexity of the domain of $\vec{c}$). Furthermore, it becomes impossible to set the initial condition by elementary percentile-based thresholding; therefore, we propose to use our method as a refining ‘post-processing tool’, after obtaining the initial guess by some relatively simple method like k-means. The “crude” method provides a guess that is somewhat correlated with the desired output I^0 , $\vec{c}^0$, and our variational method refines its output by running the gradient descent algorithm minimizing the energy E we introduced earlier, thus producing a refined output. We chose to use k-means for obtaining the initial guess because of its straightforward relation to our energy: applying k-means is equivalent to minimizing the spatial fidelity term $\mu \|u - f\|_{L^2}^2$.

For simplicity, we assume that the input color images are square and are provided in the RGB format; thus, we are to process a 3D array I of a size $N \times N \times 3$ of real values $I(x, y, clr)$, where

(x, y) is the spatial pixel coordinate, clr is an integer index between one and three corresponding to ‘red’, ‘green’ or ‘blue’, respectively. We interpret the color image I as the set of values of a function $f: [0, 1]^2 \rightarrow \mathbb{R}^3$. Just as in the 2D case, we look for the segmented version of the given image $I(x, y, clr)$ in the form:

$$u = \vec{c}_1\phi_1\phi_2 + \vec{c}_2\phi_1(1 - \phi_2) + \vec{c}_3(1 - \phi_1)\phi_2 + \vec{c}_4(1 - \phi_1)(1 - \phi_2), \text{ where } \vec{c}_1, \dots, \vec{c}_4 \in \mathbb{R}^3. \quad (18)$$

Unknown functions ϕ_1 and ϕ_2 and the vectors $\vec{c}_1, \dots, \vec{c}_4$ are determined, just as in the 2D case, as those minimizing the energy function:

$$E(\phi_1, \phi_2, \vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_4) = WGL(\phi_1) + WGL(\phi_2) + \mu\|u - f\|_{L^2}^2 + \lambda|Pr_\Lambda(u - f)|_B^2. \quad (19)$$

Here, we use the norm and the seminorm for multi-valued functions, and the set of ‘significant’ wavelet modes Λ can be determined from the grayscale version of the image or for each color channel individually.

Let us emphasize what parts of the algorithm are essentially 3D and discuss how the two characteristic functions of the specific 2D sets come into the four-phase color segmentation scenario. Below is a brief intuitive explanation of our segmentation algorithm adapted to a vector-valued case.

The main relation between the three color channels is provided by the functions ϕ_1 and ϕ_2 , functions of two (spatial) arguments, with no direct dependence on the third, color-related coordinate. The actual color information of the segmented version of the segmented image is stored in the matrix $C(j, clr)$, $j \in \{1, \dots, 4\}$ (four constant colors in the output image), $clr \in \{1, 2, 3\}$ (RGB component indexing). As we noticed before, due to the diffuse interface nature of the method, ϕ_1 and ϕ_2 are not exactly binary, but rather nearly-binary: for the purpose of our intuitive explanation, let us assume $\phi_1(x, y) \approx \chi_{E_1}(x, y)$, $\phi_2(x, y) \approx \chi_{E_2}(x, y)$; then (again, in some approximate sense) the segmented image takes on the value $C(1, :)$ on $E_1 \cap E_2$, value $C(2, :)$ on $E_1 \setminus E_2$, value $C(3, :)$ on $E_2 \setminus E_1$ and the value $C(4, :)$ on $[0, 1]^2 \setminus (E_1 \cup E_2)$.

After the initial values of the color constants stored in C and the corresponding sets E_1 and E_2 are assigned, we run the gradient descent algorithm, updating all of the above at every step, until an equilibrium is attained.

Figure 12 shows an example of four-phase color image segmentation, where the initial guess was obtained by the k-means method (with Euclidean distance).

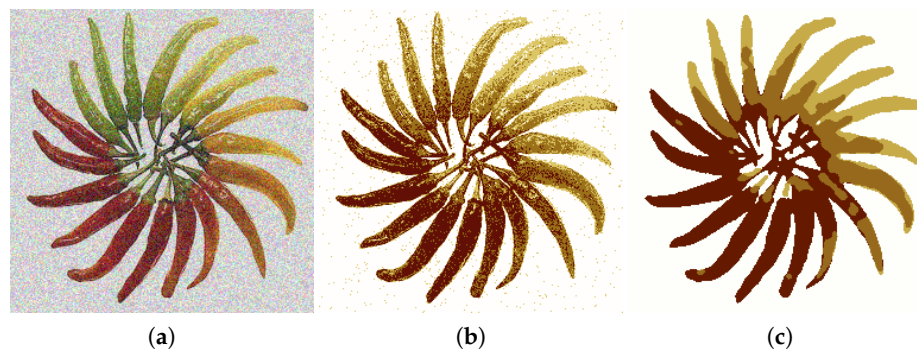


Figure 12. (a) Original image, (b) k-means segmentation that is used to initialize ϕ_1 and ϕ_2 (c) segmentation output.

5. Conclusions

We proposed a variational multi-phase segmentation approach based on the wavelet Ginzburg–Landau energy as the regularizer and thus free of partial differential equations. Our

model is based on sparse representation and explicitly involves an edge-enhancing term to better segment components. Compared to PDE-based models that are local, the proposed approach involves diffuse interface behaviors, such as coarsening and phase separation that are all nonlocal and create better connections between image segments. The proposed model works effectively on both scalar and vector-valued images that have low contrast, noise and details of multi-direction and multi-scale.

The method can be further generalized by using a system (frame) other than the classical separable wavelet frame, such as composite wavelets. Details of the design for such WGL-like regularization part of the energy can be found [58]. Even though we only discussed the specifics of the four-phase segmentation, we can extend this approach to segmentation of an image into 2^n phases using n functions ϕ and further this approach for multi-dimensional data processing and classification.

Acknowledgments: Weihong Guo is partially supported by NSF DMS1521582.

Author Contributions: The energy for the variational model discussed in this paper was designed by both authors. J.D. developed the theory and methodology, performed the numerical experiments and described those. W.G. evaluated the results against the state of the art segmentation methods, described those and highlighted the advantages. The manuscript editing was a mutual effort.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Some Additional Comments

Appendix A.1. Existence of Minimizers

The following theorem proves the existence of local minimizers of the energy introduced in Section 2.1. The proof employs the strategy of a standard “compactness argument”. Whenever the domain of integration is omitted, we assume the integration over the entire spatial domain $\Omega = [0, 1]^2$, the same applied to the function spaces with the omitted function domain, i.e., $B = B(\Omega)$, $L^p = L^p(\Omega)$, etc.

Theorem A1. *Consider the energy functional:*

$$E(\phi_1, \phi_2, \vec{p}) = WGL(\phi_1) + WGL(\phi_2) + \mu \|u - f\|_{L^2}^2 + \lambda |Pr_\Lambda(u - f)|_B^2.$$

where:

$$WGL(v) = \frac{\epsilon}{2} |v|_B^2 + \frac{1}{4\epsilon} \int W(v) dx,$$

$$u(x) = p_1^2 \phi_1(x) \phi_2(x) + p_2^2 \phi_1(1 - \phi_2(x)) + p_3^2 (1 - \phi_1(x)) \phi_2(x) + p_4^2 (1 - \phi_1(x))(1 - \phi_2(x)),$$

$\phi_1, \phi_2 \in B(\Omega)$, $\vec{p} \in [0, 1]^4$ and $f \in B \cap L^\infty$. Then, E has at least one minimizer over the space $X = B \otimes B \otimes [0, 1]^4$, which is not necessarily unique.

Remark A1. *In practice, the dependence of the recovered minimizer on the initial condition actually plays to our advantage, allowing to use our method as a refining tool, recovering the segmented output relatively close to a chosen initial guess. Our numerical examples show successful segmentation with percentile-based initial guess (Section 3.2, Appendix, B.2) or initial guess obtained by a k-means technique (for color images, Section 4.3).*

Proof. The energy functional E is non-negative and, hence, has a finite infimum $0 \leq M < \infty$. Let us prove the infimum is attained on an element from the admissible set(space) X . Consider a minimizing sequence $\{U_n\}$, where $U_n = \{\phi_{1n}, \phi_{2n}, \vec{p}_n\} : E(U_n) \rightarrow M$ as $n \rightarrow \infty$. The energy values $E(U_n)$ converge, hence are bounded by some positive constant ($\exists A > 0$ s.t. $\forall n \in \mathbb{N} E(U_n) \leq A$), and so are the seminorms $|\phi_{1n}|_B$ and $|\phi_{2n}|_B$.

The corresponding norms $\|\phi_{i_n}\|_B$ can be estimated from above using the bound on the integral of the double-well potential (the same argument works for both ϕ_1 and ϕ_2 , so we use ϕ_i to indicate the function in question in the following parts of the solution):

$$\int W(\phi_{i_n})dx = \int (\phi_{i_n}^2 - \phi_n)^2 \leq C \Rightarrow \left| \int \phi_{i_n}^2 - \phi_n \right| \leq C', \quad \|\phi_{i_n} - \frac{1}{2}\|_{L^2}^2 \leq C' + \frac{1}{4},$$

thus $\|\phi_{i_n}\|_{L^2} \leq C''$ (here, C', C'' are positive constants). Therefore,

$$\|\phi_{i_n}\|_B^2 \leq \|\phi_{i_n}\|_B^2 + \|\phi_{i_n}\|_{L^2}^2$$

is also bounded. Notice also that:

$$\|\phi_{i_n}\|_{L^4}^2 \leq \sqrt{C} + \|\phi_{i_n}\|_{L^2} \leq \sqrt{C} + C''.$$

Since X is Banach and $\{\phi_{i_n}\}$ is bounded, it has a weakly converging subsequence: $\{\phi_{i_{n_k}}\}: \phi_{i_{n_k}} \rightharpoonup \tilde{\phi}_i \in B$. As we noticed above, the sequence is also bounded in L^4 , so, WLOG, we can assume that $\phi_{i_{n_k}}$ converges weakly in L^4 , as well, and $\{\tilde{p}_{n_k}\}$ converges to an element in $\tilde{p} \in \mathbb{R}^4$. Fix some $U_{n_{k*}}$; let $u_{n,k}$ and $u_{n,k*}$ denote corresponding functions obtained as in (5). Then:

$$\begin{aligned} & \sum_{i=1,2} \left(\frac{\epsilon}{2} \langle \phi_{i_{n_k}}, \phi_{i_{n_k}} \rangle_B + \frac{1}{4\epsilon} \int (\phi_{i_{n_k}}^2 - 1)(\phi_{i_{n_{k*}}}^2 - 1)dx \right) + \frac{\mu}{2} \langle u_{n_k} - f, v - f \rangle_{L_2(\Omega)}^2 + \\ & + \frac{\lambda}{2} \langle P_\Lambda(u_{n_k} - f), P_\Lambda(v - f) \rangle_B^2 \rightarrow \frac{\epsilon}{2} \langle u, v \rangle_B + \frac{1}{4\epsilon} \int (u^2 - 1)(v^2 - 1)dx + \\ & + \frac{\mu}{2} \langle u - f, v - f \rangle_{L_2(\Omega)}^2 + \frac{\lambda}{2} \langle P_\Lambda(u - f), P_\Lambda(v - f) \rangle_B^2 \end{aligned}$$

Now, if we let $k, k* \rightarrow \infty$,

$$\begin{aligned} & \frac{\epsilon}{2} \langle \phi_{i_{n_k}}, \phi_{i_{n_{k*}}} \rangle_B + \frac{1}{4\epsilon} \int (\phi_{i_{n_k}}^2 - 1)(\phi_{i_{n_{k*}}}^2 - 1)dx + \frac{\mu}{2} \langle \phi_{i_{n_k}} - f, \phi_{i_{n_{k*}}} - f \rangle_{L_2(\Omega)}^2 + \\ & + \frac{\lambda}{2} \langle P_\Lambda(\phi_{i_{n_k}} - f), P_\Lambda(\phi_{i_{n_{k*}}} - f) \rangle_B^2 \rightarrow E(u). \end{aligned}$$

Taking the diagonal subsequence with $k = k*$, we see that $E(\phi_{i_{n_k}}) \rightarrow E(u)$ as $k \rightarrow \infty$. Hence, the minimum of the WGLF energy is achieved at the element $u \in B$. $\square$

Appendix A.2. Remarks about Determining the Initial Values of $\vec{c}$ from the Histogram of the Image Intensity Values

In some cases, analysis of a histogram of the image that we need to segment allows determining four clearly dominant values. These values can be either assigned as the constants c_1, c_2, c_3, c_4 associated with the four “phases” of the segmented output, or serve as an initial guess.

However, it is not always possible to do so. For example, after adding Gaussian noise to an image with only four intensity values in its range, the overall intensity distribution changes from four equal spikes on the histogram to the distribution shown in Figure A1.

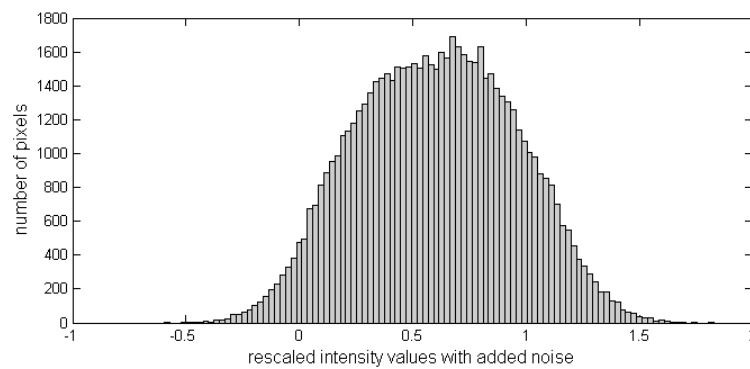


Figure A1. The histogram of the four-valued image after Gaussian noise had been added.

If we tried to segment a “clean” four-valued image after determining the values of $\vec{c}$ from the histogram, we would have recovered it perfectly. With the noisy images, however, we need to use the described percentile initialization.

The minimizing values of the vector $\vec{c}$ that we recover can deviate from the original ones, due to the fact that adding the Gaussian noise affects the range of the image intensity values. When estimating the classification accuracy, we compared the resulting four classes by matching the corresponding components (since in both the original and the recovered vectors $\vec{c}$ the components are organized in the ascending order).

Appendix B. Additional Details Regarding Numerical Simulations

Appendix B.1. More Examples

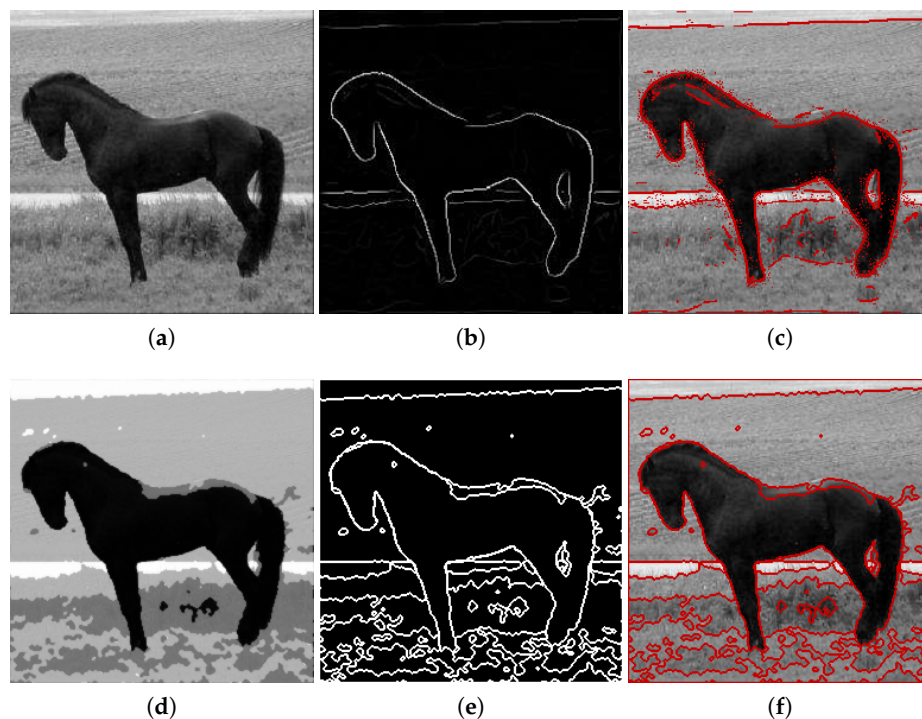


Figure A2. (a) Original ‘Horse’ image, (b) ground truth boundary information (from [59]), (c) images (a,b) overlaid, (d) our algorithm output, (e) boundaries of the segmented regions, (f) boundaries for the segmented regions overlaid on the original image.

We can see that the result of the ‘Horse’ image segmentation produced region boundaries that contain the boundaries of the horse silhouette against the background. However, in the case of the ‘Plane’ image, the sky takes most of the image, which biases the segmentation process to a more precise segmentation of the sky rather than the plane.

Both images are 256 by 256 pixels, parameters for processing: $\epsilon = 1/35$, $\mu = 3$, $\lambda = 0.75 * \epsilon$, the edge sets shown in Figure A4f,g.

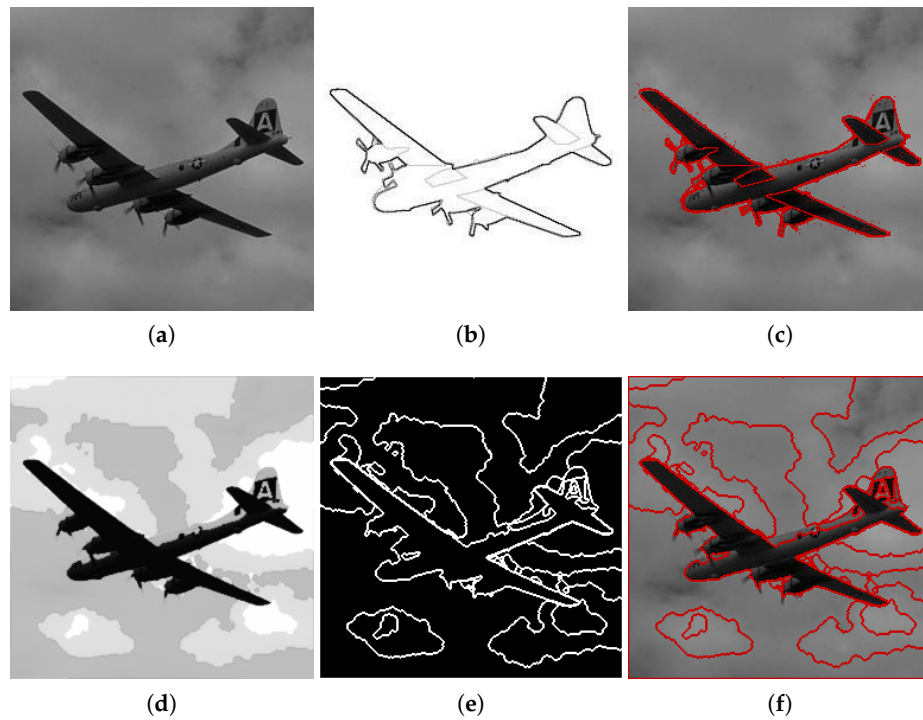


Figure A3. (a) Original ‘Plane’ image (from Berkeley Segmentation Data Set 300 [60]), (b) ground truth boundary information (also from BSDS300), (c) images (a,b) overlaid, (d) our algorithm output, (e) boundaries of the segmented regions, (f) boundaries for the segmented regions overlaid on the original image.

Appendix B.2. Edge Information Used in the Fidelity Terms for the Images from Sections and B.1

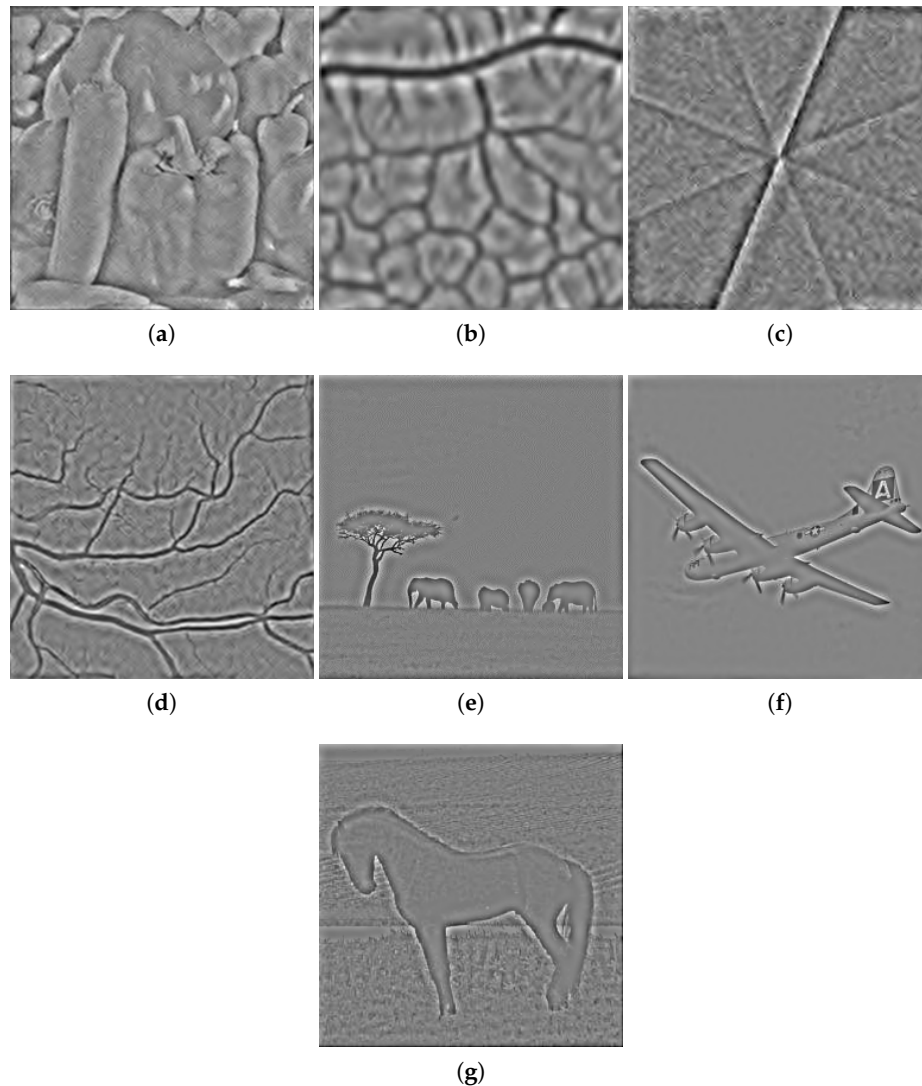


Figure A4. Edge information used in the fidelity term for (a) the ‘Peppers’ image with added noise, SNR = 15 dB (b) ‘Leaf’, (c) the ‘Sectors’ image with added noise, SNR = 10 dB, (d) retinal image, (e) ‘Elephants’, (f) ‘Plane’, (g) ‘Horse’.

References

1. Lindeberg, T.; Li, M.X. Segmentation and classification of edges using minimum description length approximation and complementary junction cues. *Comput. Vis. Image Underst.* **1997**, *1*, 144–151.
2. Boykov, Y.; Kolmogorov, V. An Overview of Image and Video Segmentation in the last 40 years. In Proceedings of the 6th International Symposium on Signal Processing and Its Applications, Kuala Lumpur, Malaysia, 13–16 August 2001; pp. 144–151.
3. Pal, S.K.; Ghosh, A.; Shankar, B.U. Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. *Int. J. Remote Sens.* **2000**, *21*, 2269–2300.
4. Roerdink, J.; Meijster, A. The Watershed Transform: Definitions, Algorithms and Parallization Strategies. *Fundam. Inform.* **2001**, *41*, 187–228.
5. Beucher, S.; Meyer, F. The Morphological Approach to Segmentation: The Watershed Transformation. *Mathematical Morphology in Image Processing*; Marcel Dekker Inc.: New York, NY, USA, 1993; pp. 433–481.

6. Cheng, J.; Rajapakse, J.C. Segmentation of Clustered Nuclei With Shape Markers and Marking Function. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 741–748.
7. Zhang, C.; Sun, C.; Su, R.; Pham, T.D. Segmentation of clustered nuclei based on curvature Weighting. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, Dunedin, New Zealand, 26–28 November 2012; IVCNZ, McCane, B., Mills, S., Deng, J.D., Eds.; ACM: New York, NY, USA, 2012; pp. 49–54.
8. Atta-Fosu, T.; Guo, W.; Jeter, D.; Mizutani, C.; Stopczynski, N.; Sousa-Neves, R. 3D Clumped Cell Segmentation Using Curvature Based Seeded Watershed. *J. Imaging* **2016**, *2*, doi:10.3390/jimaging2040031.
9. Potts, R.B. Some Generalized Order-Disorder Transformations. *Math. Proc.* **1952**, *48*, 106–109.
10. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239.
11. Boykov, Y.; Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137.
12. Appleton, B.; Talbot, H. Globally Minimal Surfaces by Continuous Maximal Flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 106–118.
13. Yuan, J.; Bae, E.; Tai, X.; Boykov, Y. A Continuous Max-Flow Approach to Potts Model. In Proceedings of the 11th European Conference on Computer Vision Heraklion, Greece, 5–11 September 2010; pp. 379–392.
14. Bae, E.; Yuan, J.; Tai, X.C. Global minimization for continuous multiphase partitioning problems using a dual approach. *Int. J. Comput. Vis.* **2011**, *92*, 112–129.
15. Yuan, J.; Bae, E.; Tai, X.; Boykov, Y. A spatially continuous max-flow and min-cut framework for binary labeling problems. *Numer. Math.* **2014**, *126*, 559–587.
16. Mumford, D.; Shah, J. Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **1989**, *42*, 557–685.
17. Caselles, V.; Kimmel, R.; Sapiro, G. Nonlinear total variation based noise removal algorithms. *Int. J. Comput. Vis.* **1997**, *22*, 61–79.
18. Vese, L.; Chan, T. Active contours without edges. *IEEE Trans. Image Process.* **2001**, *10*, 266–277.
19. Vese, L.; Chan, T. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *Int. J. Comput. Vis.* **2002**, *50*, 271–293.
20. Kang, S.; March, R. Multiphase Image Segmentation via Equally Distanced Multiple Well Potential. *J. Vis. Commun. Image Represent.* **2014**, *25*, 1446–1459.
21. Bresson, X.; Esedoglu, S.; Vandergheynst, P.; Osher, S. Fast global minimization of the active contour/snake model. *J. Math. Imaging Vis.* **2007**, *28*, 151–167.
22. Pock, T.; Schoenemann, T.; Graber, G.; Bischof, H.; Cremers, D. A convex formulation of continuous multi-label problems. In Proceedings of the 10th European Conference on Computer Vision: Part III, Marseille, France, 12–18 October 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 792–805.
23. Pock, T.; Chambolle, A.; Cremers, D.; Bischof, H. A convex relaxation approach for computing minimal partitions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 810–817.

24. Lellmann, J.; Kappes, J.; Yuan, J.; Becker, F.; Schnörr, C. Convex multi-class image labeling by simplex-constrained total variation. In Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision, Voss, Norway, 1–5 June 2009; Springer: Berlin, Germany, 2009; Volume 5567, pp. 150–162.
25. Lellmann, J.; Schnörr, C. Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.* **2011**, *4*, 1049–1096.
26. Brown, E.S.; Chan, T.F.; Bresson, X. Completely convex formulation of the Chan-Vese image segmentation model. *Int. J. Comput. Vis.* **2012**, *98*, 103–121.
27. Papadakis, N.; Aujol, J.-F.; Caselles, V.; Yildizoglu, R. High-dimension multi-label problems: Convex or non convex relaxation. *SIAM J. Imaging Sci.* **2013**, *6*, 2603–2639.
28. Li, B.; Chui, C.; Chang, S.; Ong, S. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. *Comput. Biol. Med.* **2011**, *41*, 1–10.
29. Garcia-Cardona, C.; Merkurjev, E.; Bertozzi, A.; Flenner, A.; Percus, A.G. Multiclass Data Segmentation Using Diffuse Interface Methods on Graphs. *IEEE Trans. Pattern Anal. Mach. Int.* **2014**, *36*, 1600–1613.
30. Merkurjev, E.; Garcia, C.; Bertozzi, A.L.; Flenner, A.; Percus, A. Diffuse interface methods for multiclass segmentation of high-dimensional data. *Appl. Math. Lett.* **2014**, *63*, 29–34.
31. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916.
32. Unnikrishnan, R.; Pantofaru, C.; Hebert, M. Toward Objective Evaluation of Image Segmentation Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 929–944.
33. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV), Vancouver, BC, Canada, 7–14 July 2001.
34. Shi, R.; Ngan, K.; Li, S.; Paramesran, R.; Li, H. Visual Quality Evaluation of Image Object Segmentation: Subjective Assessment and Objective Measure. *IEEE Trans. Image Process.* **2015**, *24*, 5033–5045.
35. Shi, R.; Ngan, K.; Li, S. The Objective Evaluation of Image Object Segmentation Quality. In Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS), Poznan, Poland, 28–31 October 2013.
36. Dobrosotskaya, J.; Guo, W. A PDE-free variational model for multiphase image segmentation. In Proceedings of the Conference on Wavelets and Sparsity XVI, San Diego, CA, USA, 10–12 August 2015; Volume 9597, p. 959706.
37. Emmerich, H. *Diffuse Interface Approach in Materials Science Thermodynamic Concepts and Applications of Phase-Field Models*; Springer: Berlin/Heidelberg, Germany, 2003.
38. Jung, Y.; S.H.Kang.; Shen, J. Multiphase image segmentation via Modica-Mortola phase transition. *SIAM J. Appl. Math.* **2007**, *67*, 1213–1232.
39. Modica, L.; Mortola, S. Un esempio di gamma-convergenza. *Boll. Unione Mat. Ital.* **1977**, *5*, 285–299.
40. Bertozzi, A.; Esedoglu, S.; Gillette, A. Analysis of a two-scale Cahn-Hilliard model for image inpainting. *Multiscale Model. Simul.* **2007**, *6*, 913–936.
41. Esedoglu, S. Blind Deconvolution of Bar Code Signals. *Inverse Probl.* **2004**, *20*, 121–135.
42. Wang, M.Y.; Zhou, S.; Ding, H. Non-linear diffusions in topology optimization. *Struct. Multidiscip. Optim.* **2004**, *28*, 262–276.
43. Favaro, P.; Osher, S.; Soatto, S.; Vese, L.A. 3D Shape from Anisotropic Diffusion. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Madison, WI, USA, 18–20 June 2003; Volume 1, pp. 179–186.
44. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *1*, 629–639.
45. Dobrosotskaya, J.; Bertozzi, A. Analysis of the wavelet Ginzburg–Landau energy in image applications with edges. *SIAM J. Imaging Sci.* **2013**, *6*, 698–729.
46. Dobrosotskaya, J.; Bertozzi, A. Wavelet analogue of the Ginzburg–Landau energy and its Γ -convergence. *Interfaces Free Bound.* **2010**, *12*, 497–525.
47. Chan, T.F.; Shen, J.; Zhou, H.M. Total variation wavelet inpainting. *J. Math. Imaging Vis.* **2006**, *25*, 107–125.

48. Elad, M.; Starck, J.L.; Querre, P.; Donoho, D.L. Simultaneous cartoon texture image inpainting using morphological component analysis. *Appl. Comput. Harmon. Anal.* **2005**, *19*, 340–358.
49. Chambolle, A.; DeVore, R.A.; Lee, N.; Lucier, B.J. Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal through Wavelet Shrinkage. *IEEE Trans. Image Process.* **1998**, *7*, 319–333.
50. Cai, J.F.; Dong, B.; Osher, S.; Shen, Z. Image Restoration: Total Variation, Wavelet Frames, and Beyond. *J. Am. Math. Soc.* **2012**, *25*, 1033–1089.
51. Chambolle, A.; Lions, P.L. Image recovery via total variation minimization and related problems. *Numer. Math.* **1997**, *76*, 167–188.
52. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear Total Variation based noise removal algorithms. *Phys. D.* **1992**, *60*, 259–268.
53. Maso, G.D. *An Introduction to Gamma Convergence. Progress in nonlinear Differential Equations and Their Applications*; Birkhauser Boston, Inc.: Boston, MA, USA, 1993.
54. Eyre, D.J. An Unconditionally Stable One-Step Scheme for Gradient Systems. 1997. Available online: <https://pdfs.semanticscholar.org/4f01/6a98fe25bfc06b9bcab3d85eeaa47d3ad3ca.pdf> (accessed on 4 July 2017).
55. Mallat, S. *Wavelet Tour of Signal Processing*; Academic Press, London, UK, 1999.
56. Staal, J.; Abramoff, M.; Niemeijer, M.; Viergever, M.; van Ginneken, B. Ridge based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509.
57. Dobrosotskaya, J.; Ehler, M.; King, E.; Bonner, R.; Czaja, W. Modeling of the rhodopsin bleaching with variational analysis of retinal images. *SPIE Med. Imaging* **2011**, 7962, doi:10.1117/12.878709.
58. Czaja, W.; Dobrosotskaya, J.; Manning, B. Composite wavelet representations for reconstruction of missing data. *Proc. SPIE* **2013**, 8750, doi:10.1117/12.2019219.
59. Ren, X.; Fowlkes, C.C.; Malik, J. Cue Integration for Figure/Ground Labeling. In Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'05), Vancouver, BC, Canada, 5–8 December 2005; MIT Press: Cambridge, MA, USA, 2005; pp. 1121–1128.
60. D. Martin; C. Fowlkes; D. Tal; J. Malik A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In Proceedings of the 8th International Conference on Computer Vision, July 2001, Vol. 2; pp.416–423



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).