

## REPRODUCIBLE KERNEL HILBERT SPACE BASED GLOBAL AND LOCAL IMAGE SEGMENTATION

LIAM BURROWS<sup>1</sup>, WEIHONG GUO<sup>2</sup>, KE CHEN<sup>1</sup> AND FRANCESCO TORELLA<sup>3</sup>

<sup>1</sup> Centre for Mathematical Imaging Techniques and Department of Mathematical Sciences,  
University of Liverpool, United Kingdom

<sup>2</sup> Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve  
University, Cleveland, OH 44106, USA

<sup>3</sup> Liverpool Vascular & Endovascular Service, Royal Liverpool and Broadgreen University  
Hospitals NHS Trust, Liverpool, L7 8XP, United Kingdom

**ABSTRACT.** Image segmentation is the task of partitioning an image into individual objects, and has many important applications in a wide range of fields. The majority of segmentation methods rely on image intensity gradient to define edges between objects. However, intensity gradient fails to identify edges when the contrast between two objects is low. In this paper we aim to introduce methods to make such weak edges more prominent in order to improve segmentation results of objects of low contrast. This is done for two kinds of segmentation models: global and local. We use a combination of a reproducing kernel Hilbert space and approximated Heaviside functions to decompose an image and then show how this decomposition can be applied to a segmentation model. We show some results and robustness to noise, as well as demonstrating that we can combine the reconstruction and segmentation model together, allowing us to obtain both the decomposition and segmentation simultaneously.

**1. Introduction.** Image segmentation is a fundamental problem that has numerous applications in various fields. It is still a challenging problem when the contrast between the object of interest and the rest of the image is low and when there is dominant noise. The commonly used image gradient information is not effective in detecting weak edges and thus causes leakage in segmentation.

Variational segmentation methods have been successful for images with clear contrast. Depending on whether global or local/selective segmentation is the purpose, region or edge based segmentation methods have been studied. Many variational segmentation methods have roots in the Potts model [23], which is NP hard in multiphase segmentation. Graph cut methods approximate global minimizers in computationally-efficient ways [7, 6, 3, 29, 5, 30]. The Mumford–Shah model [21] is a discrete case of the Potts model that assumes piecewise smooth image intensity function. This is however difficult to implement. One difficulty comes from the extension of the piecewise smooth function from one sub-domain to another. The other comes from the non-regularity of the length term. Ambrosio and Tortorelli proposed an elliptic approximation [2, 1] to simplify the length term. Chan and Vese [12] simplified the piecewise-smooth Mumford-Shah model into a piecewise constant case with two phases.

---

2010 *Mathematics Subject Classification.* Primary: 58F15, 58F17; Secondary: 53C35.

*Key words and phrases.* RKHS, Heaviside function, Image segmentation.

To handle topology change, a common method is to use the popular level set method proposed by Osher and Sethian [22]. However, the incorporation of Heaviside function to represent the inside and outside domain makes the corresponding optimization problem nonconvex and easy to get stuck in local minimizers. Chan, Esedoglu and Nikolova [11] proposed a convex relaxation method to overcome this issue. Cai et al. [34] introduced two-stage segmentation, involving a convex variant of the Mumford-Shah model followed by thresholding to obtain the segmentation result. Various efforts have also been made to improve upon the data fidelity terms presented in the Chan-Vese model. One such method proposed by Li et al. [19] is the Region-Scalable Fitting method who used a Gaussian kernel in the fidelity to allow segmentation of an image containing inhomogeneous intensity. Another such method was proposed by Wang et al. [27], who proposed the Local Chan-Vese (LCV) model, which uses an averaging convolution kernel to provide a higher contrast image, allowing for easier segmentation.

The methods reviewed so far are mainly region based and lead to global segmentation. Another approach, able to achieve local segmentation results, are edge based methods. The first of which was developed by Kass et al. [18], which was further refined by Caselles et al. [9] in the Geodesic Active Contour model (GAC). The GAC model was used by Gout et al. [14], who added a distance constraint to encourage the contour to stay near a marker set input by the user to indicate which object to segment. Further improvements came from Badshah and Chen [4] who merged the Gout model with region terms in the form of the Chan-Vese fidelity, and this was later worked on by Spencer and Chen [26] who unbounded the distance constraint from the regulariser and added it as a separate term. Spencer and Chen used the Euclidean distance to put a penalty on objects far away from the target. However this method is rather sensitive to the placement of the markers, often requiring the markers to cover the entirety of the object and a close tuning of the parameter. In addition to proposing the above non-convex version, the authors also used the convex relaxation technique from [11] to obtain an equivalent convex model.

More recently, Roberts and Chen [24] proposed to change the Euclidean distance constraint from [26] to an edge weighted geodesic distance constraint. Moreover, in [25] Roberts and Spencer introduced a reformulated Chan-Vese fitting term for selective segmentation and making use of the geodesic distance.

Using this as a fitting term provides a much more suitable term for selective segmentation when compared with the classic Chan-Vese term, allowing for the background to be composed of multiple regions. In addition, the advantage of using the geodesic distance over the Euclidean distance from [26] is that it is much more robust to parameter selection and placement of the marker set. The geodesic distance increases when an edge is detected, thus is able to apply a penalty on objects outside the region of interest, making for a much more intuitive distance constraint.

Another recent approach for selective segmentation was proposed by Liu et al. [33], who have users to input markers along boundaries of the region of interest and put a weight related to distance to the marker points on the fidelity term of the method by Cai et al. [34]. Clearly edge information plays an important role in all these mentioned models.

The majority of region and edge based segmentation methods cannot handle images with low contrast/missing edge information. They often lead to over segmentation caused by leakage. In this work, we present how to use reproducible Kernel Hilbert space (RKHS) and approximated Heaviside functions to make faint edges more prominent, and thus enhance the segmentation results. We present both global and local versions of the proposed method. Extensive numerical results demonstrate their effectiveness and superiority compared to the state of the arts.

In the next section, we provide a brief review of some modeling tools to be used in the proposed methods.

## 2. Background review.

**2.1. Global segmentation.** As this paper aims to obtain an automatic edge enhancement method through RKHS, it is feasible to apply the idea to most of the commonly-used segmentation models. We demonstrate the global segmentation idea in the framework of convex relaxed model [11]:

$$\begin{aligned} F(u, c_1, c_2) = & \int_{\Omega} |\nabla u| d\mathbf{x} + \lambda_1 \int_{\Omega} (z(\mathbf{x}) - c_1)^2 u d\mathbf{x} \\ & + \lambda_2 \int_{\Omega} (z(\mathbf{x}) - c_2)^2 (1 - u) d\mathbf{x} + \xi \int_{\Omega} \nu(u) d\mathbf{x}, \end{aligned} \quad (1)$$

where  $c_1, c_2$  are respectively the mean intensities within and outside the segmentation boundary  $\Gamma$ ,  $u(\mathbf{x}) \in [0, 1]$  is the relaxed function to replace the Heaviside in a level set formulation. Here  $\nu(u) = \max\{0, 2|u - \frac{1}{2}| - 1\}$  is an exact penalty term defined in [15], and it is proven in [11] that the inclusion of this term ensures the constraint  $u \in [0, 1]$  holds. We can then define the segmentation result to be the set  $\Sigma = \{\mathbf{x} : u(\mathbf{x}) \geq \gamma^*\}$  for  $\gamma^* \in (0, 1)$ . We typically fix  $\gamma^* = 0.5$ .

**2.2. Local segmentation.** We can make use of a distance constraint in a model to segment a particular region of interest. The particular object (or objects) of interest are indicated by a marker set  $\mathcal{M}$ , which is a set of points that lie in the object, typically prescribed by the user. We briefly review the work in [25], which achieved local segmentation by using a geodesic distance constraint from [24] (similarly to [26] which used Euclidean distance), and proposed reformulated Chan-Vese like fitting terms more suitable for local segmentation.

$$\begin{aligned} F(u) = & \int_{\Omega} g(|\nabla z|) |\nabla u| d\mathbf{x} + \int_{\Omega} (\lambda_1 f_1(z(\mathbf{x})) - \lambda_2 f_2(z(\mathbf{x}))) u d\mathbf{x} \\ & + \theta \int_{\Omega} \mathcal{D}_G(\mathbf{x}) u d\mathbf{x} + \xi \int_{\Omega} \nu(u) d\mathbf{x}, \end{aligned} \quad (2)$$

where  $g(s) = \frac{1}{1+ts^2}$  is an edge indicator function,  $\mathcal{D}_G$  is the geodesic distance constraint,  $f_1(z(\mathbf{x})) = (z(\mathbf{x}) - c_1)^2$  and

$$f_2(z(\mathbf{x})) = \begin{cases} 1 + \frac{z(\mathbf{x}) - c_1}{\gamma_1}, & c_1 - \gamma_1 \leq z(\mathbf{x}) \leq c_1, \\ 1 - \frac{z(\mathbf{x}) - c_1}{\gamma_2}, & c_1 \leq z(\mathbf{x}) \leq c_1 + \gamma_2, \\ 0, & \text{else.} \end{cases}$$

The selection of parameters  $\gamma_1$  and  $\gamma_2$  is detailed in [25].

**2.3. Geodesic distance.** Compared to Euclidean distance used for local segmentation [26], the geodesic distance is found to perform better in [25] because it is edge aware. This indicates that edge enhancement can make a big impact through geodesic distance. Calculating the geodesic distance involves finding the solution to the Eikonal equation. This is a classic problem in wave propagation, leading to the time it takes for a wave to travel through a certain medium. A general formulation of the two-dimensional Eikonal equation (although easily generalised to one dimension, or easily further extended to higher dimensions) is given as follows (note  $q(\mathbf{x}) = 0$  with  $\Omega^* = \mathcal{M}$ )

$$|\nabla \mathcal{D}(\mathbf{x})| = \frac{1}{p(\mathbf{x})}, \mathbf{x} \in \Omega; \quad \mathcal{D}(\mathbf{x}) = q(\mathbf{x}), \mathbf{x} \in \Omega^*. \quad (3)$$

In [24],  $\frac{1}{p(x)}$  is chosen to be  $\frac{1}{p(\mathbf{x})} = \epsilon + \beta_G |\nabla z(\mathbf{x})|^2 + \theta_G \mathcal{D}_E(\mathbf{x})$ , where the parameters are fixed by the authors of [24] to be  $\epsilon = 10^{-3}$ ,  $\beta_G = 1000$  and  $\theta_G = 0.1$ . A fast sweeping method proposed by [32] is used to solve the Eikonal equation. Then, if we denote the solution to (3) with this special  $p$  and  $q$  as  $D_G^0$ , the geodesic distance,  $\mathcal{D}_G$ , is then defined as

$$\mathcal{D}_G(x) = \frac{\mathcal{D}_G^0(\mathbf{x})}{\|\mathcal{D}_G^0(\mathbf{x})\|_{L^\infty}}. \quad (4)$$

The solution to the Eikonal equation can be physically interpreted as the minimum time it takes to get to a point  $\mathbf{x} \in \Omega$  from  $\Omega^* \subset \Omega$ , where  $p(\mathbf{x})$  is the speed of travel. In a classical wave propagation problem,  $p(\mathbf{x})$  would change depending on the medium of the domain  $\Omega$ . For our application, the domain  $\Omega$  contains all the pixels in an image, and so to penalise pixels on the other side of an edge we must assign a speed  $p(x)$  that takes edge information into consideration. In particular, we would like  $\mathcal{D}$  to be larger near edges, so we assign smaller values to  $p(\mathbf{x})$  when  $\mathbf{x}$  is near edges and larger values otherwise. Or, equivalently we let  $\frac{1}{p(\mathbf{x})}$  to be large at pixels containing edges, and smaller otherwise. It is useful to note that if  $p(\mathbf{x}) = 1$  and  $q(\mathbf{x}) = 0$  then the solution of (3) is simply the Euclidean distance.

Intuitively, the geodesic distance can be understood like this: the distance from point A to B is small when there is no edge in between A and B as it takes not much effort to travel from A to B. However, the presence of an edge can be seen as having to climb a steep hill, requiring more effort. Figure 1 shows the differences between the two distances on a synthetic image. One can see that the geodesic distance takes edge information into consideration seriously.

The one flaw of using geodesic distance is that it is heavily reliant on the method of edge detection. If an edge goes undetected, or even just part of an edge is slightly obscure, the geodesic distance will fail to penalise regions outside of the region of interest. Note that the majority of the edge detection methods rely on the intensity gradient which is very sensitive to noise and low contrast. We propose to use a reproducible kernel Hilbert space and approximated Heaviside function (abbreviated as RKHS method) to handle the problem. In the next section, we review RKHS proposed in [13].

**2.4. RKHS and Approximated Heaviside Functions.** In [13], Deng, Guo and Huang use RKHS and approximated Heaviside functions for image super resolution. Here, we adopt them to separate edge features from smooth components. This will then be incorporated into global and selective image segmentation in Chapter 2 and Chapter 3 respectively. It is able to detect weak edges where using the image

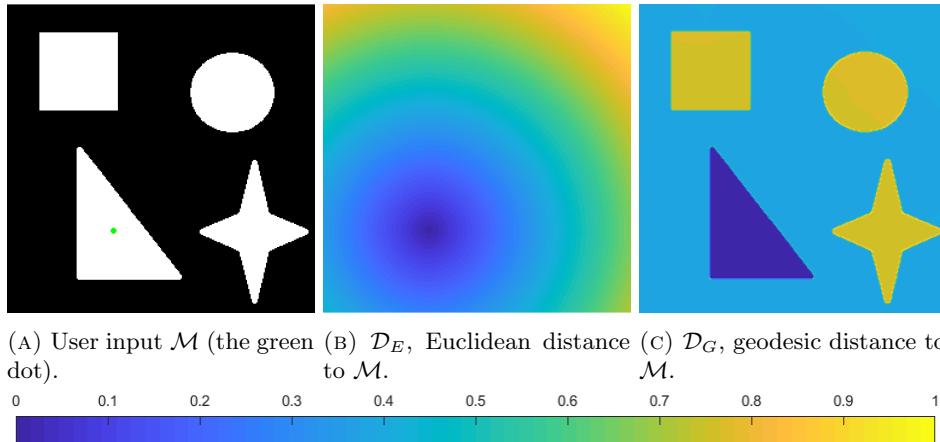


FIGURE 1. Comparison of Geodesic and Euclidean distance constraints.

gradient may fail to detect anything significant. This is typically a problem in images with noise and images in which separate objects are present with little contrast between them. Using RKHS to perform segmentation was used in [17], in which they make use of a limited set of labelled pixels to perform multi-class segmentation on the same image or similar images. The RKHS methods were able to segment medical images of low contrast effectively. RKHS theory was also used in [31], in a segmentation model in order to simplify the computation of solving the model. Alternatively, recently Li et al. [20] proposed a non-local total variation model for restoring weak structures in images, demonstrating applications to inpainting, zooming and denoising. This method does not perform contour detection however, which can be achieved by taking the gradient of the output.

It is shown in [13] that it is reasonable to model an image to be composed of two parts: the smooth parts and the edges. RKHS is used to model the smooth parts of an image, while edges can be represented by a set of approximated Heaviside functions. In this section we will review some key details about RKHS, show the kernel we use and lay out the details of the approximated Heaviside function.

Consider an arbitrary set  $X$ , with a Hilbert space  $\mathcal{H}$  of real functions on  $X$ . Let  $\mathcal{H}$  be the family of functions  $f : X \rightarrow \mathbb{R}$ . Then  $\mathcal{H}$  is an RKHS on  $X$  if there exists a symmetric function  $\mathbb{K} : X \times X \rightarrow \mathbb{R}$  such that: *i*) for all  $x \in X$ ,  $\mathbb{K}(\cdot, x) \in \mathcal{H}$ , *ii*) there exists a reproducing relation  $f(x) = \langle f, \mathbb{K}(\cdot, x) \rangle$  for all  $f \in \mathcal{H}$ . The function  $\mathbb{K}$  is called a Kernel function if, for every distinct  $n$  points  $\{x_1, x_2, \dots, x_n\} \subset X$ ,  $\mathbb{K}(x_i, x_j) \geq 0$ , i.e.  $\mathbb{K}$  is positive semi-definite function. For every RKHS there is a unique reproducing kernel, and conversely for every positive definite kernel  $\mathbb{K} : X \times X \rightarrow \mathbb{R}$ , there is a unique RKHS on  $X$  such that  $\mathbb{K}$  is its reproducing kernel.

Therefore, the choice of kernel determines which space one wishes to work in. Choosing  $\mathbb{K}$  to be polynomials of degree  $d$ :  $\mathbb{K}(\mathbf{x}, \tilde{\mathbf{x}}) = (1 + \langle \mathbf{x}, \tilde{\mathbf{x}} \rangle)^d$  will correspond to a polynomial space. Whereas, if a Sobolev space is sought, we can use  $\mathbb{K}(\mathbf{x}, \tilde{\mathbf{x}}) = \frac{1}{2}e^{-\gamma|\mathbf{x}-\tilde{\mathbf{x}}|}$ . In addition, kernels such as  $\mathbb{K}(\mathbf{x}, \tilde{\mathbf{x}}) \propto |\mathbf{x} - \tilde{\mathbf{x}}|$ ,  $\mathbb{K}(\mathbf{x}, \tilde{\mathbf{x}}) \propto |\mathbf{x} - \tilde{\mathbf{x}}|^3$  correspond to one dimensional piecewise linear and cubic splines respectively.

For our case we use the popular Gaussian kernel corresponding to  $C^\infty$  space to model the smooth parts. It is given by:

$$\mathbb{K}(\mathbf{x}, \tilde{\mathbf{x}}) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^2 e^{-\frac{|\mathbf{x}-\tilde{\mathbf{x}}|^2}{2\sigma^2}}. \quad (5)$$

Without loss of generality, we may assume that the image intensity function is defined on a unit domain  $[0, 1] \times [0, 1]$ . Given an image  $z$  of size  $n \times m$ , we may treat it as a matrix obtained by evaluating the underlying image intensity function at the coordinates  $(j/(n-1), k/(m-1))$ ,  $j = 0, 1, \dots, n-1, k = 0, 1, \dots, m-1$ , denoted by  $\mathbf{x}_i$ . The smooth component of the underlying image intensity function is represented by  $\sum_i d_i K(\mathbf{x}, \mathbf{x}_i)$ .

Additionally, we extend the typical one dimensional approximated Heaviside function,

$$\psi(t) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{t}{\delta}\right), \quad (6)$$

to two dimensions by considering  $\mathbf{x} \in \mathbb{R}^2$  and the variation  $\psi(\mathbf{v}_i \cdot \mathbf{x} + c_j)$ . With  $\mathbf{v}_i = (\cos \theta_i, \sin \theta_i)$ , this two dimensional approximated Heaviside function can describe an edge with orientation  $\theta_i$  at position  $c_j$ . This will allow us to recover edges from an image at different orientations and at different locations. We can then consider the edge part of an image to be modelled from a collection of these functions

$$h(\mathbf{x}) = \sum_{i=1}^{\ell} \sum_{j=1}^N b_{ij} \psi(\mathbf{v}_i \cdot \mathbf{x} + c_j), \quad (7)$$

where  $\theta_i$  are equally partitioned into  $\ell$  segments between  $[0, 2\pi)$ . In [13],  $\ell = 24$ , leading to:

$$\theta_i \in \{0, \pi/12, 2\pi/12, \dots, 23\pi/12\},$$

and  $c_j$ , describing the position of each edge, is taken from  $c_j \in \{0, \frac{1}{N-1}, \frac{2}{N-1}, \dots, 1\}$ , where  $N = nm$  is the number of pixels in the image. Note  $\{\psi(\mathbf{v}_i \cdot \mathbf{x} + c_j)\}$  is a redundant dictionary to represent edges, thus most of the  $b_{ij}$  should be zero. In Figure 2, we show some examples of  $\psi(v_i \cdot \mathbf{x} + c_j)$  from [13].

Combining the RKHS and approximated Heaviside function together, we model the underlying intensity function as

$$f(x) = \sum_i d_i K(x, \mathbf{x}_i) + \sum_{i=1}^{\ell} \sum_{j=1}^N b_{ij} \psi(\mathbf{v}_i \cdot \mathbf{x} + c_j).$$

In what follows, we elaborate how to incorporate the RKHS and approximated Heaviside functions into global (Chapter 3) and selective local image segmentation (Chapter 4). We will present both two stage (feature separation followed by segmentation) and one stage (combined feature separation and segmentation) for each.

We first describe the two-stage approach. Given  $z$ , an image of size  $n \times m$  to be segmented, we assume that it is an estimation of the discretisation of  $f(x)$ , i.e.,  $z = Kd + \Psi\beta + \epsilon_n$ , in which  $K_{j,k} = K(x_j, \mathbf{x}_k)$ ,  $\Psi$  collects values of  $\psi(v_i \cdot \mathbf{x} + c_j)$  evaluated on discrete grids,  $\beta$  is a vector collecting entries  $b_{ij}$  lexicographically, and  $\epsilon_n$  is a random noise.

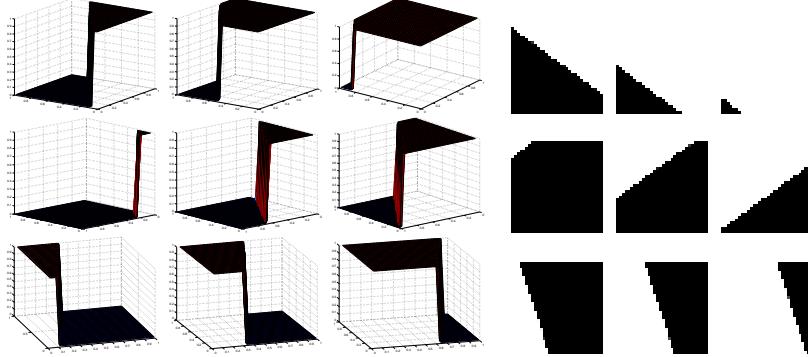


FIGURE 2. Left panel: 3D surface images of  $\psi$  of nine parameter pairs  $(\theta, c)$ ; right panel: the corresponding 2D images. In each panel, from left to right and then from top to bottom:  
 $(\frac{4\pi}{5}, \frac{51}{1024}), (\frac{4\pi}{5}, \frac{25}{64}), (\frac{4\pi}{5}, \frac{175}{256}), (\frac{6\pi}{5}, \frac{135}{1024}), (\frac{6\pi}{5}, \frac{1}{2}), (\frac{6\pi}{5}, \frac{25}{32}), (\frac{8\pi}{5}, \frac{5}{64}), (\frac{8\pi}{5}, \frac{75}{256}), (\frac{8\pi}{5}, \frac{75}{128})$

**3. Proposed global segmentation method.** In the following section, we propose to use RKHS based decomposition to aid global segmentation (selective region of interest segmentation will follow in the next section). We start with a two-stage method, including a reconstruction model prior to segmentation, and then we propose unifying the two stages into a single, combined model.

**3.1. Two stage global segmentation.** We first propose a two stage model to conduct segmentation of images with low contrast and noise. The first stage is to run an RKHS based model (8) to obtain a clean approximation to the given input noisy image to be segmented, and to separate the edge components from the smooth components. In the second stage, we conduct segmentation on the clean image with guidance from the edge information (model (25) as an example). In particular, we aim to use  $\Psi\beta$  in cases where typically the image gradient is used in order to detect edges. The problem with the image gradient is that it can be sensitive to noise and low contrast. Particularly when both noise and regions of low contrast are present, the image gradient struggles to detect the region of low contrast without detecting noise. It is common to use a method of denoising when noise is present, however denoising runs the risk of smoothing out already sensitive edges. The  $\Psi\beta$  term aims to improve on this flaw of the image gradient.

We first use the following novel model to separate edge features and to remove noise, similar to the one proposed in [13]:

$$\min_{d, \beta} \frac{1}{2} \|z - (Kd + \Psi\beta)\|^2 + \gamma d^\top Kd + \alpha \|\beta\|_1 + \nu g^\top |\nabla(Kd + \Psi\beta)|, \quad (8)$$

where  $g = g(\Psi\beta) = \frac{1}{1+\epsilon|\Psi\beta|^2}$  is the edge stopping function based on  $\Psi\beta$  which is better than a gradient based one. The final term encourages the contrast to be low in homogeneous regions and high near edges.

We implement a block coordinate descent (BCD) scheme [28] and perform the computation on small overlapping patches of the image. By doing this, details which might otherwise go unnoticed to the human eye when zoomed out may become enhanced.

**Implementation:** We solve the model (8) by first introducing three auxiliary variables:  $\theta$  such that  $\theta = \beta$ ,  $W$  such that  $W = Kd + \Psi\beta$ , and  $\mathbf{v}$  such that  $\mathbf{v} = \nabla W$ , so that we have the following scheme:

$$\begin{aligned} \min_{d, \beta, \theta, W, \mathbf{v}} & \frac{1}{2} \|z - (Kd + \Psi\beta)\|^2 + \gamma d^\top Kd + \alpha \|\theta\|_1 + \nu g^\top |\mathbf{v}| + \frac{\rho_1}{2} \|\theta - \beta + b_1\|^2 \\ & + \frac{\rho_2}{2} \|W - (Kd + \Psi\beta) + b_2\|^2 + \frac{\rho_3}{2} \|\mathbf{v} - \nabla W + \mathbf{b}_3\|^2. \end{aligned} \quad (9)$$

To implement a BCD scheme, we start with an initial for each of the  $d, \beta, \theta, W, v$  denoted by  $d^{(0)}, \beta^{(0)}, \theta^{(0)}, W^{(0)}, v^{(0)}$  and update them alternatively and iteratively as follows.

The  $d$  problem in proximal form;

$$\begin{aligned} d^{(k)} = \arg \min_d & \frac{1}{2} \|z - (Kd + \Psi\beta^{(k-1)})\|^2 + \gamma d^\top Kd + \frac{\zeta_1}{2} \|d - d^{(k-1)}\|^2 \\ & + \frac{\rho_2}{2} \|W^{(k-1)} - (Kd + \Psi\beta^{(k-1)}) + b_2^{(k-1)}\|^2, \end{aligned} \quad (10)$$

that is solved by

$$d^{(k)} = A^{-1} \left( K^\top z - (1 + \rho_2) K^\top \Psi \beta^{(k-1)} + \zeta_1 d^{(k-1)} + \rho_2 K^\top (W^{(k-1)} + b_2^{(k-1)}) \right), \quad (11)$$

where  $A = (1 + \rho_2)K^\top K + 2\gamma K + \zeta_1 I$ , ( $I$  is identity).

Due to  $g = g(\Psi\beta)$  being non-linear, we linearise the  $\beta$  problem so that we have the following proximal linear form:

$$\begin{aligned} \beta^{(k)} = \arg \min_\beta & \langle \hat{p}^{(k)}, \beta - \hat{\beta}^{(k-1)} \rangle + \frac{\rho_1}{2} \|\theta^{(k-1)} - \beta + b_1^{(k-1)}\|_2^2 \\ & + \frac{\zeta_2}{2} \|\beta - \hat{\beta}^{(k-1)}\|^2, \end{aligned} \quad (12)$$

where  $\hat{\beta}^{(k-1)} = \beta^{(k-1)} + \omega^{(k-1)}(\beta^{(k-1)} - \beta^{(k-2)})$ , and  $\hat{p}^{(k)} = \nabla f(\hat{\beta}^{(k-1)})$ , where  $f$  is given by:

$$\begin{aligned} f(\hat{\beta}^{(k-1)}) = & \frac{1}{2} \|z - (Kd^{(k)} + \Psi\hat{\beta}^{(k-1)})\|^2 + \mu g^\top |\mathbf{v}^{(k-1)}| \\ & + \frac{\rho_2}{2} \|W^{(k-1)} - (Kd^{(k)} + \Psi\hat{\beta}^{(k-1)}) + b_2^{(k-1)}\|^2, \end{aligned} \quad (13)$$

$$\beta^{(k)} = \frac{1}{(\rho_1 + \zeta_2)} (\rho_1(\theta^{(k-1)} + b_1^{(k-1)}) + \zeta_2 \hat{\beta}^{(k-1)} - \hat{p}^{(k)}). \quad (14)$$

The  $\theta, W$  and  $\mathbf{v}$  problems are given as follows:

$$\theta^{(k)} = \arg \min_\theta \alpha \|\theta\|_1 + \frac{\rho_1}{2} \|\theta - \beta^{(k)} + b_1^{(k-1)}\|_2^2, \quad (15)$$

$$\begin{aligned} W^{(k)} = \arg \min_W & \frac{\rho_2}{2} \|W - (Kd^{(k)} + \Psi\beta^{(k)}) + b_2^{(k-1)}\|^2 \\ & + \frac{\rho_3}{2} \|\mathbf{v}^{(k-1)} - \nabla W + \mathbf{b}_3^{(k-1)}\|^2, \end{aligned} \quad (16)$$

$$\mathbf{v}^{(k)} = \arg \min_{\mathbf{v}} \nu g^T |\mathbf{v}| + \frac{\rho_3}{2} \|\mathbf{v} - \nabla W^{(k)} + \mathbf{b}_3^{(k-1)}\|^2. \quad (17)$$

Solutions are given by

$$\theta^{(k)} = \text{shrink}(\beta^{(k)} - b_1^{(k-1)}, \frac{\alpha}{\rho_1}), \quad (18)$$

$$W^{(k)} = \Re \left[ \mathcal{F}^* \left( \frac{\rho_3 \mathcal{F}(\nabla^*(\mathbf{v}^{(k-1)} + \mathbf{b}_3^{(k-1)})) + \rho_2 \mathcal{F}(Kd^{(k)} + \Psi\beta^{(k)} - b_2^{(k)})}{\rho_2 + \rho_3 \mathcal{F}(\nabla^2)} \right) \right], \quad (19)$$

$$\mathbf{v}^{(k)} = \text{shrink}(\nabla W^{(k)} - \mathbf{b}_3^{(k-1)}, \frac{\nu}{\rho_3} \cdot g). \quad (20)$$

Bregman parameters are updated as follows:

$$b_1^{(k)} = b_1^{(k-1)} + \theta^{(k)} - \beta^{(k)}, \quad (21)$$

$$b_2^{(k)} = b_2^{(k-1)} + W^{(k)} - (Kd^{(k)} + \Psi\beta^{(k)}), \quad (22)$$

$$\mathbf{b}_3^{(k)} = \mathbf{b}_3^{(k-1)} + \mathbf{v}^{(k)} - \nabla W^{(k)}. \quad (23)$$

The convergence for this model follows similarly to the discussion later in theorem 3.1 in the next subsection, guaranteed by using a BCD scheme [28] to solve.

Figure 3 shows a CT image in which we are interested in segmenting the blood vessel (pointed by the red arrow), its smooth parts described by  $Kd$  (top middle), and the edge parts described by  $\Psi\beta$  (top right). In the region indicated by the red arrow, we can see clear edges shown in  $\Psi\beta$  in (c) even though it is hard to visualise in (a). This is of key interest as we aim to use this in a segmentation model to improve reliability. In particular, for the region indicated by the red arrow in (a), we see that the edge indicator function making use of  $|\nabla z|$  has failed to detect the very small edge at this point.

Now that we have the decomposition of an input image, we can explore how we can use this information in a segmentation framework. To demonstrate the effectiveness of  $\Psi\beta$  to aid segmentation of objects of low contrast, we use the segmentation framework proposed in [8], which unifies the geodesic active contours model [9] (GAC), and the convex relaxed Chan-Vese model from [11], given by:

$$\begin{aligned} F(u) = & \int_{\Omega} g(|\nabla z|) |\nabla u| d\mathbf{x} + \lambda_1 \int_{\Omega} (z - c_1)^2 u d\mathbf{x} \\ & + \lambda_2 \int_{\Omega} (z - c_2)^2 (1 - u) d\mathbf{x} + \xi \int_{\Omega} \nu(u) d\mathbf{x}, \end{aligned} \quad (24)$$

The first stage model (8) provides not only a separation of edge features from the rest, but also a clean image  $M = Kd + \Psi\beta$ . In the next stage, we use  $M$  as the input and  $g(\Psi\beta)$  as a spatially adaptive weight for segmentation:

$$\begin{aligned} F(u) = & \int_{\Omega} g(|\Psi\beta|) |\nabla u| d\mathbf{x} + \lambda_1 \int_{\Omega} (M - c_1)^2 u d\mathbf{x} \\ & + \lambda_2 \int_{\Omega} (M - c_2)^2 (1 - u) d\mathbf{x} + \xi \int_{\Omega} \nu(u) d\mathbf{x}. \end{aligned} \quad (25)$$

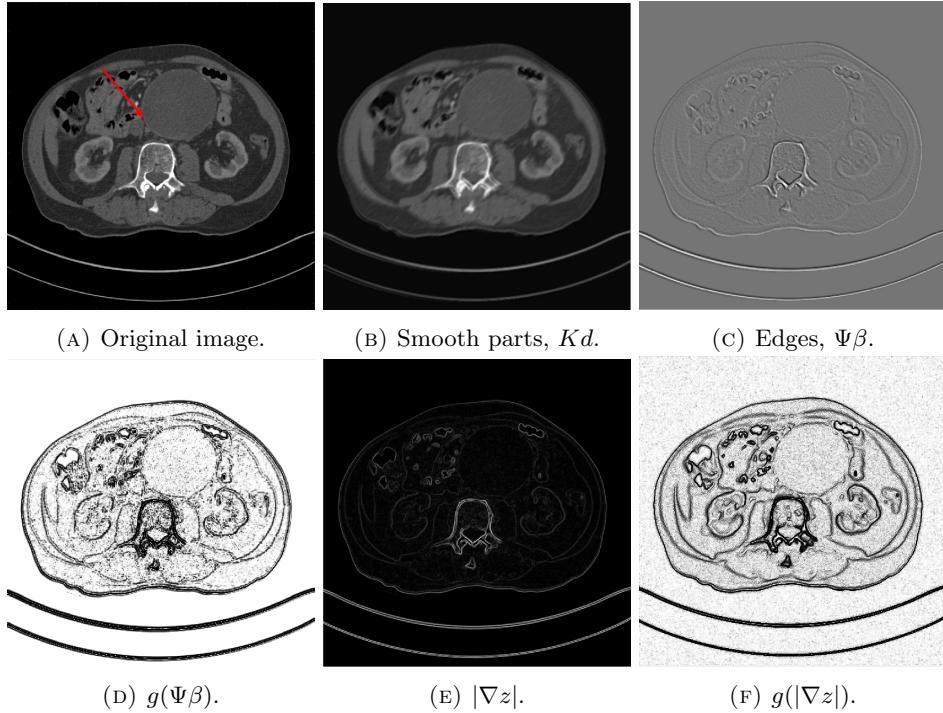


FIGURE 3. The decomposition of the input image (top left), into the smooth parts (top middle) and edge parts (top right), obtained from RKHS model (8), and edge indicator function  $g(\Psi\beta)$  (bottom left). In comparison, we display the gradient of the image,  $|\nabla z|$  and the edge indicator function  $g(|\nabla z|)$  in the bottom middle and bottom right respectively.

**3.2. A combined global segmentation model.** So far we have discussed a two-phase model, which runs the RKHS model (8) as a pre-process to obtain  $Kd$  and  $\Psi\beta$ , then implementing them into our segmentation model of choice. We have chosen to use the framework given in (24), however our  $\Psi\beta$  can be implemented into any segmentation model which uses information about the edges by simply replacing  $|\nabla z|$  with  $\Psi\beta$ .

We now consider unifying the RKHS model with a segmentation model to obtain a simultaneous output. To achieve this, we merge our RKHS model and our chosen segmentation framework and propose the following unified model:

$$\begin{aligned} \min_{d, \beta, u \in [0, 1], c_1, c_2} & \frac{1}{2} \|z - (Kd + \Psi\beta)\|^2 + \gamma d^\top Kd + \alpha \|\beta\|_1 + \mu g^\top |\nabla u| \\ & + \lambda \left[ u^\top (Kd + \Psi\beta - c_1)^2 + (1-u)^\top (Kd + \Psi\beta - c_2)^2 \right], \end{aligned} \quad (26)$$

where  $u$  and  $g = g(|\Psi\beta|)$  are now discretised  $N \times 1$  vectors where  $N$  is the total number of pixels in the given discrete to be segmented image  $z$ .

Due to the non-differentiability of the two  $\ell_1$  terms, we introduce two auxiliary variables,  $\theta$  such that  $\theta = \beta$ , and  $\mathbf{w} = (w_1, w_2)$  such that  $\mathbf{w} = \nabla u$ . Our model then takes the form

$$\begin{aligned} & \min_{d, \beta, \theta, \mathbf{w}, u \in [0, 1], c_1, c_2} \frac{1}{2} \|z - (Kd + \Psi\beta)\|^2 + \gamma d^\top Kd + \alpha \|\beta\|_1 + \mu g^\top \mathbf{w} \\ & + \lambda \left[ u^\top (Kd + \Psi\beta - c_1)^2 + (1-u)^\top (Kd + \Psi\beta - c_2)^2 \right], \\ & \text{such that } \theta = \beta \text{ and } \mathbf{w} = \nabla u. \end{aligned} \quad (27)$$

(27) is approximated using the following augmented form:

$$\begin{aligned} & \min_{d, \beta, \theta, \mathbf{w}, u \in [0, 1], c_1, c_2} \frac{1}{2} \|z - (Kd + \Psi\beta)\|^2 + \gamma d^\top Kd + \alpha \|\beta\|_1 + \mu g^\top \mathbf{w} \\ & + \lambda \left[ u^\top (Kd + \Psi\beta - c_1)^2 + (1-u)^\top (Kd + \Psi\beta - c_2)^2 \right] \\ & + \frac{\rho_1}{2} \|\theta - \beta + b_1\|_2^2 + \frac{\rho_2}{2} \|\mathbf{w} - \nabla u + \mathbf{b}_2\|_2^2. \end{aligned} \quad (28)$$

This gives us 7 sub-problems to solve. In order to guarantee the convergence of our model, we implement a BCD scheme [28] that has convergence. In particular, we change the  $d, c_1, c_2$  and  $u$  sub-problems to the proximal form, and we change the  $\beta$  sub-problem to the prox-linear form.

In the following, we explicitly discuss the solution to each of the sub-problems.

#### *d sub-problem:*

$$\begin{aligned} d^{(k)} = & \arg \min_d \frac{1}{2} \|z - (Kd + \Psi\beta^{(k-1)})\|^2 + \gamma(d)^\top Kd + \frac{\zeta_1}{2} \|d - d^{(k-1)}\|^2 \\ & + \lambda \left[ (u^{(k-1)})^\top (Kd + \Psi\beta^{(k-1)} - c_1^{(k-1)})^2 \right. \\ & \left. + (1-u^{(k-1)})^\top (Kd + \Psi\beta^{(k-1)} - c_2^{(k-1)})^2 \right]. \end{aligned} \quad (29)$$

The solution is given by:

$$\begin{aligned} d^{(k)} = & A^{-1} \left( K^\top z - (1+2\lambda)K^\top \Psi\beta^{(k-1)} + \zeta_1 d^{(k-1)} \right. \\ & \left. + 2\lambda K^\top [c_1^{(k-1)} u^{(k-1)} + c_2^{(k-1)} (1-u^{(k-1)})] \right), \end{aligned} \quad (30)$$

where  $A = (1+2\lambda)K^\top K + 2\gamma K + \zeta_1 I$ .

#### *$\beta$ sub-problem:*

$$\begin{aligned} \beta^{(k)} = & \arg \min_\beta \langle \hat{p}^{(k)}, \beta - \hat{\beta}^{(k-1)} \rangle + \frac{\rho_1}{2} \|\theta^{(k-1)} - \beta + b_1^{(k-1)}\|_2^2 \\ & + \frac{\zeta_2}{2} \|\beta - \hat{\beta}^{(k-1)}\|^2, \end{aligned} \quad (31)$$

where  $\hat{\beta}^{(k-1)} = \beta^{(k-1)} + \omega^{(k-1)}(\beta^{(k-1)} - \beta^{(k-2)})$ , and  $\hat{p}^{(k)} = \nabla f(\hat{\beta}^{(k-1)})$ , where  $f$  is given by:

$$\begin{aligned} f(\hat{\beta}^{(k-1)}) &= \frac{1}{2} \|z - (Kd^{(k)} + \Psi\hat{\beta}^{(k-1)})\|^2 + \mu g^\top |\mathbf{w}^{(k-1)}| \\ &\quad + \lambda \left[ (u^{(k-1)})^\top (Kd^{(k)} + \Psi\hat{\beta}^{(k-1)} - c_1^{(k-1)})^2 \right. \\ &\quad \left. + (1 - u^{(k-1)})^\top (Kd^{(k)} + \Psi\hat{\beta}^{(k-1)} - c_2^{(k-1)})^2 \right]. \end{aligned} \quad (32)$$

$$\begin{aligned} \nabla f(\hat{\beta}^{(k-1)}) &= -\Psi^\top (z - (Kd^{(k)} + \Psi\hat{\beta}^{(k-1)})) \\ &\quad - 2\mu\Psi^\top (|\mathbf{w}^{(k-1)}| \odot (g(\Psi\hat{\beta}^{(k-1)}))^2 \odot (\Psi\hat{\beta}^{(k-1)})) \\ &\quad + 2\lambda\Psi^\top [Kd^{(k)} + \Psi\hat{\beta}^{(k-1)} - c_1^{(k-1)}u^{(k-1)} - c_2^{(k-1)}(1 - u^{(k-1)})], \end{aligned} \quad (33)$$

where  $\odot$  denotes the Hadamard product between vectors (component wise multiplication). Thus the  $\beta$  update is given as follows:

$$\beta^{(k)} = \frac{1}{(\rho_1 + L_2)} (\rho_1(\theta^{(k-1)} + b_1^{(k-1)}) + \zeta_2\hat{\beta}^{(k-1)} - \hat{p}^{(k)}). \quad (34)$$

#### $\theta$ sub-problem:

$$\theta^{(k)} = \arg \min_{\theta} \alpha \|\theta\|_1 + \frac{\rho_1}{2} \|\theta - \beta^{(k)} + b_1^{(k-1)}\|_2^2, \quad (35)$$

whose solution to this is given by the shrinkage operator

$$\theta^{(k)} = \text{shrink}(\beta^{(k)} - b_1^{(k-1)}, \frac{\alpha}{\rho_1}), \quad (36)$$

and we update the Bregman parameter as follows:

$$b_1^{(k)} = b_1^{(k-1)} + \theta^{(k)} - \beta^{(k)}. \quad (37)$$

#### $c_1$ and $c_2$ sub-problems:

$$c_1^{(k)} = \arg \min_{c_1} \lambda (u^{(k-1)})^\top (Kd^{(k)} + \Psi\beta^{(k)} - c_1)^2 + \frac{\zeta_3}{2} \|c_1 - c_1^{(k-1)}\|^2, \quad (38)$$

$$c_2^{(k)} = \arg \min_{c_2} \lambda (1 - u^{(k-1)})^\top (Kd^{(k)} + \Psi\beta^{(k)} - c_2)^2 + \frac{\zeta_4}{2} \|c_2 - c_2^{(k-1)}\|^2, \quad (39)$$

and the solutions are given by the following:

$$c_1^{(k)} = \frac{\zeta_3 c_1^{(k-1)} + 2\lambda(u^{(k-1)})^\top (Kd^{(k)} + \Psi\beta^{(k)})}{\zeta_3 + 2\lambda(u^{(k-1)})^\top I}, \quad (40)$$

$$c_2^{(k)} = \frac{\zeta_4 c_2^{(k-1)} + 2\lambda(1 - u^{(k)})^\top (Kd^{(k)} + \Psi\beta^{(k)})}{\zeta_4 + 2\lambda(1 - u^{(k)})^\top I}. \quad (41)$$

#### $u$ sub-problem:

$$\begin{aligned} u^{(k)} = \arg \min_{u \in [0,1]} &+ \frac{\rho_2}{2} \|\mathbf{w}^{(k-1)} - \nabla u + \mathbf{b}_2^{(k-1)}\|_2^2 + \frac{\zeta_5}{2} \|u - u^{(k-1)}\|^2 \\ &+ \lambda \left[ (u)^\top (Kd^{(k)} + \Psi\beta^{(k)} - c_1^{(k)})^2 \right. \\ &\quad \left. + (1 - u)^\top (Kd^{(k)} + \Psi\beta^{(k)} - c_2^{(k)})^2 \right]. \end{aligned} \quad (42)$$

The solution to this is given by:

$$u^{(k)} = \Re \left[ \mathcal{F}^* \left( \frac{\rho_2 \mathcal{F}(\nabla^*(\mathbf{w}^{(k-1)} + \mathbf{b}_2^{(k-1)})) - \lambda \mathcal{F}(r^{(k)}) + \zeta_5 \mathcal{F}(u^{(k-1)})}{\zeta_5 + \rho_2 \mathcal{F}(\nabla^2)} \right) \right], \quad (43)$$

where  $r^{(k)} = (Kd^{(k)} + \Psi\beta^{(k)} - c_1^{(k)})^2 - (Kd^{(k)} + \Psi\beta^{(k)} - c_2^{(k)})^2$ , and  $\mathcal{F}$  is the fast Fourier transform operator and  $\mathcal{F}^*$  is its inverse.

**w sub-problem:**

$$\mathbf{w}^{(k)} = \arg \min_{\mathbf{w}} \mu g^T |\mathbf{w}| + \frac{\rho_2}{2} \|\mathbf{w} - \nabla u^{(k)} + \mathbf{b}_2^{(k-1)}\|_2^2. \quad (44)$$

We solve it using the same shrinkage operator from the  $\theta$  sub-problem, given by:

$$\mathbf{w}^{(k)} = \text{shrink}(\nabla u^{(k)} - \mathbf{b}_2^{(k-1)}, \frac{\mu}{\rho_2} \cdot g), \quad (45)$$

with the Bregman parameter updated as follows:

$$\mathbf{b}_2^{(k)} = \mathbf{b}_2^{(k-1)} + \mathbf{w}^{(k)} - \nabla u^{(k)}. \quad (46)$$

We summarise the combined global segmentation algorithm in Algorithm 1.

---

**Algorithm 1** Combined global segmentation algorithm.

---

- 1: **Input:** Image  $z$ , parameters  $\sigma, \gamma, \alpha, \lambda, \rho_1, \rho_2, \zeta_1, \zeta_2, \zeta_3, \zeta_4$  and  $\zeta_5$ .
  - 2: Initialise  $u$ , and  $d, \beta, \theta, \mathbf{w}$  on each patch arbitrarily.
  - 3: **while**  $\text{relativeerror} > \text{tolerence}$  **do**
  - 4: For each patch, update  $d, \beta, \theta$  and  $b_1$  as in (30), (34), (36) and (37) respectively. Stitch patches together to obtain  $Kd$  and  $\Psi\beta$  by averaging intensities from overlapping patches.
  - 5: Calculate  $g(\Psi\beta) = \frac{1}{1+\epsilon|\Psi\beta|^2}$ .
  - 6: Update  $c_1$  and  $c_2$  as in (40) and (41).
  - 7: Update  $u$  as in (43).
  - 8: Update  $\mathbf{w}$  and  $\mathbf{b}_2$  as in (45) and (46) respectively.
  - 9: **end while**
  - 10: **Output:** Reproduced image  $Kd + \Psi\beta$ . Segmentation result  $\Sigma = \{\mathbf{x} : u(\mathbf{x}) \geq \gamma^*\}$  where we usually fix  $\gamma^* = 0.5$ .
- 

**Theorem 3.1.** For any initialisation  $\beta^{(0)}, \theta^{(0)}, b_1^{(0)}, c_1^{(0)}, c_2^{(0)}, u^{(0)}, \mathbf{w}^{(0)}, \mathbf{b}_2^{(0)}$ , the sequence obtained by Algorithm 1 converges to a set of stationary points  $(d^*, \beta^*, \theta^*, b_1^*, c_1^*, c_2^*, u^*, \mathbf{w}^*, \mathbf{b}_2^*)$ .

The proof is contained in the work of [28], which recommends three different updates for each sub-problem: original, proximal and prox-linear. Each update requires certain conditions to hold in the corresponding functional to be minimised.

The  $\theta$  and  $\mathbf{w}$  sub-problems are straightforward as the associated functionals given in (35) and (44) are strongly convex, thus no changes to the problem are needed and we can use the original (first) update from [28].

For the  $d, c_1, c_2$  and  $u$  problems, (29), (38), (39), (42), we have used the proximal update, which involves adding the  $l_2$  term with coefficient  $\zeta_i$ . The only condition required is that the  $\zeta_i$  parameters are required to be uniformly lower bounded from

zero and uniformly upper bounded. In practice we typically fix these parameters to be less than or equal to  $10^{-7}$ .

We implement the prox-linear update for the  $\beta$  problem (31) which involves adding a proximal term as in the proximal update, as well as linearising the existing terms. The necessary requirement for the prox-linear update is that (33) is Lipschitz-continuous. We prove this by definition:

$$\begin{aligned}
|\nabla f(\alpha) - \nabla f(\beta)| &= |\Psi^\top \Psi(1 + 2\lambda)(\alpha - \beta) \\
&\quad - 2\mu \Psi^\top (|\mathbf{w}| \odot ((g(\Psi\alpha)^2 - g(\Psi\beta)^2) \odot (\Psi\alpha - \Psi\beta)))| \\
&\leq (1 + 2\lambda) |\Psi^\top \Psi(\alpha - \beta)| \\
&\quad + 2\mu |\Psi^\top (|\mathbf{w}| \odot ((g(\Psi\alpha)^2 - g(\Psi\beta)^2) \odot (\Psi\alpha - \Psi\beta)))| \\
&\leq (1 + 2\lambda) |\Psi^\top \Psi(\alpha - \beta)| + 2\mu |\Psi^\top (|\mathbf{w}| \odot (\Psi\alpha - \Psi\beta))| \\
&= (1 + 2\lambda) |\Psi^\top \Psi(\alpha - \beta)| + 2\mu |\Psi^\top \mathbf{W} \Psi(\alpha - \beta)| \\
&= ((1 + 2\lambda) |\Psi^\top \Psi| + 2\mu |\Psi^\top \mathbf{W} \Psi|) |\alpha - \beta|,
\end{aligned}$$

where  $\mathbf{W} = \text{diag}(|\mathbf{w}|)$  and we have used the fact that  $0 \leq g(s) \leq 1$ .

**3.3. Numerical experiments.** In this section we demonstrate the performance of our proposed global segmentation models (**two-stage global segmentation** given by solving (8) followed by (25), and **combined one stage global segmentation** given by Algorithm 1). We compare with two models relying on the gradient, given by the model (24), where the first simply relies on the gradient  $|\nabla z|$  in  $g$ , and the second model is the same as (24), except that  $g = g_C = 1 - \mathcal{C}$ , where  $\mathcal{C}$  is the binary output of Canny edge detection (so that  $g = 0$  on edges and  $g = 1$  elsewhere). In addition, we make a comparison with the two-stage method by Cai et al. [34].

Figure 4 shows performance on two different images, where we can see that the first two columns show the relevant models, corresponding to (24) using just gradient  $\nabla z$  and (24) using Canny respectively, struggling to separate the white matter from the grey matter. The third column corresponds to the method by Cai et al., which is a noticeable improvement over the first two. In comparison, the fourth and fifth columns, corresponding to our proposed two-stage method and combined one-stage method respectively, do an improved job at capturing finer details of the white matter which are smoothed out in the first stage of the Cai method. We note that the combined global segmentation algorithm is slightly cleaner and sharper than the two-stage approach, and so is our preferred method. We are able to detect fine edge details without distracted too much by noise with the combined method.

In Figure 5 we look at the same image but with varying levels of noise, ranging between 3% to 9%. It is clear to see the combined model in the final two columns doing a good job at all noise levels at separating the white matter from grey matter. In comparison to the first two columns ((24) depending only on the gradient  $|\nabla z|$ ) which struggles to segment effectively particularly at higher noise levels. We do note that the reconstruction part of the combined model provides a slight denoising effect, which helps with the result.

In conclusion while the two-stage method is easier to implement, and easier to tune parameters, the combined one-stage global segmentation method (Algorithm

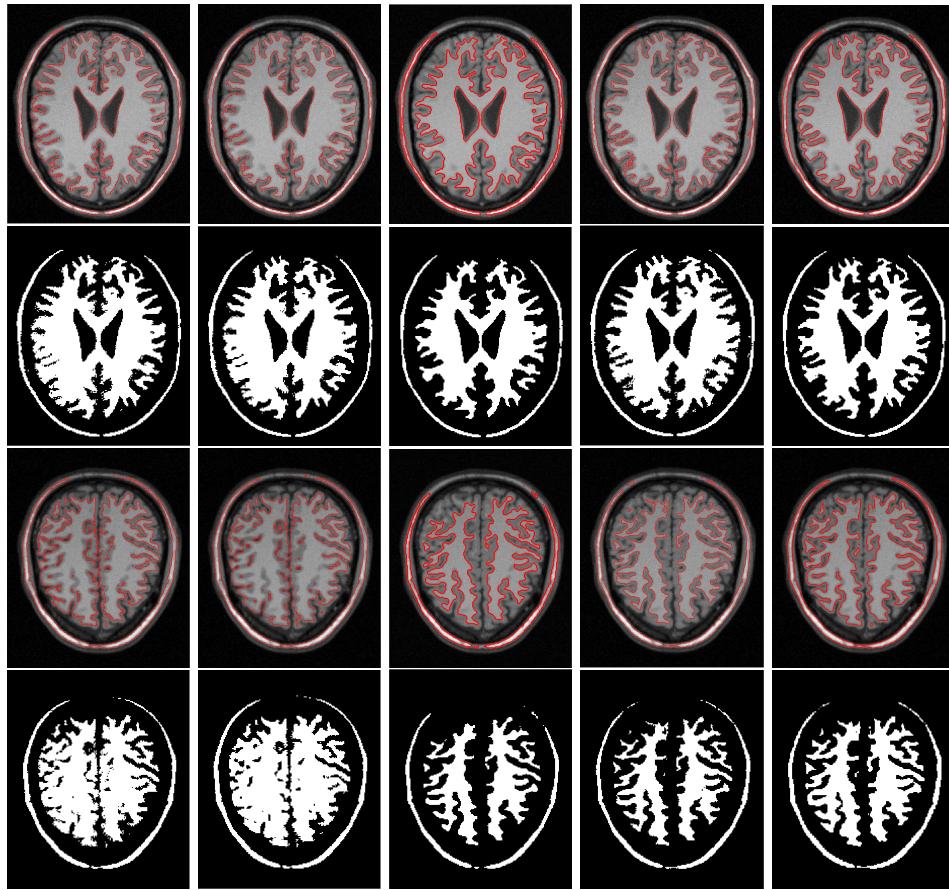


FIGURE 4. Segmentation of white matter from MRI brain image. We show both the contour overlaid on the original image, and the associated binary segmentation result for each example. First column: segmentation using edge weighted Chan-Vese using  $|\nabla z|$  model (24). Second column: Edge weighted Chan-Vese model (24) using Canny ( $g = g_c$ ). Third column: Cai et al. [34]. Fourth column: Proposed two stage global model (8) and (25). Final column: The proposed combined global segmentation model (26).

1) is preferred. The combined method is overall quicker and results are better, due to its ability to use the decomposition method directly to inform the segmentation.

**4. Proposed selective segmentation method.** In certain applications, one is only interested in segmenting some local regions of interest. We may do this in two manners: two-stage and one-stage. The former is to run RKHS and Heaviside feature separation and then use the edge information to help with selective segmentation. The second is to fuse both feature separation and segmentation in one model. We will address the advantages and disadvantages of each.

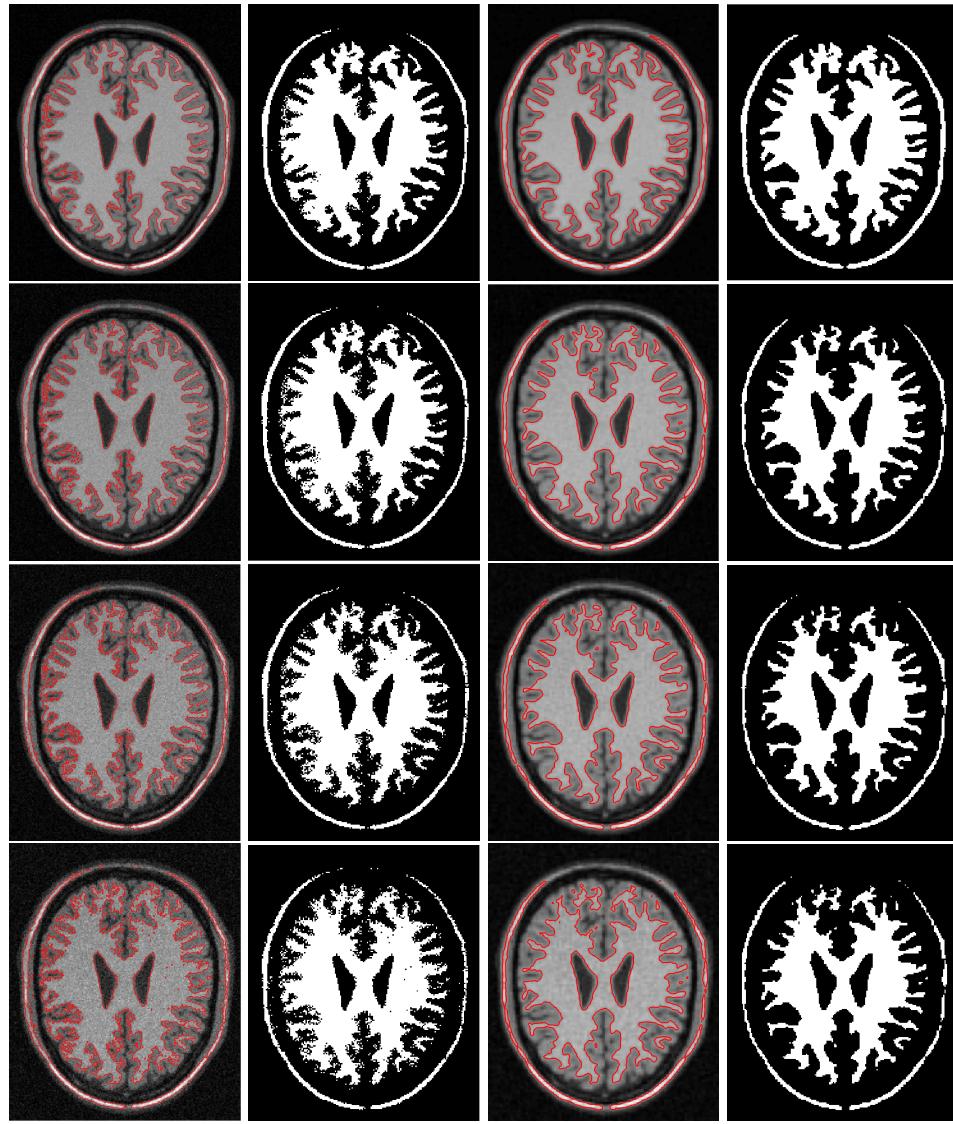


FIGURE 5. Row 1: 3% noise, Row 2: 5% noise, Row 3: 7% noise, Row 4: 9% noise. We show both the contour overlaid on the original image, and the associated binary segmentation result for each example. First two columns: Using model (24). Final two columns: Combined model, algorithm (1).

**4.1. Two stage selective segmentation.** We demonstrate the idea of selective segmentation using the framework of model from Roberts and Spencer (2), but it could be applied in any segmentation framework that relies on image gradient.

In model (2), there are two places that use gradient: one is in the edge stopping function, and one is in the geodesic distance function calculation (3). We will replace the gradient by  $\Psi\beta$ . Specifically, we replace  $g(|\nabla z|)$  with  $g(|\Psi\beta|)$  and redefine the

geodesic distance to be based on the solution of

$$|\nabla \mathcal{D}_G^0(\mathbf{x})| = \epsilon + \beta_G |\Psi\beta| + \theta_G \mathcal{D}_E(\mathbf{x}). \quad (47)$$

The geodesic distance is calculated by normalising as in (4).

In Figure 6, we see the geodesic distance penalty using the image gradient fails to penalise the region indicated by the red arrow, where there is low contrast and not much of an edge to be detected. The geodesic distance using  $\Psi\beta$  improves on this, showing a higher penalty to the bottom left region.

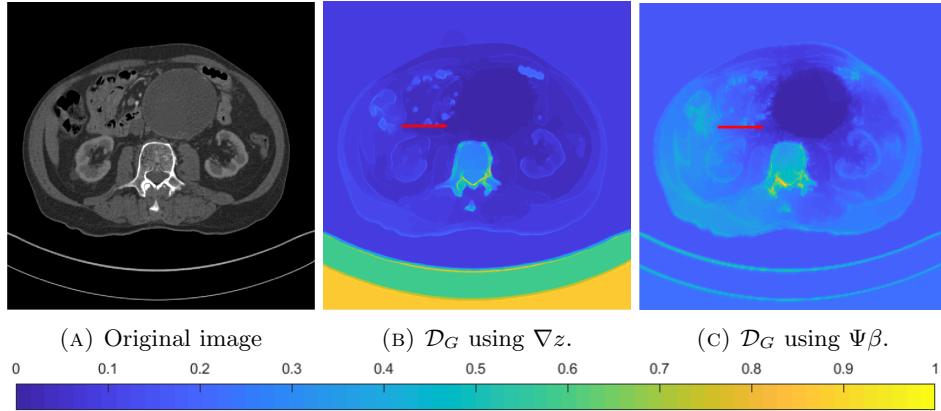


FIGURE 6. Geodesic distances computed using gradient (middle) and  $\Psi\beta$  (right)

Similarly to the global two-stage case, we run the first stage model (8) to obtain the clean image  $M = Kd + \Psi\beta$  and the decomposition, so that we can use  $\Psi\beta$  as a weight for segmentation, in addition to making use of it in the geodesic distance:

$$\begin{aligned} F(u) &= \int_{\Omega} g(|\Psi\beta|) |\nabla u| d\mathbf{x} + \int_{\Omega} (\lambda_1 f_1(M) - \lambda_2 f_2(M)) u d\mathbf{x} \\ &\quad + \theta \int_{\Omega} \mathcal{D}_G(\mathbf{x}) u d\mathbf{x} + \xi \int_{\Omega} \nu(u) d\mathbf{x}, \end{aligned} \quad (48)$$

where  $\mathcal{D}_G(\mathbf{x})$  is computed using (47) and (4) and  $g(\Psi\beta) = \frac{1}{1+\epsilon|\Psi\beta|^2}$ .

In [25], Roberts and Spencer find the associated Euler-Lagrange equations to solve their model, using a modified AOS algorithm first introduced in [26] to solve the resulting time dependent PDE. Instead of taking that approach, we use the dual formulation method of Chambolle [10], first applied to segmentation problems in the convex relaxed setting in [8].

Firstly, we fix  $c_1$  to be the average intensity of pixels in the region indicated by our input,  $\mathcal{M}$ . In order to minimise with respect to  $u$ , we introduce an auxiliary variable  $w$  so that our problem looks as follows:

$$\begin{aligned} \min_{u,w} & \int_{\Omega} g(|\Psi\beta|) |\nabla u| d\mathbf{x} + \frac{1}{2\rho} \int_{\Omega} (u - w)^2 d\mathbf{x} \\ & + \lambda \int_{\Omega} (f_1(M) - f_2(M)) w d\mathbf{x} + \theta \int_{\Omega} \mathcal{D}_G(\mathbf{x}) w d\mathbf{x}, \end{aligned} \quad (49)$$

where we have set  $\lambda_1 = \lambda_2 = \lambda$ . Additionally, it is proven in [8] that the penalty term  $\nu(u)$  in this framework is not needed as the constraint  $u \in [0, 1]$  (as well as  $w \in [0, 1]$ ) holds provided that the initialisation  $u^{(0)} \in [0, 1]$ , and so we can omit it from our functional. We solve by alternating between minimising  $u$  and  $w$ . Our first sub-problem is the  $u$  subproblem, given by:

$$u^{(k)} = \arg \min_{u^{(k)}} \int_{\Omega} g(|\Psi\beta|) |\nabla u^{(k)}| d\mathbf{x} + \frac{1}{2\rho} \int_{\Omega} (u^{(k)} - w^{(k-1)})^2 d\mathbf{x}. \quad (50)$$

The solution is given by

$$u^{(k)} = w^{(k-1)} - \rho \operatorname{div} p^{(k)}, \quad (51)$$

where  $p^{(k)} = (p_1^{(k)}, p_2^{(k)})$ , which is solved iteratively with  $p^{(0)} = 0$  and

$$p^{(k+1)} = \frac{p^{(k)} + \delta t \nabla(\operatorname{div}(p^{(k)}) - \frac{w^{(k)}}{\rho})}{1 + \frac{\delta t}{g} |\nabla(\operatorname{div}(p^{(k)}) - \frac{w^{(k)}}{\rho})|}. \quad (52)$$

After updating for  $u^{(k)}$ , we solve the following problem for  $w^{(k)}$ :

$$w^{(k)} = \arg \min_{w^{(k)}} \frac{1}{2\rho} \int_{\Omega} (u^{(k)} - w^{(k)})^2 d\mathbf{x} + \int_{\Omega} r w^{(k)} d\mathbf{x}, \quad (53)$$

where  $r = \lambda(f_1(M) - f_2(M)) + \theta \mathcal{D}_G$ .

The solution is given by

$$w^{(k)} = \min\{\max\{u^{(k)} - \rho r, 0\}, 1\}. \quad (54)$$

---

**Algorithm 2** (Two Stage Selective Segmentation)

---

- 1: **Input:** Image  $z$ , parameters  $\gamma, \alpha, \nu, \sigma$  for RKHS and parameters  $\mu, \lambda, \theta, \rho, \delta t$  and  $\iota$  for segmentation. A set of marker points  $\mathcal{M}$  that are inside the region of interest. Fix  $c_1$  to be the average intensity of pixels inside the region indicated by  $\mathcal{M}$ .
  - 2: Solve RKHS model (8) to obtain coefficients  $d$  and  $\beta$  as well as a clean image  $M = Kd + \Psi\beta$  that is the input for segmentation.
  - 3: Initialise  $u$  such that  $u = 1$  inside  $\mathcal{M}$ , and  $u = 0$  outside  $\mathcal{M}$ .
  - 4: Calculate  $g(\Psi\beta) = \frac{1}{1+\iota|\Psi\beta|^2}$  and compute  $\mathcal{D}_G$  according to (47) (4).
  - 5: **while**  $\text{relativeerror} > \text{tolerance}$  **do**
  - 6:   Update  $u$  and  $w$  as in (51) and (54) respectively
  - 7: **end while**
  - 8: **Output:** Segmentation result  $\Sigma = \{\mathbf{x} : u(\mathbf{x}) \geq \gamma^*\}$  where we usually fix  $\gamma^* = 0.5$ .
- 

**4.2. A combined selective segmentation model.** In section 2.2 we have discussed combining reconstruction and segmentation models into a unified framework, allowing us to obtain both the decomposition of an image and the global segmentation result simultaneously. It is a natural extension to consider a unified framework to achieve selective segmentation, which we could do for example on the basis of model (8) (48). This would give us the following model:

$$\begin{aligned} \min_{d,\beta,u \in [0,1]} & \frac{1}{2} \|z - (Kd + \Psi\beta)\|^2 + \gamma d^\top Kd + \alpha \|\beta\|_1 + \mu g(\Psi\beta)^\top |\nabla u| \\ & + u^\top [\lambda_1 f_1(Kd + \Psi\beta) - \lambda_2 f_2(Kd + \Psi\beta)] + \theta u^\top \mathcal{D}_G, \end{aligned} \quad (55)$$

where  $\mathcal{D}_G$  is the geodesic distance defined using  $\Psi\beta$ . This is highly non-linear however, and very computationally expensive to implement as stated above. This is mainly due to the inclusion of the geodesic distance term  $\mathcal{D}_G$ . The geodesic distance involves solving an Eikonal equation, and to do that on each iteration drastically slows down the computation speed. In addition, the inclusion of this term makes the  $\beta$  problem highly non-linear.

To simplify this, we can fix the  $\beta$  used in the calculation of  $\mathcal{D}_G$  to be the  $\beta$  from the previous iterate, and implement a BCD scheme [28], similar to how we implemented the global combined model in section 2.2. Implementing in this way achieves convergence.

**4.3. Numerical experiments.** In this section we only show some results of the proposed selective segmentation methods (two stage approach Algorithm 2, and unified approach (55)), demonstrating how  $\Psi\beta$  can be used as an improvement over the traditional  $|\nabla z|$  method. Our proposed global segmentation models are tested in §3.3. The unified one stage approach (55) is slow and difficult to compute and the results are not as good as those of the two stage approach Algorithm 2, so we focus on Algorithm 2 first.

To begin with, we look only at our two stage approach, contrasting the model (2), which we will denote as method 1 (**M1**). For simplicity, we denote the proposed Algorithm 2 as method 2 (**M2**). The differences between **M1** and **M2** is **M1** uses  $|\nabla z|$  for edge detection, whereas in **M2**, we run (8) as a preprocess, and use the  $\Psi\beta$  output (and reconstructed image) into the segmentation framework.

First, in Figure 7, we present a synthetic image of two neighbouring ellipses of very slightly different intensities. To test selective segmentation, we set the initial marker as shown in Figure 7b with the aim to selectively segment the ellipse on the left but not both of them as there is an edge between them, and we start with an initial inside the left ellipse. We add increasing levels of noise to demonstrate that the  $\Psi\beta$  is still capable of capturing the edges in the presence of excessive noise. Using the image gradient in the case of noisy images is a case of choosing between smoothing (using a preprocess such as total generalised variation (TGV)) to eliminate noise in order to capture the edges, but not smoothing too much without blurring the edge. Even at 20% noise in this synthetic example, the image gradient is not well equipped to detect the small edge between the two objects without detecting a lot of noise, where as  $\Psi\beta$  performs well even at 40% noise. The leaking in gradient based edge detection leads to segmentation of two ellipses instead of one.

Then we show some examples on medical images. Segmentation of an abdominal aortic aneurysm is a difficult task as there is often a lack of contrast between the boundary of the blood vessel and neighbouring objects. Often, the gradient of the image fails to significantly detect an edge in these regions, however using the proposed RKHS based Algorithm 2 on small patches, these small edges can become more prominent, allowing us to segment the blood vessel more effectively. Figures 8 and 9 show the comparison for two different CT images. One can see that the

gradient based approach **M1** overshoots at the lower left in Figure 8 and at the lower left and right in Figure 9.

In addition to visual results, we have supplied heatmaps in figures 8 and 9 to visualise the robustness of parameters. As a quantitative measure to assess the accuracy of our segmentation results, we use the Tanimoto coefficient (or Jaccard coefficient [16]) defined as

$$TC(\Sigma, GT) = \frac{|\Sigma \cap GT|}{|\Sigma \cup GT|},$$

where  $GT$  is the ground truth segmentation, and  $\Sigma$  is the binary segmentation output from Algorithm 2. A value  $TC = 1$  corresponds to perfect accuracy to the ground truth, and as  $TC$  decreases so does the quality of the segmentation result. In our heatmaps, we represent as green to be  $TC = 1$ , and red to be  $TC = 0$ .

In the heatmaps, we vary  $\lambda \in [0, 20]$  (on y axis) and  $\theta \in [0, 800]$  (on x axis), and the associated colour in the chart is the measure of how accurate that parameter pair  $(\lambda, \theta)$  segments the blood vessel. It is clear to see that there is a greater amount of green ( $TC \approx 1$ ) on our proposed **M2** heatmaps, compared to **M1** which uses the gradient.

It is clear to see that the two-stage approach (Algorithm 2) is an improvement on the existing method using the gradient for edge guidance. We now discuss the performance of the unified selective model, showing a result in Figure 10. This image is the same image as Figure 8, except that it is cropped. We crop it for two reasons: the first was to increase the computation speed of the model, as recalculating the geodesic distance on every iteration is computationally expensive. The second reason is that it was difficult to isolate the object from the background. The parameter selection is very sensitive, and achieving a result on the full image was difficult - cropping simplified this task, and we can see in Figure 10 that we have successfully segmented the blood vessel, without the segmentation result leaking on the object in the bottom left.

In conclusion, unlike for global segmentation in which we recommend the unified approach (§3.3), for selective segmentation we recommend the two-stage approach. This allows for a quicker computation time and less sensitive parameter selection. The combined approach for selective segmentation takes a long time to process due to recomputing the geodesic distance at each iteration, and in addition the parameters are sensitive.

**Parameter selection:** There are some parameters in the definition of the Gaussian kernel ( $\sigma$ ), the approximated Heaviside function ( $\delta$ ), the edge stopping function ( $\iota$ ) and in (8), (47) and (49). Some of the parameters are not so sensitive to the natures of the underlying image and are pre-fixed while others need to be tuned. For parameter that need to be tuned, we usually test on orders of different magnitude and choose the best. We typically fix  $\sigma = 12$ ,  $\delta = 10^{-4}$ ,  $\iota = 1000$ , fix the patchsize to be  $4 \times 4$  with the overlap of 3. In model (8), we can fix  $\alpha = 1$ ,  $\nu$  is varied between  $10^{-4}$  and  $10^{-2}$ , and  $\gamma$  is tuned by ranging between  $10^{-11}$  and  $10^{-5}$ , choosing the best for different test images. Similar choices for these parameters can be made in the respective combined models (global combined and selective combined.) The parameter in (47) are chosen as:  $\epsilon = 10^{-3}$ ,  $\beta_G = 1000$  and  $\theta_G = 0.1$ . For the model (49), we can fix  $\rho = 1$ , whereas  $\lambda$  and  $\theta$  are tuned empirically depending on the test image (and similarly for global combined and selective combined.) In (52), we introduce a time-step, which we fix as  $\delta t = 1/8$ .

**5. Conclusion.** In this paper we have proposed RKHS methods to improve segmentation results, particularly in images in which low contrast is present. We began by introducing a two stage process, involving first an image reconstruction model (decomposition process) using RKHS and approximated Heaviside functions, and then using the output from the first model into a segmentation framework. In addition to the two stage approach, we have proposed unifying to the two stages into a single, combined model, achieving reconstruction and segmentation simultaneously. We compared the effectiveness of both variants (two-stage and unified) against segmentation frameworks using the traditional  $|\nabla z|$  as edge guidance.

In addition to global segmentation, we then turned our focus to looking at region of interest selective segmentation, again showing the effectiveness of  $\Psi\beta$  over the traditional gradient approach in segmenting objects with low contrast at the boundary. We proposed a similar two-stage approach and unified combined model for selective segmentation, although we note that the unified approach in the selective case is less preferred than the two-stage, unlike global segmentation, due to the increased computation speed of the geodesic distance.

**Acknowledgments.** L. Burrows is grateful to the UK EPSRC, the Smith Institute for Industrial Mathematics, and the Royal Liverpool and Broadgreen University Hospitals NHS Trust (RLBUHT) for supporting the work through an Industrial CASE award. W. Guo is partially supported by USA National Science Foundation (DMS-1521582). K. Chen is thankful for the UK EPSRC grant EP/N014499/1. F. Torella thanks the RLBUHT for their approval to participate this work and for providing partial funding. The authors would like to thank professor Wotao Yin from UCLA for the discussion on the convergence of the algorithms.

## REFERENCES

- [1] L. Ambrosio and V. Tortorelli. On the approximation of free discontinuity problems. *BOLLETTINO DELL'UNIONE MATEMATICA ITALIANA. B*, 6(1):105–123, 1992.
- [2] L. Ambrosio and V. M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via  $t$ -convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990.
- [3] B. Appleton and H. Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):106–118, 2006.
- [4] N. Badshah and K. Chen. Image selective segmentation under geometrical constraints using an active contour approach. *Communications in Computational Physics*, 7(4):759, 2010.
- [5] E. Bae, J. Yuan, and X.-C. Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, 92(1):112–129, 2011.
- [6] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):1124–1137, 2004.
- [7] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 1, pages 377–384. IEEE, 1999.
- [8] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and vision*, 28(2):151–167, 2007.
- [9] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.
- [10] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

- [11] T. F. Chan, S. Esedoglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.
- [12] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- [13] L.-J. Deng, W. Guo, and T.-Z. Huang. Single-image super-resolution via an iterative reproducing kernel hilbert space method. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(11):2001–2014, 2016.
- [14] C. Gout, C. Le Guyader, and L. Vese. Segmentation under geometrical conditions using geodesic active contours and interpolation using level set methods. *Numerical Algorithms*, 39(1-3):155–173, 2005.
- [15] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer Science & Business Media, 2013.
- [16] P. Jaccard. The distribution of flora in the alpine zone.1. New Phytologist, 11(2): 37-50, 1912.
- [17] S. H. Kang, B. Shafei, and G. Steidl. Supervised and transductive multi-class segmentation using p-Laplacians and RKHS methods *Journal of Visual Communication and Image Representation*, 25(5):1136–1148, 2014.
- [18] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [19] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding. Minimization of region-scalable fitting energy for image segmentation. *IEEE Transactions on Image Processing*, 17(10):1940–1949, 2008.
- [20] Z. Li, and F. Malgouyres. Regularized non-local total variation and application in image restoration. *Journal of Mathematical Imaging and Vision*, 59(2):296–317, 2017.
- [21] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.
- [22] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.
- [23] R. B. Potts. Some generalized order-disorder transformations. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 48, pages 106–109. Cambridge University Press, 1952.
- [24] M. Roberts, K. Chen, and K. L. Irion. A convex geodesic selective model for image segmentation. *Journal of Mathematical Imaging and Vision*, pages 1–22, 2018.
- [25] M. Roberts and J. Spencer. Chan–vese reformulation for selective image segmentation. *Journal of Mathematical Imaging and Vision*, Aug 2019.
- [26] J. Spencer and K. Chen. A convex and selective variational model for image segmentation. *Communications in Mathematical Sciences*, 13(6):1453–1472, 2015.
- [27] X.-F. Wang, D.-S. Huang, and H. Xu. An efficient local chan–vese model for image segmentation. *Pattern Recognition*, 43(3):603–618, 2010.
- [28] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.
- [29] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov. A continuous max-flow approach to potts model. In *European Conference on Computer Vision*, pages 379–392. Springer, 2010.
- [30] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov. A spatially continuous max-flow and min-cut framework for binary labeling problems. *Numerische Mathematik*, 126(3):559–587, 2014.
- [31] H. Zhang, Y. Chen, and J. Shi. Nonparametric image segmentation using Renyi’s statistical dependence measure. *Journal of Mathematical Imaging and Vision*, 44(3):330–340, 2012.
- [32] H. Zhao. A fast sweeping method for Eikonal equations. *Mathematics of Computation*, 74(250):603–627, 2005.
- [33] , C. Liu and M. Ng and T. Zeng. Weighted variational model for selective image segmentation with application to medical images *Pattern Recognition*, 76: 367-379,2018.
- [34] , X. Cai, R. Chan, and T. Zeng. A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding. *SIAM Journal on Imaging Sciences*, 6(1), 368-390, 2013.

Received xxxx 20xx; revised xxxx 20xx.

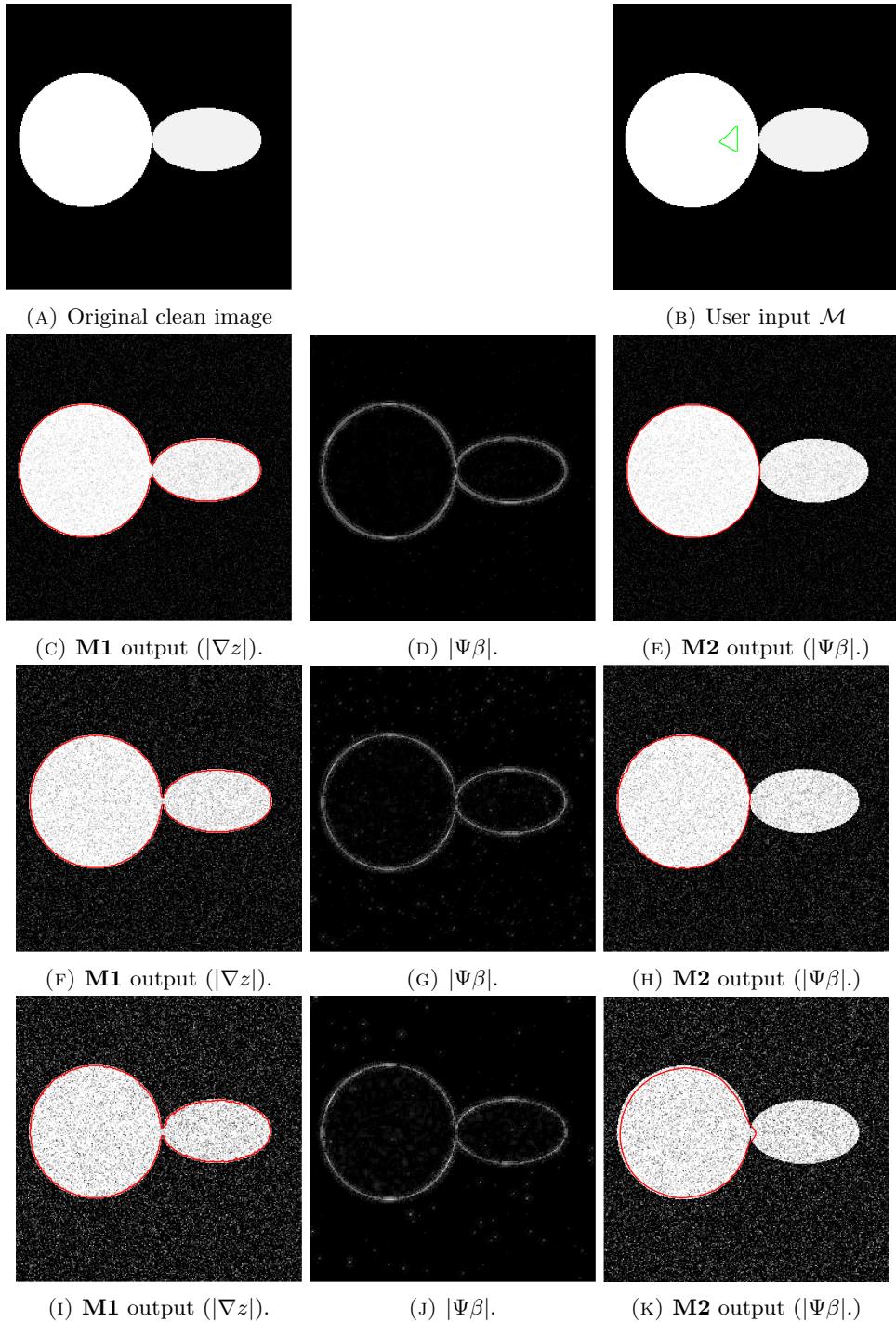


FIGURE 7. Row one: Original clean image and the user input. Row two: Image and result with 20% noise. Row three: Image and result with 40% noise. Row four: Image and result with 60% noise.

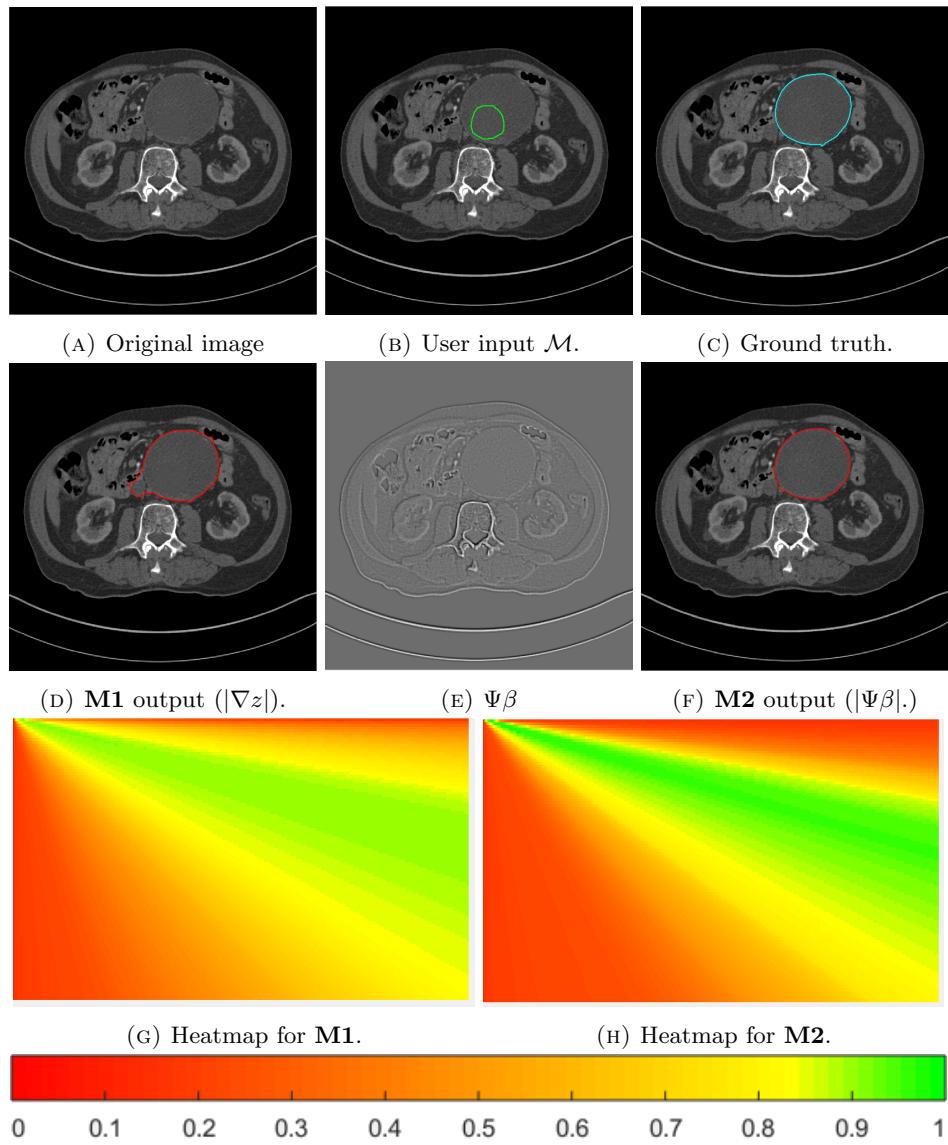


FIGURE 8. Segmentation of blood vessel.

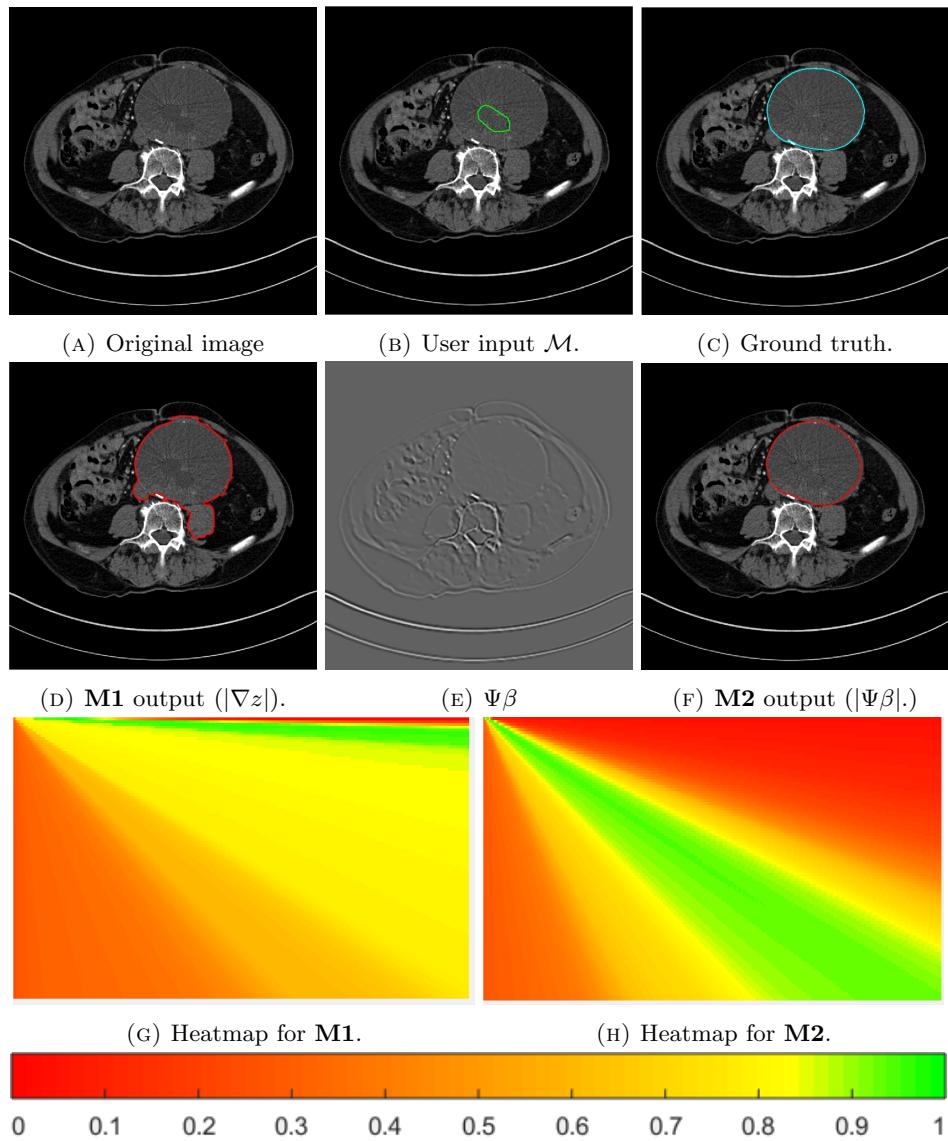


FIGURE 9. Segmentation of blood vessel.

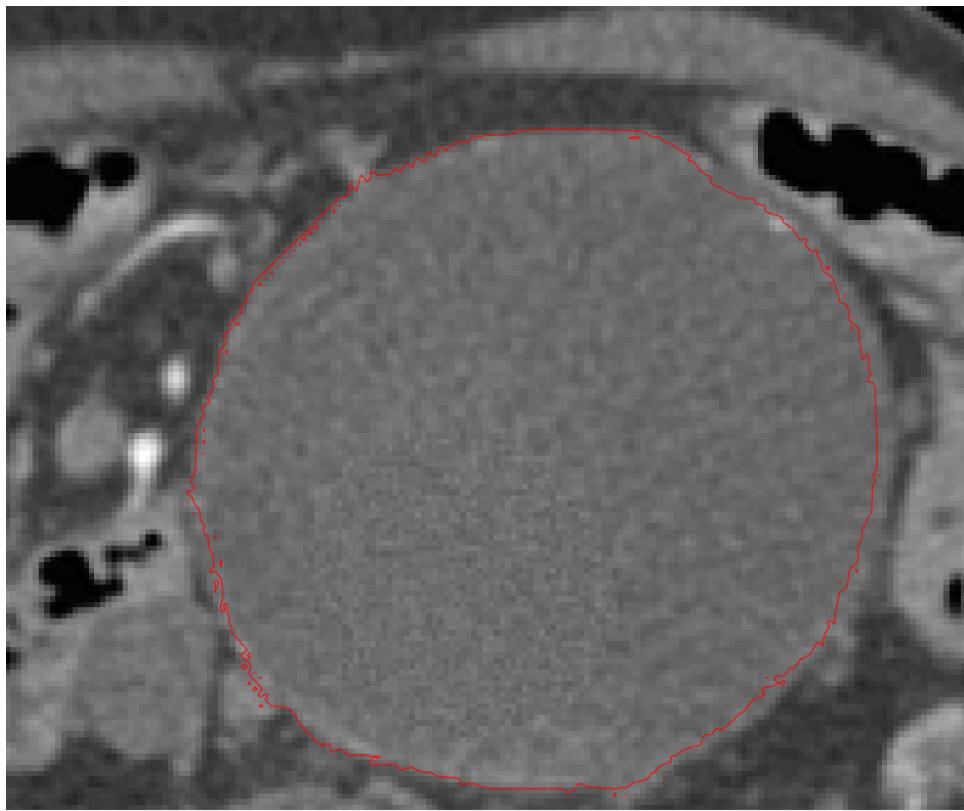


FIGURE 10. Result of the combined selective model.