To my advisor and my parents
ACKNOWLEDGMENTS

First of all, I would like to express my sincere appreciation to my advisor, Professor Yunmei Chen. I am very grateful for her professional guidance, invaluable advice and consistent support throughout my life at UF. Without these, this thesis would not have been possible. She is one of the greatest professors I have ever met, very knowledgeable and friendly. I have learned a lot from her and I really enjoy the experience of working with her.

Next, I want to thank my other excellent committee members, Professor Murali Rao, Professor Scott McCullough, Professor Lei Zhang and Professor Keith White. Thank them for being my committee members, attending my oral qualifying exam and the final defense, providing constructive advice and suggestions to my research. In addition, Dr. Chen, Dr. Rao, Dr. McCullough and Dr. Zhang have prepared recommendation letters for my job hunting, and I really appreciate their kindness and help.

I would also like to thank my parents, my brother, my sister-in-law and my little twin nephews in China for their continuous love and unconditional support throughout my life. They are the most important people in my life and they have given me much joy and encouragement.

Finally, I thank all my group members, Fuhua Chen, Xiaojing Ye, Ouyang Yuyuan, Jiangli Shi, for their excellent comments and discussions about my research. I would also like to express my gratefulness to all the staff members in our department, Margaret Somers, Sandy Gagnon, Kristen Cason, and Conie Doby for their patience and valuable assistance to my graduate studies at UF.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>8</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 Introduction</td>
<td>11</td>
</tr>
<tr>
<td>1.1 Image Segmentation</td>
<td>11</td>
</tr>
<tr>
<td>1.1.1 Methods Overview</td>
<td>11</td>
</tr>
<tr>
<td>1.1.2 Why Nonparametric Image Segmentation?</td>
<td>13</td>
</tr>
<tr>
<td>1.1.3 Segmentation with Bias Correction</td>
<td>15</td>
</tr>
<tr>
<td>1.2 Image Restoration</td>
<td>18</td>
</tr>
<tr>
<td>2 Nonparametric Image Segmentation Using Rényi’s Statistical Dependence Measure</td>
<td>21</td>
</tr>
<tr>
<td>2.1 Problem Statement and Related Works</td>
<td>21</td>
</tr>
<tr>
<td>2.2 Rényi’s Statistical Measure</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Proposed Model and Numerical Method</td>
<td>24</td>
</tr>
<tr>
<td>2.3.1 Level set formulation and numerical method</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2 A soft formulation and numerical method</td>
<td>28</td>
</tr>
<tr>
<td>2.4 Experimental Results</td>
<td>30</td>
</tr>
<tr>
<td>2.4.1 Comparison with two parametric models</td>
<td>32</td>
</tr>
<tr>
<td>2.4.2 Test on real medical images</td>
<td>34</td>
</tr>
<tr>
<td>2.5 Conclusion</td>
<td>35</td>
</tr>
<tr>
<td>3 An Efficient Algorithm for Multiphase Image Segmentation with Intensity Bias Correction</td>
<td>38</td>
</tr>
<tr>
<td>3.1 Model Formulation</td>
<td>38</td>
</tr>
<tr>
<td>3.1.1 Multiphase Segmentation and MAP Approach</td>
<td>38</td>
</tr>
<tr>
<td>3.1.2 Modeling the Intensity Inhomogeneity</td>
<td>39</td>
</tr>
<tr>
<td>3.2 Numerical Algorithm</td>
<td>42</td>
</tr>
<tr>
<td>3.2.1 First Variations of $b$, $c$, and $\sigma$</td>
<td>43</td>
</tr>
<tr>
<td>3.2.2 Solution to $u$</td>
<td>44</td>
</tr>
<tr>
<td>3.2.3 Algorithm</td>
<td>46</td>
</tr>
<tr>
<td>3.3 Experimental Results</td>
<td>46</td>
</tr>
<tr>
<td>3.3.1 Experiment Settings</td>
<td>46</td>
</tr>
<tr>
<td>3.3.2 Quantitative Evaluation and Comparison with Existing Methods</td>
<td>48</td>
</tr>
<tr>
<td>3.3.2.1 Methods Overview</td>
<td>48</td>
</tr>
<tr>
<td>Table Number</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2-1</td>
<td>Number of iterations and CPU time (s) for the experiments (Figure 2-1, 2-2, and 2-3).</td>
</tr>
<tr>
<td>3-1</td>
<td>Quantitative evaluation of Figure 3-2</td>
</tr>
<tr>
<td>3-2</td>
<td>Quantitative evaluation of Figure 3-3</td>
</tr>
<tr>
<td>3-3</td>
<td>Quantitative evaluation of Figure 3-4</td>
</tr>
<tr>
<td>3-4</td>
<td>Quantitative evaluation of Figure 3-5</td>
</tr>
<tr>
<td>3-5</td>
<td>Quantitative evaluation of Figure 3-6</td>
</tr>
<tr>
<td>4-1</td>
<td>PSNR values for Figure 4-4, 4-5.</td>
</tr>
<tr>
<td>4-2</td>
<td>PSNR values for Figure 4-6.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1-1</td>
<td>Image segmentation with intensity bias correction.</td>
</tr>
<tr>
<td>2-1</td>
<td>Segmentation results of three synthetic images.</td>
</tr>
<tr>
<td>2-2</td>
<td>Segmentation results of a clean brain image (first row) and its noisy version (second row).</td>
</tr>
<tr>
<td>2-3</td>
<td>Segmentation results of a clean lung image (first row) and a noisy one (second row).</td>
</tr>
<tr>
<td>3-1</td>
<td>Comparison of the proposed model with WKVLS, SVMLS and CLIC on an MR brain image.</td>
</tr>
<tr>
<td>3-2</td>
<td>Comparison of the proposed model, WKVLS, SVMLS and CLIC on an MR brain image with severe noise and intensity inhomogeneity.</td>
</tr>
<tr>
<td>3-3</td>
<td>Robust to different initializations.</td>
</tr>
<tr>
<td>3-4</td>
<td>Robustness test of the proposed algorithm on different intensity inhomogeneity.</td>
</tr>
<tr>
<td>3-5</td>
<td>Robustness test of the proposed algorithm on different image noise levels.</td>
</tr>
<tr>
<td>3-6</td>
<td>Efficiency of local intensity estimation in FastSEG when applied to images with strong noise.</td>
</tr>
<tr>
<td>4-1</td>
<td>True logarithmic histogram of $n$ and its corresponding approximation.</td>
</tr>
<tr>
<td>4-2</td>
<td>Samples of training images.</td>
</tr>
<tr>
<td>4-3</td>
<td>Part of the clean test images and their corresponding blurry ones.</td>
</tr>
<tr>
<td>4-4</td>
<td>Test on salt-and-pepper noise.</td>
</tr>
<tr>
<td>4-5</td>
<td>Test on random-valued impulse noise.</td>
</tr>
<tr>
<td>4-6</td>
<td>Test on different level of noise.</td>
</tr>
</tbody>
</table>
This thesis is devoted to the formulation, implementation, and improvement of mathematical models and efficient algorithms for image segmentation and image restoration using tools based on partial differential equation (PDE), calculus of variations, numerical optimizations, and statistical methods.

Image segmentation or pattern classification is to partition an image domain into disjointed subregions such that each region corresponds to a single subject. This problem is of fundamental importance in digital image processing and has wide applications. During the last few decades, a considerable amount of approaches have emerged to tackle this issue. However, the difficulties caused by intensity inhomogeneity, higher level of noise, and unevenly distributed illumination still need to be addressed.

We have two main results for image segmentation.

The first one is a nonparametric image segmentation model, which partitions an image by maximizing the similarity between the image and a label image, generated by setting different constants as the image intensities of partitioned subregions. Rényi’s statistical dependence measure, is selected as a criterion to measure the similarity, and the computation is further simplified by applying the theory of reproducing kernel Hilbert spaces.

Another one is a fast algorithm for multiphase image segmentation with bias correction, which is designed to segment images corrupted by strong noise and intensity
inhomogeneity. In our algorithm, all the variables have closed-form solutions. Moreover, the algorithm only involves regular convolutions and point-wise projections onto the unit ball and canonical simplex. Therefore, the computation complexity is very low.

Our second focus is image restoration. Due to limitations of hardware or erroneous transmission, real world images may contain certain kind of distortion and noise. The goal of image restoration is to recover the latent clear image from observed contaminated data. The fields of image restoration are very broad, and it is impossible to cover all the related fields. In this thesis, we only deal with joint image deblurring and denoising. This is in general an ill-posed problem, and certain regularization technique must be imposed. In this work, we choose sparse representation theory, and propose a sparseland model for deblurring images in the presence of impulse noise.
CHAPTER 1
INTRODUCTION

Chapter 1 contains the background information regarding image segmentation and image restoration.

1.1 Image Segmentation

Image segmentation or pattern classification is to partition the image domain into disjointed subregions such that each region corresponds to a single subject. Mathematically speaking, given an image \( I : \Omega \rightarrow \mathbb{R} \), where \( \Omega \subset \mathbb{R}^D \) is a closed and bounded region that represents the domain of \( I \), the purpose of image segmentation is to partition its domain \( \Omega \) into several (say \( M \)) regions, such that each region delineates an image pattern distinct from those by other regions. Namely, we need to solve for a set of regions \( \{ \Omega_i \}_{i=1}^M \) such that \( \Omega = \bigcup_{i=1}^M \Omega_i \), and \( \{ \Omega_i \}_{i=1}^M \) are disjoint, where \( \Omega_i \) indicates the support of the \( i \)-th pattern in image \( I \). In the above notation, \( D \) is the dimension of the image (usually 2 or 3). For ease of presentation, we only consider rectangular gray-valued images.

Image segmentation has long been a challenging and fundamental problem in image processing and computer vision with a wide range of applications. In particular, the emerging developments in medical imaging demand more effective and robust algorithms for image segmentation. Approaches to image segmentation can be roughly classified into two categories: edge-based models (e.g. \([1–5]\)) and region-based models (e.g. \([6–11]\)). Edge-based models rely on edge information to locate the boundaries of regions. Region based models partition the image domain into several disjoint regions such that each region exhibits distinct statistical properties from those by others.

1.1.1 Methods Overview

The problem of image segmentation can be formulated as an energy minimization problem. Most of the known models in this field are closely related to the Mumford-Shah
model [10] proposed by Mumford and Shah. Mumford-Shah model can be written as

$$\min E_{MS}(J, \Gamma) = \lambda \int_{\Omega} (J(x) - I(x))^2 \, dx + \mu \int_{\Omega \setminus \Gamma} |\nabla J(x)|^2 \, dx + |\Gamma|, \quad (1-1)$$

where $I$ is the original image to be segmented, $J$ is a smooth approximation of $I$, $|\Gamma|$ represents the length the segmenting curve $\Gamma$.

The first term in (1–1) ensures the closeness of the original image $I$ and its approximation $J$, the second term guarantees the smoothness of the $J$, and the last term minimizes the edge length. $\lambda, \mu > 0$ are parameters used to balance the strength of these three terms. Mumford-Shah model is a well-known model for simultaneous smoothing and segmentation, however, the well-posedness has not been established yet and it is not commonly used in most real-world segmentation applications. Most real-world segmentation problems seek for objects with close boundaries, while model (1–1) may result in open boundaries.

In real-world segmentation problems, a variant of Mumford-Shah model is more applicable. This variant is the so-called the piecewise constant Mumford-Shah model or the Chan-Vese Model, which is proposed in [6, 9]. Instead of seeking for smooth approximations, Chan-Vese model looks for a piecewise constant approximation. More specifically, this model assumes $J$ is a constant $c_i$ in each $\Omega_i$, and in this case, model (1–1) becomes

$$\min E_{CV}(\Omega_i, c_i) = \sum_{i=1}^{M} \lambda \int_{\Omega_i} (I(x) - c_i)^2 \, dx + |\partial \Omega_i|, \quad (1-2)$$

where $M$ refers to the number of partitioned subregions. Note that the second term in (1–1) disappears because $J$ is constant on $\Omega_i$.

The Chan-Vese model (1–2) is shown to be well-posed, and it can separate two relatively homogeneous regions without using any edge information. However, the homogeneity assumption limits its applications.

A more general approach is parametric region based active contour method. It assume that at each $x \in \Omega_i$, the image intensity $I(x)$ is an independent random
variable drawn from the probability density function (p.d.f.) \( P(l_i(x)|\lambda_i) \), where \( l_i \) is the restriction of \( l \) to \( \Omega_i \) and \( \lambda_i \) is a parameter vector which needs to be estimated. Then the segmentation is obtained by minimizing the negative log-likelihood functional together with the length term, i.e.,

\[
\min_{C,\lambda_i/s} - \sum_{i}^{M} \int_{\Omega_i} \log P(l_i(x)|\lambda_i) \, dx + \beta |C|. \tag{1–3}
\]

Model (1–3) is more general and has various specific formulations depending how the probability density function, \( P(l_i(x)|\lambda_i) \), is selected. The region competition model by Zhu et al. [11] and geodesic active region models by Rousson et al. [12] and Paragios et al. [13] assume \( P(l_i(x)|\lambda_i) \) is a Gaussian distribution:

\[
P(l_i(x)|c_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp \left( -\frac{(l_i(x) - c_i)^2}{2\sigma_i^2} \right). \tag{1–4}
\]

If all the \( \sigma_i \)'s are the same and prefixed, model (1–3)(1–4) reduces to model (1–2). We want to mention that model (1–4) (1–4) is a global Gaussian model as it assumes all random variables \( l(x) \) in the region \( \Omega_i \) share the same mean \( c_i \) and variance \( \sigma_i \).

1.1.2 Why Nonparametric Image Segmentation?

Note that most of the above mentioned methods are parametric ones in that they explicitly or implicitly assume images share certain kind of special characteristics. Some may assume images are homogeneous, while others may assume image intensities obey specific distribution. However, a specific assumption of the intensity distribution can be a significant restriction in real applications, especially when the image has heavy noise or is of multi-modal intensity distributions.

To overcome this problem, nonparametric models [14] have been developed to increase the robustness and successfully applied to image segmentation and registration. These methods are featured by using nonparametric density estimation to replace the parametric density estimation. For instance, the nonparametric active contour model [15] is driven by the disparity of the foreground and background p.d.f.'s,
which are approximated by Parzen window density method. In [16], the dynamic segmentation of video image sequences is obtained by minimizing the disparity of the p.d.f. of the current frame with the previous one and the p.d.f.'s are also estimated using Parzen window method. The variational segmentation model in [17] incorporates boundary information with region information, where the boundary information is obtained from the edge map image and the interior region information is represented by the intensity p.d.f. captured using Parzen window density estimation. Mory et al. [18] consider the foreground and background p.d.f.'s to be unknown, which are integrated in the region computation model. The proposed model could simultaneously perform segmentation and nonparametric density estimation, which are updated using the Parzen window density method. [19] regards the foreground and background cumulative distribution function (c.d.f.) as unknown and utilizes the Wasserstein distance to measure the disparity of local c.d.f. with the estimated c.d.f.'s. A work closely related to ours is [20], in which Kim et al. segment images through maximizing mutual information between the image to be segmented and its corresponding label image defined by setting different constants as image intensities of partitioned subregions (ref. Chapter 2), which turns out to be minimizing the displacement of the Logarithmic of the foreground and background p.d.f.'s and the p.d.f.'s are again estimated using Parzen window method.

Borrowing the idea from [20], we propose a new approach of nonparametric image segmentation that uses Rényi’s statistical dependence measure, maximum correlation coefficient, as a similarity measure of two images in different modalities. By using this measure as an alternative choice of dependence measure to mutual information, we do not need to estimate the continuous joint probability density function of two images, which is sensitive to image quantization and also makes the optimization process complicated. Moreover, the computation is further simplified by applying the theory of reproducing kernel Hilbert spaces.
1.1.3 Segmentation with Bias Correction

In many real world applications, images may encounter significant intensity inhomogeneity due to spatial variations in illuminations and physical constraints in acquisition sensitivities. A particular example is magnetic resonance (MR) imaging, where the inhomogeneity is presented as bias field mainly caused by nonuniform magnetic fields [21]. In these cases, the same object in a given image may exhibit various contrasts at different locations of the image domain. These large variations in image intensities can cause false identification of regions as a consequence of ambiguous statistics presented by pixel intensities. For example, the widely used Chan-Vese model [9] failed to generate correct segmentation due to intensity inhomogeneity as shown in Figure 1-1(b) whereas the desired segmentation should be the one shown in Figure 1-1(c). Therefore, segmentation for images with inhomogeneous intensities is a challenging problem, and usually requires a combination of segmentation and intensity bias correction.

There have been a series of work proposed to tackle the segmentation problem with bias field correction (e.g. [21–30] and references therein). Due to the space limitation,
we here only review several very recent models that are closely related to the present work. In [29], the authors proposed a variational model for tissue classification of MR images. In their model, local intensities of different tissues within a neighborhood are used to form separable clusters, where the centers are approximated by the product of the bias within the neighborhood and a tissue-dependent constant. This local clustering criterion is combined with membership functions to form an energy functional. Then the tissue classification and bias field estimation are simultaneously achieved by estimating the membership functions, bias field, and the parameters that approximate the true signals in each region via minimizing the energy functional. In [31], a minimization framework was developed for multiphase segmentation and bias correction. This model used the same local clustering criterion as that in [29] to define an energy functional in level set formulation. Later, Li et al. extended their model in [30] to simultaneous multi-phase segmentation and bias correction. In [30], multiple level set functions were used to represent the subregions. Minimization of the energy was achieved by an interleaved process of level set evolution [32] and the estimation of the bias field. Recently, Zhang et al. in [33] also proposed a level set approach for simultaneous tissue segmentation and bias correction for MR images. But different from the work in [29–31], where the intensities in each cluster were approximated by its mean in $L^2$ sense, in their model local intensities of different tissues were assumed to be distributed as Gaussians with the means as the centers of the cluster and variances to be optimized.

We aim to propose a joint image segmentation and bias field correction framework that unifies these recent models, and develop an efficient numerical algorithm to solve the model. We consider a multiplicative structure of intensity density function for center points of image patches, then we utilize the maximum-a-posteriori (MAP) principle to construct a generalized model for joint image segmentation and bias field estimation.

Besides the modeling aspect, computation is also a critical issue of segmentation in real applications. Many variational segmentation models use level set formulation [34]
to deal with the problem of topology change during contour evolution, and have showed promising results. However, the computational cost of level set based approaches can be high in the commonly used semi-implicit implementation. Also, it has been demonstrated that the method is sensitive to the initial condition and may require periodic reinitialization of the level set function [9]. As an alternative, we directly work on the characteristic functions of the partition regions. Then the regularization term that penalizes lengths of partition curves is equal to the total variations (TV) of the characteristic functions, and the new objective function is convex with respect to these functions. However, the minimization is still difficult to carry out due to the non-convexity of the solution set, which is later shown to be the tensor product of vertices of a canonical simplex. To resolve this problem, we relax the constraint to the entire simplex before solving the minimization problem. Then we truncate the final result and obtain a characteristic function as our segmentation result. This is an extension of the idea proposed in [35, 36] for two-phase segmentation. In that case, there is a single characteristic function with binary constraint appeared in the minimization. After relaxation, the constraint becomes the unit interval and hence can be handled relatively easily, see, e.g. [35, 37]. Moreover, theoretical results on the equivalence between the original and relaxed problems can be established. However, the situation becomes much more complicated when the problem levels up to multiphase segmentation, mainly due to the simplex constraints and multiple non-smooth TV terms involved in optimization.

In Chapter 3, we propose an effective numerical algorithm that utilizes the primal-dual formulation of TV norms and special properties of canonical simplex to quickly approximate a solution. The primal-dual formulation has been successfully applied to TV based image reconstruction to achieve very promising efficiency [38, 39]. In this thesis, we utilize the similar idea to derive a fast segmentation algorithm which involves only convolutions using kernel function with small support, and pointwise projections onto the unit ball and canonical simplex. To demonstrate the effectiveness of
our method, we show the numerical experiments on various kinds of images, and make comparisons with the state-of-the-art algorithms.

1.2 Image Restoration

Image restoration is another fundamental problem in image processing and it has been extensively studied in the past several years. Basically, it is to recover the latent clear image from degraded data. The degraded process can be mathematically modeled as

\[ g = k \ast f + n, \]

where \( g, f \) and \( n \) represent the observed noisy and blurred image, the original clear image as well as the additive noise respectively, \( k \) stands for a known blur kernel which is space invariant and the symbol \( \ast \) refers to the convolution operator. The problem of image recovery deals with how to get the clear image \( f \) based on the observed data \( g \).

This is generally an ill-posed inverse problem, and we need to employ some sort of regularization to make it well-posed. One of the famous one is the ROF model [40], proposed by Rudin, Osher, and Fatemi, which uses total variation regularization and recovers \( f \) by minimizing the functional

\[
E(f) = \lambda \int_{\Omega} |\nabla f| \, dx + \frac{1}{2} \| k \ast f - g \|_2^2.
\]

In (1–5), the first term refers to the total variation of \( f \), which is a very famous regularization approach due to its excellent ability in preserving edges while still keeping the smoothness of the recovered image. The second term is a data-fidelity term, and \( \lambda > 0 \) is a parameter which balances the strength of these two terms.

For the Gaussian white noise case, the ROF model could be quite efficient because the second term turns out to be the log-likelihood of the noise if we assume the variance is fixed. However, in real applications, the noise may not obey Gaussian distribution and the ROF model would fail. A typical example is impulse noise, which is often generated by malfunctioning pixels in camera sensors, faulty memory locations in hardware, or
erroneous transmission. Impulse noise has two common types, salt-and-pepper noise and random-valued impulse noise. Please refer to Chapter 4 for a detailed definition.

Classical impulse noise removal techniques are generally based on median-based filters [41] and its various modifications [42–47]. The problem with these filter-based denoising methods is that they tend to modify pixels that are not affected by noise and thus bring blur to the recovered images especially when the noise ratio is high. To improve performance, various decision based methods have been proposed, like the median filter based on homogeneity information [48], the boundary discriminative noise detection [49], directional weighted median filter [50], modified adaptive center-weighted median filter [51] and regularized data-preserving methods [52]. These decision based methods generally include two steps: they first identify possible noisy pixels and then replace them using certain type of filters or some regularization method while leaving all other pixels unchanged. All of the above mentioned methods only consider the noise removal problem, and thus they could not handle the case when the image is contaminated by blur and impulsive noise simultaneously.

There are also some interesting work on image deblurring in the presence of impulse noise. In [53–55], the authors proposed a model involving a L1 data fidelity and the Mumford-Shah regularization term. The two-phase approach is employed in [56–58]. After identifying the possible noisy pixels using median-type filters, the authors in [56, 57] use Mumford-Shah regularization while [58] use TV based regularization only on the data samples that are not outlier candidates. In [59], Yang et al. proposed an efficient algorithm, FTVd (fast total variation deblurring), and the authors in [60] used augmented lagrange method to for TVL1 model to deblur images with impulse noise. Liu et al. [61, 62] developed a TV regularized and adaptive L2-based weighted model to deblur images in the presence of impulse noise.

We also want to mention some novel image recovered models based on sparse representation theory. The first are decision based methods [63, 64] aiming at removing
impulse noise. The novelty of these two methods is that they first detect the noisy pixels using the known methods ([63] uses [49], and [64] employs [51]), and then apply the sparse representation theory to recover the noisy pixels. In [65], Lou et al. proposed a method that directly deals with the sparse representation of blurred images. Their intuition is that the sparse coefficients of a latent image with respect to an over-complete basis are the same as those that encode the blurred version of the image with respect to the blurred version of the over-complete basis. However, it is only used for Gaussian noise and not applicable for impulse noise.

In Chapter 4, we propose a novel approach for deblurring images corrupted by impulse noise. The key point is to approximate the probability density function of impulse noise by mixed Gaussian distributions. We also use the sparse representation theory to regularize the image. Experimental results are provided to demonstrate the effectiveness of the proposed method.
2.1 Problem Statement and Related Works

Let \( \Omega \) be a bounded Lipschitz domain, \( I : \Omega \to \mathbb{R} \) be a given image and \( C \) be an arbitrary curve in the domain \( \Omega \). The segmentation problem can be formulated as move \( C \) such that it separates the foreground from the background. To do so, we generate a binary image \( L \) corresponding to the position of \( C \) in the following way

\[
L(x) = \begin{cases} 
F & \text{if } x \in R; \\
B & \text{if } x \in R^c. 
\end{cases}
\]

(2–1)

\( R \) and \( R^c \) denote the region inside and outside \( C \) respectively.

Note that the label image \( L \) in this setting changes as the curve \( C \) evolves. When \( C \) reaches the position of the right segmentation, the intensities inside \( C \) and outside \( C \) have different statistics, for instance, different means or (and) variances. Obviously, the intensities inside \( C \) and outside \( C \) for the label image \( L \) are always two different constants. Therefore, when \( C \) provides a good segmentation, the image \( I \) and label image \( L \) should be better matched statistically. This is the basic idea of using matching \( I \) and \( L \) to assist segmentation.

However, it is usually not easy to match these two images as they are of different modalities and it does not make sense to directly compare their intensities. To cope with this difficulty, a number of similarity measures based on statistical dependence have been proposed. For instance, [20] chooses to maximize the mutual information between original image \( I \) and its label image \( L \), together with a constraint of a length term, i.e.,

\[
E(C) = -MI(I, L) + \lambda \int_C ds. 
\]

(2–2)

In (2–2), \( I \) and \( L \) are viewed as random variables. At each point \( x \in \Omega \), the image intensity \( I(x) \) (or \( L(x) \)) is a sample drawn from the random variable \( I \) (or \( L \)) and all the
samples \( \{I(x)|x \in \Omega\} \) (or \( \{L(x)|x \in \Omega\} \)) are assumed to be independent. The mutual information \( MI(I, L) \) is defined as follows:
\[
MI(I, L) = h(I) - h(I|L)
\]
\[
= h(I) - Pr(L = F)h(I|L = F) - Pr(L = B)h(I|L = B),
\]
where \( h(\cdot) \) refers to the the entropy. For a continuous random variable \( Z \), the entropy of \( Z \) is defined as:
\[
h(Z) = - \int_{\mathbb{R}^n} p_Z(z) \log p_Z(z) dz.
\]
Since \( h(I) \) is independent of the curve \( C \), we only need to estimate \( h(I|L = F) \) and \( h(I|L = B) \), which are estimated by using the nonparametric Parzen window density strategy, i.e.,
\[
h(I|L = F) \approx - \frac{1}{|R|} \int_{\mathbb{R}} \log \left( \frac{1}{|R|} \int_{\mathbb{R}} K(I(x) - I(\hat{x})) d\hat{x} \right) dx,
\]
\[
h(I|L = B) \approx - \frac{1}{|R^c|} \int_{R^c} \log \left( \frac{1}{|R^c|} \int_{R^c} K(I(x) - I(\hat{x})) d\hat{x} \right) dx.
\]
In the above equations, \( K(\cdot) \) is the so-called window function or kernel, which is positive, symmetric, vanishing at infinity and satisfies
\[
\int_{\mathbb{R}^2} K(s) ds = 1.
\]
For instance, we can choose \( K \) to be the Gaussian p.d.f., i.e.,
\[
K(s) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left( -\frac{s^2}{2\sigma^2} \right),
\]
which is the Parzen-window density estimation kernel. Note that this kernel is infinitely differentiable and thus leading to the same property for the estimated p.d.f.

The mutual information could be effectively used as a similarity measure to match the image to be segmented and its label image. However, it requires to estimate the joint p.d.f. of \( I \) and \( L \), which is sensitive to image quantization and also increases the complexity of computation. In this work, we choose to use Rényi’s statistical measure,
maximum cross correlation, as a similarity measure. This measure deals directly with samples and does not need to estimate the continuous joint p.d.f.. A brief review of this measure is provided in the next section.

2.2 Rényi’s Statistical Measure

In [66], Rényi proposed a set of postulates for a suitable dependence measure \( Q \) of two random variables/vectors \( X \) and \( Y \), which has drawn much attention ever since. These postulates include

1. \( Q(P_{X,Y}) \) is well-defined;
2. \( 0 \leq Q(P_{X,Y}) \leq 1 \);
3. \( Q(P_{X,Y}) = 0 \) if and only if \( X, Y \) are independent;
4. \( Q(P_{X,Y}) = 1 \) if \( Y = f(X) \) or \( X = g(Y) \), where \( f \) and \( g \) are Borel measurable functions.

\( P_{X,Y} \) refers to the joint density function of \( X \) and \( Y \). If \( Q(P_{X,Y}) \) satisfies the above postulations, then it can be used as a dependence measure. Rényi also showed that one function satisfying these conditions is

\[
Q(P_{X,Y}) = \sup_{f, g \in V} CC(f(X), g(Y)),
\]

where \( V \) is the space of all Borel measurable functions with finite positive variance, and \( CC(f(X), g(Y)) \) is the correlation coefficient of \( f(X) \) and \( g(Y) \), i.e.,

\[
CC(f(X), g(Y)) = \frac{\text{cov}(f(X), g(Y))}{\sqrt{\text{var}(f(X))} \sqrt{\text{var}(g(Y))}}.
\]

(2–3)

In (2–3), the covariance between \( f(X) \) and \( g(Y) \) is defined as

\[
\text{cov}(f(X), g(Y)) = \text{E}[(f(X) - E[f(X)])(g(Y) - E[g(Y)])],
\]

where \( E[f(X)] \) is the expectation value of \( f(X) \).

The difficulty of using Rényi’s measure lies in the fact that we need to find the optimal \( f \) and \( g \) in the space \( V \), which is the set of all Borel measurable functions with
finite positive variance. It is extremely difficult to search $f$ and $g$ in such a huge space. Fortunately, we have shown in [67] that the supremum in $\mathcal{V}$ could be attained in a much smaller space, which is a reproducing kernel Hilbert space (RKHS) associated with a reproducing kernel that is continuous, symmetric, positive definite and vanishing at infinity. For ease of presentation, we include the RKHS theory in the Appendix. In this work, we choose the Gaussian function to be the reproducing kernel, i.e.,

$$K(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right).$$

According to the theory of RKHS (see [67] or Appendix), any two functions $f$ and $g$ in the RKHS associated with the Gaussian kernel can be approximated by functions $p$ and $q$ of the form,

$$p(x) = \sum_{i=1}^{n} \frac{\alpha_i}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - y_i)^2}{2\sigma^2}\right),$$

and

$$q(x) = \sum_{j=1}^{m} \frac{\beta_j}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - z_j)^2}{2\sigma^2}\right),$$

for some parameters $\sigma, y_i, \alpha_i, z_j, \beta_j$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. In practice, we can choose $f$ and $g$ to be of the above form and fix $\sigma, y_i, z_j$. Thus we only need to estimate the coefficients $\alpha_i$ and $\beta_j$, which could significantly simplify the computation.

### 2.3 Proposed Model and Numerical Method

In section 2.3, we propose our model and corresponding numerical schemes. Our aim is to find a curve $C$ such that the resulted label image $L$ defined in (2–1) matches the best with the original image $I$. In this work we use the Rényi’s statistical dependence measure, maximum correlation coefficient, as a similarity measure to align $I$ and $L$. By using this measure we don’t need to estimate the continuous joint p.d.f. of the two images as in the models based on mutual information.

As shown in the previous postulates, when the image $I$ and its label image $L$ are functions of each other, i.e., $L = f(I)$ or $I = g(L)$ for some function $f$ or $g$, the maximum
cross correlation attains its maximum value 1. Note that \(L\) is piecewise constant, so is the resulting image \(g(L)\). It does not make a big difference by maximizing the cross correlation between \(f(I)\) and \(L\) or \(g(L)\), so in the following we choose to maximize the cross correlation between \(f(I)\) and the label image \(L\). The objective energy functional is obtained by combining the cross correlation of \(f(I)\) and \(L\) and the length of \(C\), i.e.,

\[
E(C, a_1, \ldots, a_n) = \int_C \left| \nabla H(\phi(x)) \right| dx + \lambda \left( 1 - CC(f(I), L) \right)^2 \tag{2–4}
\]

where

\[
f(I(x)) = \sum_{i=1}^{n} a_i K(I, y_i),
\]

and the corresponding label image \(L(x)\) is defined by

\[
L(x) = \begin{cases} 
  c_1 & \text{if } x \in R; \\
  c_2 & \text{if } x \in R^c. 
\end{cases}
\]

2.3.1 Level set formulation and numerical method

Energy functional (2–4) can be minimized using the level set approach [6, 34, 68, 69]. The curve \(C\) is represented by the zero level of a Lipschitz function \(\phi : \Omega \to \mathbb{R}\) and the resulting energy functional becomes

\[
E(\phi, a_1, \ldots, a_n) = \int_{\Omega} |\nabla H(\phi(x))| dx + \lambda \left( 1 - CC(f(I), L) \right)^2 \tag{2–5}
\]

where \(H\) is the Heaviside function and

\[
L(x) = c_1 H(\phi(x)) + c_2 (1 - H(\phi(x))).
\]

The alternate minimization (AM) approach [70] is employed to solve this problem. First, we keep \(a_1, \ldots, a_n\) fixed and solve for \(\phi\) using the gradient descent approach, i.e.,

\[
\frac{\partial \phi}{\partial t} = \delta(\phi) \left[ \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (1 - CC(f(I), L)) F \right], \tag{2–6}
\]

25
where $\delta$ is the regularized Dirac function and

$$F = \frac{(f(I) - \overline{f(I)})\text{var}(L) - \text{cov}(f(I), L)(L - \overline{L})}{\text{var}(f(I))^{\frac{1}{2}}\text{var}(L)^{\frac{3}{2}}} \cdot (c_1 - c_2). \quad (2\text{-}7)$$

In (2–6), $\delta(\phi)\text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right)$ is the first variation of the smoothing term (TV-term) in (2–5) with respect to $\phi$. In the same way, $\delta(\phi)\lambda(1 - CC(f(I), L))F$ is the first variation of the fidelity term in (2–5) with respect to $\phi$.

Note that $\delta(\phi)F$ is the first variation of $CC(f(I), L)$ with respect to $\phi$ and this derivation can be obtained by applying the chain rule, that is, the first variation of $CC(f(I), L)$ with respect to $L$ times the first variation of $L$ with respect to $\phi$, i.e.,

$$\delta(\phi)F = \frac{dCC(f(I), L)}{d\phi} = \frac{dCC(f(I), L)}{dL} \cdot \frac{dL}{d\phi}. \quad (2\text{-}8)$$

Note that

$$CC(f(I), L) = \frac{\text{cov}(f(I), L)}{\sqrt{\text{var}(f(I))}\sqrt{\text{var}(L)}} = \frac{\mathbb{E}[(f(I) - \overline{f(I)})(L - \overline{L})]}{\sqrt{\mathbb{E}[(f(I) - \overline{f(I)})^2]} \sqrt{\mathbb{E}[(L - \overline{L})^2]}},$$

thus

$$\frac{dCC(f(I), L)}{dL} = \frac{(f(I) - \overline{f(I)})\text{var}(L) - \text{cov}(f(I), L)(L - \overline{L})}{\text{var}(f(I))^{\frac{1}{2}}\text{var}(L)^{\frac{3}{2}}} \cdot (c_1 - c_2). \quad (2\text{-}9)$$

On the other hand, $\frac{dL}{d\phi} = \delta(\phi)(c_1 - c_2)$. Therefore, we can get (2–7) through (2–8) and (2–9).

Then we keep $\phi$ fixed and minimize (2–5) with respect to $a_i$‘s. In this thesis, we present two algorithms to update $a_i$‘s. The first one is to use the gradient descent method

$$\frac{\partial a_i}{\partial t} = (1 - CC(f(I), L))E.$$

In the above equation, $E$ is the first variation of $CC(f(I), L)$ with respect to $a_i$, i.e.,

$$E = \frac{\text{cov}(p_i, L)\text{var}(f(I)) - \text{cov}(f(I), L)\text{cov}(f(I), p_i)}{\text{var}(f(I))^{\frac{1}{2}}\text{var}(L)^{\frac{3}{2}}},$$
where
\[ p_i = K(I, y_i). \]

The second method is based on the formulation of the energy functional. To minimize (2–5) with respect to \( a_i \)'s is equivalent to maximizing \( CC(f(I), L) \) with respect to \( a_i \)'s as the length term is independent of \( a_i \)'s.

For simplicity, we first introduce some notations. The image \( I \) and \( L \) are viewed as vectors of length \( N \), where \( N \) is the number of pixels. For each \( i = 1, 2, \cdots, n \), set \( p_i = K(I, y_i) \) be a vector of length \( N \), and define \( P = [p_1, p_2, \cdots, p_n] \), \( a = [a_1, a_2, \cdots, a_n] \).

For any vector \( x \), we denote the mean of \( x \) to be \( \bar{x} \). Thus
\[
 f(I) = \sum_{i=1}^{n} a_i K(I, y_i) = \sum_{i=1}^{n} a_i p_i = Pa,
\]
and
\[
 CC(f(I), L) = \frac{\langle Pa - \bar{P}a, L - \bar{L} \rangle}{|Pa - \bar{P}a||L - \bar{L}|} = \frac{\langle P_0 a, L_0 \rangle}{|P_0 a||L_0|},
\]
where \( P_{0ij} = P_{ij} - \frac{1}{N} \sum_{k=1}^{N} P_{k,j} \) and \( L_0 = L - \bar{L} \).

Let \( \hat{P} \) be a matrix with orthonormal column vectors which spans the column space of \( P_0 \), then we have \( P_0 a = \hat{P} \alpha \) for some vector \( \alpha \). Under these formulations, we get
\[
 CC(f(I), L) = \frac{\langle P_0 a, L_0 \rangle}{|P_0 a||L_0|} = \frac{\langle \hat{P} \alpha, L_0 \rangle}{|\alpha||L_0|} = \frac{\alpha^T \hat{P}^T L_0}{|\alpha||L_0|},
\]
(2–10)

In (2–10), \( U \Lambda V^T \) is the singular value decomposition of the matrix \( \hat{P}^T \), where \( U \) is an \( n \times n \) unitary matrix, \( V^T \) is an \( N \times N \) unitary matrix and \( \Lambda \) is an \( n \times N \) diagonal matrix with nonnegative diagonal entries \( \lambda_1, \lambda_2, \cdots, \lambda_n \), which are listed in decreasing order.

Now let \( x = \frac{U^T \alpha}{|U^T \alpha|}, y = \frac{V^T L_0}{|V^T L_0|} \), then \( x \in \mathbb{R}^n, y \in \mathbb{R}^N, |x| = |y| = 1 \), thus
\[
 CC(f(I), L) = x^T \Lambda y = x^T \Theta \Theta z = (\Theta x, \Theta z),
\]
27
where \( z = [y_1, y_2, \cdots, y_n] \) and \( \Theta \) is an \( n \times n \) diagonal matrix with diagonal entries 
\[
\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n}.
\]

Therefore, \( CC(f(I), L) \) is maximized when \( \Theta x \) and \( \Theta z \) have the same direction. Note that \( |x| = 1 \), we get 
\[
x = \Theta^+ (\Theta z),
\]
where \( \Theta^+ \) refers to the pseudo-inverse of \( \Theta \). From the definition of \( x \), we can pick a particular \( \alpha = Ux \) and then we can update \( a \) from the relation 
\[
P_0 a = \hat{P} \alpha.
\]

In this work, \( \phi \) and \( a_i \)'s are alternatively updated until we reach a satisfactory result. To increase the rate of convergence, the semi-implicit difference scheme is applied in (2–6), i.e.,
\[
\frac{\phi^{n+1} - \phi^n}{\Delta t} = \delta(\phi^n) \left[ \text{div} \left( \frac{\nabla \phi^{n+1}}{|\nabla \phi^n|} \right) + \lambda (1 - CC(f(I)^n, L^n)) F^n \right],
\]
This equation could be effectively solved by using the additive operator splitting (AOS) method. Note that the curvature term in (2–6) is approximated by
\[
\frac{\partial}{\partial x} \left( \frac{\phi_x}{\sqrt{\phi_x^2 + \phi_y^2 + \epsilon^2}} \right) + \frac{\partial}{\partial y} \left( \frac{\phi_y}{\sqrt{\phi_x^2 + \phi_y^2 + \epsilon^2}} \right),
\]
where \( \epsilon \) is a small positive number in case that the denominators become zero. However, it may still cause stability issues and limit the convergence rate.

2.3.2 A soft formulation and numerical method
To avoid local minimum problem we propose a soft formulation of the energy functional (2–4) by using the same strategy in [36], and use Chambolle’s dual method [35, 71] to solve it.
Let \( u : \Omega \rightarrow [0, 1] \) be a fuzzy membership function and rewrite the energy functional as
\[
E(u, a_1, \cdots, a_n) = \int_{\Omega} |\nabla u(x)| dx + \frac{\lambda}{2} (1 - CC(f(I), L))^2,
\]
where
\[
L(x) = c_1 u(x) + c_2 (1 - u(x))
\]
In real applications, we want $L$ to be a binary image so that it could give a reasonable segmentation. So during each iteration, we reset

$$L(x) = c_1 \chi(u > 0.5) + c_2(1 - \chi(u > 0.5)),$$

and $\chi$ refers to the characteristic function.

Following the strategy in [35, 71], we introduce an auxiliary variable $\nu : \Omega \rightarrow [0, 1]$ and consider the following approximated energy functional

$$E(u, \nu, a_1, a_2, \cdots, a_n) = \int_\Omega |\nabla u(x)|dx + \frac{1}{2\theta} \| u - \nu \|^2 + \lambda(1 - CC(f(I), L))^2,$$

where

$$L(x) = c_1 \nu(x) + c_2(1 - \nu(x)),$$

and $\theta$ is chosen to be small enough such that the minimizers $u^*$ and $\nu^*$ are close to each other.

We still employ the alternate minimization (AM) approach to solve this minimization problem, i.e., we go on to alternatively solve the following two problems:

$$\min_u \int_\Omega |\nabla u(x)|dx + \frac{1}{2\theta} \| u - \nu \|^2 \quad (2-11)$$

and

$$\min_{0 \leq \nu \leq 1} \frac{1}{2\theta} \| u - \nu \|^2 + \lambda(1 - CC(f(I), L))^2. \quad (2-12)$$

The minimization problem (2-11) could be effectively solved by applying Chambolle’s method [71] and the solution is

$$u(x) = \nu(x) - \theta \text{div} p(x),$$

where $p = (p^1, p^2)$ is given by

$$- \nabla(\theta \text{div} p - \nu) + |\nabla(\theta \text{div} p - \nu)|p = 0. \quad (2-13)$$
Equation (2–13) could be solved by a fixed point method, i.e.,

\[ p^{n+1} = \frac{p^n + \tau \nabla (\text{div} p^n - v/\theta)}{1 + \tau \nabla |\text{div} p^n - v/\theta|}. \]

Following the same strategy in [35], the solution \( v \) of (2–12) is given by

\[ v = \min(\max(u + \lambda \theta G, 0), 1), \]

where

\[ G = (1 - CC(f(I), L)) \cdot (c_1 - c_2) \cdot \frac{(f(I) - \overline{f(I)}) \text{var}(L) - \text{cov}(f(I), L)(L - \overline{L})}{\text{var}(f(I))^{\frac{1}{2}} \text{var}(L)^{\frac{3}{2}}}. \]

\( a_i \)'s are only included in the cross correlation term, so the optimization scheme is exactly the same as the level set approach.

We want to mention that both of these two methods (level set method and the soft segmentation by using Chambolle’s method) are quite effective. However, it is still possible that we may result in a local minimum. This is due to the non-convexity of the cross correlation term in the energy functional (2–4). Therefore, (2–4) may have more than one minimizer and we cannot guarantee this approach converges to a global minimizer.

### 2.4 Experimental Results

In section 2.4, we show our experimental results on various images to demonstrate the performance of the proposed model for segmentation. All the simulations are performed in Matlab 7.9 (R2009b) on a PC with an Intel Core 2 Duo CPU at 2.4 GHz and 3 GB RAM.

We compare the proposed nonparametric model with two parametric models, namely, the Chan-Vese model (1–2) and also the parametric Gaussian model (1–3)(1–4).

For completeness, we rewrite these models as the following

\[ \min_{c,c_1,c_2} \int_{\Omega_1} (c_1 - I)^2 dx + \int_{\Omega_2} (c_2 - I)^2 dx + \beta |C|, \quad (2–14) \]
Table 2-1. Number of iterations and CPU time (s) for the experiments (Figure 2-1, 2-2, and 2-3).

<table>
<thead>
<tr>
<th>Images</th>
<th>Size</th>
<th>Model (2–14)</th>
<th>Model (2–15)</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iteration</td>
<td>CPU</td>
<td>Iteration</td>
<td>CPU</td>
</tr>
<tr>
<td>Fig.1(a)</td>
<td>100 × 100</td>
<td>50</td>
<td>2.93</td>
<td>63</td>
</tr>
<tr>
<td>Fig.1(e)</td>
<td>100 × 100</td>
<td>176</td>
<td>4.48</td>
<td>182</td>
</tr>
<tr>
<td>Fig.1(i)</td>
<td>256 × 256</td>
<td>86</td>
<td>7.07</td>
<td>90</td>
</tr>
<tr>
<td>Fig.2(a)</td>
<td>210 × 180</td>
<td>-</td>
<td>-</td>
<td>98</td>
</tr>
<tr>
<td>Fig.2(f)</td>
<td>210 × 180</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Fig.3(a)</td>
<td>336 × 406</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fig.3(e)</td>
<td>336 × 406</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

and

\[
\min_{c, c_1, c_2, \sigma_1, \sigma_2} \sum_{i=1}^{2} \int_{\Omega_i} \left( \frac{(l - c_i)^2}{2\sigma_i^2} + \log \sigma_i \right) \, dx + \beta |C|. \tag{2–15}
\]

We apply the same soft formulation to the above two models and compare the results with those got from the proposed model. The numerical algorithm would be terminated once \( \| u^{k+1} - u^k \| / \| u^k \| < 10^{-5} \), where \( u^k \) refers to the level set function or the membership function during the \( k^{th} \) iteration. The corresponding number of iterations and CPU time for all the following experiments are summarized in Table . From this table, we can see that compared to the parametric models, the proposed nonparametric model needs less time and iterations to obtain good results.

For the level set approach, the initial \( \phi \) is set to be the signed distance function of the initial circles in the image. However, Chambolle’s approach is more robust to the initialization. The initial \( u \) and \( v \) can be generated as random fields in the range \([0, 1]\). All of the test images are rescaled to the interval \([0, 1]\) and the parameters \( c_1, c_2 \) in the label image are fixed to be 1 and 2. Unless otherwise stated, in each figure, we include the test image \( l \), the final transformed image \( f(l) \), the final label image \( L \) and the segmentation result, i.e., the contours \((\phi = 0 \text{ or } u = 0.5)\) superimposed on the original image.
2.4.1 Comparison with two parametric models

Figure 2-1. Segmentation results of three synthetic images. (a)(f)(k) Test images $I$; (b)(g)(l) Segmentation results of the piecewise constant Mumford-Shah model; (c)(h)(m) Results of the parametric Gaussian model; (d)(i)(n) Results of the proposed model; (e)(j)(o) Distributions of the foreground (blue) and background (red).

The purpose of Experiment I (Figure 2-1) is to test the ability of the proposed model on three synthetic images, where the foreground and background are generated by two different distributions. Meanwhile, we also compare the segmentation results with the Chan-Vese model (2–14) and the parametric Gaussian model (2–15).

In each row, from left to right, we present the original synthetic test image, corresponding boundary overlaid on the image obtained from the the Chan-Vese model(2–14), the parametric Gaussian model (2–15), the proposed model, the distributions of the foreground (blue) and background (red).
In the first test image (Figure 2-1(a)), the intensities of foreground and background are drawn from two Gaussian distributions with different mean and the same variance (Figure 2-1(e)). While the intensities of the foreground and background in the second test image (Figure 2-1(f)) share the same mean and different variance (Figure 2-1(j)). The last test image (Figure 2-1(k)) shows a uni-modal Gaussian foreground over a bi-modal Gaussian background (Figure 2-1(o)).

We can easily see that all the above-mentioned three models work effectively for the first test image. However, the Chan-Vese model fails for the second and third test images as the piecewise constant assumption does not hold for these two cases. As for the parametric Gaussian model, it could automatically estimate the mean and variance between and outside the curve and then use these information to aid the segmentation. Therefore, the result for the second test image (Figure 2-1(f)) is satisfactory even if the image intensities overlaps. However, it fails for the third test image (Figure 2-1(k)) as the background is bi-modal Gaussian distributed, which makes the assumption for the parametric Gaussian model do not hold.

Our proposed nonparametric model successfully separates the foreground from the background in all these three cases. The image intensities of object and background in the first test image vary a lot in the first test image (Figure 2-1(a)) and it would be easily segmented. Regarding the last two images, the background or the foreground is comparatively homogeneous. Hence, we could select $y_i$’s from the intervals where the intensities of the background or the background lie. Thus after applying the function $f$, linear combination of a series Gaussian functions centered at $y_i$’s, to the image $I$, the values of $f(I)$ in the inhomogeneous region become extremely small and the transformed images $f(I)$ are more homogenous. Therefore, by carefully choosing $y_i$’s in the function $f$ and maximizing the correlation coefficient of the transformed images $f(I)$ with the label image $L$, we get the desirable results.
The comparison with these two models indicates that the parametric models fail when the image intensity distribution is more complicated. That is why we need to explore for more general nonparametric models.

**2.4.2 Test on real medical images**

We test the proposed model on two real medical images, an MR brain image and a lung image.

In Experiment 2 (Figure 2-2), the first test image (Figure 2-2(a)) is a clean brain image and the second test image (Figure 2-2(f)) is generated by adding Gaussian noise with zero mean and variance 0.1 to the clean image (Figure 2-2(a)). Meanwhile, we also compare the results with those obtained from the parametric Gaussian model (2–15).

The segmentation results of the parametric Gaussian model (2–15) are placed at the end of each row (Figure 2-2(e), 2-2(j)). We can see that it successfully separates background, cerebrospinal fluid (csf) from white matter and gray matter. However, this is quite different from the results (Figure 2-2(d), 2-2(i)) obtained from the proposed model. From (Figure 2-2(d), 2-2(i)), we can see that gray matter is separated from the rest and the whole image is actually segmented into three parts even if we only do the two phase segmentation.

This is reasonable because we utilize the histogram information (Figure 2-2(k)) of the test image. As indicated in the graph (Figure 2-2(k)), there are three peaks, which from left to right stand for background and csf, gray matter, white matter. Note that background and csf are considered as a whole and the intensities of the gray matter lie in the interval \([0.4, 0.6]\). We uniformly select \(y_i's\) from the interval \([0.4, 0.6]\), then after applying the function \(f\) to the test images \(I\), intensities which are not in this interval (white matter, background and csf), would become almost zero while intensities of the gray matter are enlarged. That is why white matter, csf and background look dark while the gray matter looks bright in the transformed images \(f(I)\) (Figure 2-2(b), 2-2(g)).
other words, by doing this transformation, we can view white matter, csf and background as a whole and separate them from the gray matter.

The last image (Figure 2-2(l)) shows that the cross correlation between the image \( f(I) \) and the label image \( L \), i.e., \( CC(f(I), L) \), keeps increasing as the iteration process goes. And this trend coincides with the mutual information between the test image \( I \) and the label image \( L \). So we can conclude that the proposed method is consistent with the mutual information based nonparametric image segmentation method [20].

Figure 2-1 and 2-2 indicate that by choosing specific \( y_i \)'s, the proposed model works well for images with inhomogeneity, unevenly distributed illumination and it can get multiphase segmentation results while only using two phases. In the following experiment, we do not pay too much attention on the selection of \( y_i \)'s and let them to be equally spaced in the interval \([0, 0.5]\).

Experiment 3 (Figure 2-3) aims to test whether the proposed model works for images with fine structures. We choose the test image to be a lung image with lots of fine details. This first test image (Figure 2-3(a)) is a clean image and the second one (Figure 2-3(e)) is more inhomogeneous, which is generated by adding Gaussian noise with zero mean and variance 0.1 to the clean image. After applying the function \( f \) to the original test images \( I \) (Figure 2-3(a), 2-3(e)), the resulted images \( f(I) \) (Figure 2-3(b), 2-3(f)) have more strong contrast between different features while still preserving the detailed structures. The same parameters are applied for these two tests. The final results (Figure 2-3(d), 2-3(h)) show that most of the fine structures are captured and the noise inhomogeneity does not exert a big difference.

2.5 Conclusion

In Chapter 2, we propose a novel image segmentation framework based on Rényi’s statistical dependence measure. The computation is greatly simplified by applying the theory of reproducing kernel Hilbert space. Two numerical approaches, the levelset method and Chambolle’s dual approach, are employed during the implementation.
Figure 2-2. Segmentation results of a clean brain image (first row) and its noisy version (second row). From left to right: Test image $I$ (a)(f); Final transformed images $f(I)$ (b)(g); Final label images $L$ (c)(h); Segmentation results of the proposed model (d)(i); Segmentation results obtained from the parametric Gaussian model (e)(j). Third row: (k) Histogram of the test image (a); (l) the information of $CC(f(I), L)$ and $MI(I, L)$ during each iteration.
Figure 2-3. Segmentation results of a clean lung image (first row) and a noisy one (second row). From left to right: Test images I (a)(e), transformed images $f(I)$ (b)(f), label images $L$ (c)(g) and the final contour ($u = 0.5$) superimposed on the test images (d)(h).

Finally, the proposed model is applied to different kinds of images and gets satisfactory results.
CHAPTER 3
AN EFFICIENT ALGORITHM FOR MULTIPHASE IMAGE SEGMENTATION WITH INTENSITY BIAS CORRECTION

3.1 Model Formulation

In Section 3.1, we establish a generalized multiphase image segmentation framework for joint region partitioning and intensity bias correction.

3.1.1 Multiphase Segmentation and MAP Approach

Let \( I : \Omega \rightarrow \mathbb{R} \) be the input image to be segmented, the image segmentation problem is to find a set of regions \( \{\Omega_i\}_{i=1}^M \) such that \( \Omega = \bigcup_{i=1}^M \Omega_i, \{\Omega_i\}_{i=1}^M \) are disjoint. This is equivalent to solving for the collection of characteristic functions \( \chi_i(x) \) of \( \Omega_i \), where

\[
\chi_i(x) = \begin{cases} 
1 & \text{if } x \in \Omega_i, \\
0 & \text{otherwise},
\end{cases}
\]

for \( i = 1, \cdots, M \), and \( \sum_i \chi_i(x) = 1, \forall x \in \Omega \).

As addressed in the introduction section, the input image \( I \) can be corrupted by noise and unknown intensity bias field \( b : \Omega \rightarrow \mathbb{R} \). This process is usually modelled as

\[ I(x) = b(x)l_0(x) + n(x), \]

where \( l_0 \) is the ideal clean image. Consider the simple case where the ideal image is constant \( c_i \) in each region \( \Omega_i \), and the noise \( n(x) \) is normally distributed and independent of those at other locations. More precisely, if \( x \in \Omega_i \), then

\[ I(x) = b(x)c_i + n_i(x), \]

where \( n_i(x) \) is normally distributed with mean zero and unknown variance \( \sigma_i^2 \). It is worth noting that different applications may yield changes in the modeling of (3–2) [21]. Nevertheless, the derivation and resulting algorithms given below still work with appropriate modifications accordingly.
To this end, we can see that a complete solution package to an image segmentation with bias field estimation problem is \( \{ \chi, b, c, \sigma \} \), where \( \chi = (\chi_1, \cdots, \chi_M)^T : \Omega \to \{0, 1\}^M \), \( c = (c_1, \ldots, c_M)^T \in \mathbb{R}^M \) and \( \sigma = (\sigma_1, \ldots, \sigma_M)^T \in \mathbb{R}^M_+ \). Here \( \chi \), the \( i \)-th component of \( \chi \), is the characteristic function of \( \Omega_i \) as defined in (3–1), \( c_i \) and \( \sigma_i^2 \) represent the original mean intensity and noise variance in region \( \Omega_i \), respectively. Finally, \( b \) is unknown bias field that causes intensity inhomogeneity in the image.

We first find the posterior probability distribution \( p(\{\chi, b, c, \sigma\}|l) \) of \( \{\chi, b, c, \sigma\} \) given image \( l \), and then obtain an optimal segmentation and bias field estimation by the maximum-a-posteriori (MAP). Note that the Bayes’ rule implies that

\[
p(\{\chi, b, c, \sigma\}|l) \propto p(l|\{\chi, b, c, \sigma\})p(\{\chi, b, c, \sigma\}). \tag{3–3}
\]

Therefore, we need to determine \( p(\{\chi, b, c, \sigma\}) \), the prior information imposed to \( \{\chi, b, c, \sigma\} \), and \( p(l|\{\chi, b, c, \sigma\}) \), the joint distribution of pixel intensities given \( \{\chi, b, c, \sigma\} \).

### 3.1.2 Modeling the Intensity Inhomogeneity

Based on model (3–2), one can readily see that \( l(x) \) is normally distributed as

\[ N(b(x)c_i, \sigma_i^2) \]

if \( x \in \Omega_i \) (or equivalently \( \chi_i(x) = 1 \)) given the segmentation \( \{\chi, b, c, \sigma\} \). However, the observed intensity \( l(x) \) is merely one realization and it is usually not reliable to recover \( \chi, b, c \) and \( \sigma \) simultaneously. To overcome this difficulty, we estimate intensity density function based on pixel intensities in a local image patch. More precisely, we assume a multiplicative density structure of \( l(x) \) as follows,

\[
p(l(x)|\{\chi, b, c, \sigma\}) \propto \prod_{y \in W_x^\rho} (p(l(y)|\{\chi, b, c, \sigma\})^\pi_x(y), \tag{3–4}
\]

where \( W_x^\rho = \{y \in \Omega : |y - x| \leq \rho\} \) is a round image neighborhood with prescribed radius \( \rho \) and centered at \( x \). On the right hand side of (3–4), we consider that \( l(y) \) closely follows the model (3–2) and contributes to the density function of \( l(x) \) via a weighted product as in (3–4). Note that (3–4) is an ad hoc modification of the log-likelihood function and thus the method is not really a MAP method in the rigorous sense.
In (3–4), \( \pi_x : W^\rho_x \to [0, 1] \) gives the weights of intensity distributions of the points in \( W^\rho_x \) such that \( \sum_{y \in W^\rho_x} \pi_x(y) = 1 \). One can simply choose \( \pi_x(y) = 1/|W^\rho_x| \) for all \( y \in W^\rho_x \) if the intensities of neighbor points in \( W^\rho_x \) make equal contributions to the probability distribution \( p(I(x)|\{\chi, b, c, \sigma\}) \). In this thesis, we use more adaptive weights \( \pi_x(y) \) according to the distance from \( y \) to the center \( x \) via

\[
\pi_x(y) = K_s(y - x),
\]

where \( K_s \) a (truncated) Gaussian kernel function defined by

\[
K_s(z) = \begin{cases} 
C \exp \left(-|z|^2/2s^2\right), & \text{if } |z| \leq \rho \\
0, & \text{otherwise.}
\end{cases}
\]

for some \( s > 0 \). In (3–5), \( C \) is a normalizing constant that makes \( \int_{|z|\leq\rho} K_s(z)dz = 1 \).

We observe that the intensity bias field in practical applications usually varies gradually across the image domain. In other words, the value \( b(y) \) is nearly constant for points \( y \) in an image patch \( W^\rho_x \) provided that \( \rho \) is not too large. Therefore, we approximate \( b(y) \) by \( b(x) \), the bias at the center point \( x \), and obtain that \( I(y) \sim N(b(x)c, \sigma^2_i) \) for \( y \in W^\rho_x \cap \Omega_i \). Hence, the joint distribution \( p(I|\{\chi, b, c, \sigma\}) \) in (3–4) can be written as

\[
p(I|\{\chi, b, c, \sigma\}) = \prod_{x \in \Omega} \prod_{y \in W^\rho_x} p(I(y)|\{\chi, b, c, \sigma\})^{K_s(y-x)}
\]

where \( p(I(y)|\{\chi, b, c, \sigma\}) \) is Gaussian-type

\[
\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left(-\frac{|I(y) - b(x)c|^2}{2\sigma_i^2}\right)
\]

for those points \( y \) that \( \chi_i(y) = 1 \). By now, we have established the conditional probability density \( p(I|\{\chi, b, c, \sigma\}) \) in (3–3).

On the other hand, we set the prior of \( \chi \) according to the descriptive length of the boundaries \( \partial \Omega \), to exponential distribution with parameter \( \alpha \), which implicitly penalizes
undesired irregular and zigzag partition curves. The priors of $b$, $c$ and $\sigma$ are imposed non-informatively. In addition, terms in $\{\chi, b, c, \sigma\}$ are assumed to be independent. Consequently, the prior $p(\{\chi, b, c, \sigma\})$ can be simplified to

$$
p(\{\chi, b, c, \sigma\}) \propto \prod_{i=1}^{M} \exp(-\alpha|\partial\Omega_i|).
$$

(3–7)

Based on (3–6) and (3–7), the MAP of (3–3) is equivalent to the following minimization after we apply negative logarithm to both sides of (3–3),

$$
\min_{\chi, b, c, \sigma} \left\{ \alpha \sum_{i=1}^{M} |\partial\Omega_i| + L(\{\chi, b, c, \sigma\}) \right\}.
$$

(3–8)

Here $L(\{\chi, b, c, \sigma\})$ is the negative log-likelihood function

$$
L(\{\chi, b, c, \sigma\}) = -\log p(I|\{\chi, b, c, \sigma\})
$$

$$
= \int_{\Omega} \sum_{i=1}^{M} \int_{W^\rho_i \cap \Omega_i} K_s(y-x) l_i(y; x) dy dx
$$

(3–9)

and $l_i(y; x)$ is defined for $y \in \Omega_i \cap W^\rho_x$ by

$$
l_i(y; x) := \frac{|I(y) - b(x)| c_i^2}{2\sigma_i^2} + \frac{1}{2} \log(2\pi\sigma_i^2).
$$

(3–10)

In (3–9), we substitute the summation by integral to accommodate the continuous setting of our derivation, and omit $W^\rho_x$ and $\Omega_i$ in the last equality according to the definitions of $K_s$ and $\chi_i$ in (3–5) and (3–1), respectively.

To write the first term in (3–8) using the characteristic functions $\chi_i$, we recall that the total variation of a function $f : \Omega \to \mathbb{R}$ is defined by

$$
TV(f) = \sup_{\rho \in \mathcal{Y}} \left\{ -\int_{\Omega} f \text{div} \rho dx \right\}
$$

(3–11)
where the admissible set \( Y \) is
\[
Y := \{ p \in C_0^\infty(\Omega; \mathbb{R}^d) : |p(x)| \leq 1, \forall x \in \Omega \}.
\] (3–12)

Since \( \chi_i \) is the characteristic function of \( \Omega_i \), the total variation of \( \chi_i \) is then the descriptive length of \( \partial \Omega_i \), namely,
\[
TV(\chi_i) = |\partial \Omega_i|.
\] (3–13)

Plug (3–13) and (3–9) into (3–8), we obtain a generalized multiphase segmentation model as follows,
\[
\min_{\chi, b, c, \sigma} \sum_{i=1}^M \left\{ \alpha \cdot TV(\chi_i) + \int_\Omega \chi_i(x) h_i(x) dx \right\}
\] (3–14)
subject to the constraint that only one component in \( \chi(x) = (\chi_1(x), \ldots, \chi_M(x))^T \) is one and the rest are zeros at each \( x \in \Omega \). In (3–14), the function \( h_i \) is defined by
\[
h_i(x) = \int_\Omega K_s(y-x) l_i(x; y) dy.
\] (3–15)

Note that the second term in (3–14) is obtained by exchanging the symbols \( x \) and \( y \), followed by switching the order of integrations in (3–9).

### 3.2 Numerical Algorithm

Although the segmentation problem has been unified to (3–14), this minimization problem cannot be solved efficiently in general due to the nondifferentiability of the TV term, and the nonconvexity of the objective function with respect to \( \{\chi, b, c, \sigma\} \). Conventional approaches based on level set formulation require extensive computations and suffer the local minimums severely. In this thesis, we develop a numerical algorithm to tackle problem (3–14) effectively.

We first relax the constraint on the characteristic function \( \chi = (\chi_1, \ldots, \chi_M)^T \) in (3–14) to \( X \) defined by
\[
X := \{ u : \Omega \rightarrow [0, 1]^M | u(x) \in \Delta^M, \forall x \in \Omega \}
\] (3–16)
and the canonical simplex $\Delta^M$ is defined by

$$
\Delta^M = \{(z_1, \cdots, z_M)^T \in \mathbb{R}_+^M : z_1 + \cdots + z_M = 1\}
$$

The relaxed model of (3–14), using notation $u$ instead of conventional binary function $\chi$, becomes

$$
\min_{u, b, c, \sigma} \sum_{i=1}^M \left\{ \alpha TV(u_i) + \int_\Omega u_i(x) h_i(x) dx \right\} \quad (3–17)
$$

subject to $u \in X$. One can readily see that the original constraint in (3–14) further requires $u(x)$ to be one of the vertexes of $\Delta^M$. This relaxation substitutes extends the solution set of $u$ to a continuous and convex set. Note that this relaxation is similar to the idea proposed in [35–37] and please refer to the corresponding papers for more detailed information.

In the rest part of this section, we use alternating minimizations to construct an iterative algorithm. Namely, we need to minimize the objective function with respect to one of the variables in $\{u, b, c, \sigma\}$ with others being fixed.

### 3.2.1 First Variations of $b$, $c$, and $\sigma$

First of all, we observe that the variables $b$, $c$, and $\sigma$ only appear in the second term of the objective function in (3–17), and their solutions can be obtained by first variations.

Fix $u$, $c$ and $\sigma$, we compute the Euler-Lagrangian (E-L) equation for $b$ and obtain

$$
b(x) = \frac{\sum_{i=1}^M (c_i/\sigma_i^2)[K_s * (u, l)](x)}{\sum_{i=1}^M (c_i^2/\sigma_i^2)[K_s * u_i](x)}, \quad x \in \Omega, \quad (3–18)
$$

where $*$ is the convolution operator. Next, we fix $u$, $b$, and $\sigma$, and obtain the E-L equation of $c_i$ for each $i = 1, \cdots, M$ as

$$
c_i = \frac{\int_\Omega [K_s * (u, b^2)](x) dx}{\int_\Omega [K_s * (u, b^2)](x) dx}. \quad (3–19)
$$
Finally we have the E-L equation of $\sigma$, as
\[
\sigma_i^2 = \frac{\int_\Omega \left( [K_s * (u_i l^2)] - 2 c_i b [K_s * l] + c_i b^2 \right) dx}{\int_\Omega [K_s * u_i] dx}.
\] (3–20)

Therefore, the updates of $b$, $c$ and $\sigma$ have closed forms and the main computations are regular convolutions using kernel function $K_s$ defined in (3–5).

3.2.2 Solution to $u$

Now we turn to the minimization of the objective function (3–17) with respect to $u$. For fixed $b$, $c$ and $\sigma$, the minimization can be written as
\[
\min_{u \in X} \sum_{i=1}^M \left\{ \alpha TV(u_i) + \int_\Omega u_i(x) h_i(x) dx \right\},
\] (3–21)

where $h_i$ does not depend on $u$ according to its definition in (3–15). We remark that (3–21) is a constrained nonsmooth optimization problem due to the constraint on $u(x) \in \Delta^M$ for each $x \in \Omega$ and the nondifferentiable TV term in the objective function. So we need to find an effective way to tackle these two issues.

For each $u_i$, we introduce the dual variable $p_i \in Y$ according to the definition in (3–11), and reformulate the minimization problem (3–21) as a min-max problem
\[
\min_{u \in X} \max_{p \in Y} \sum_{i=1}^M \left\{ -\alpha \int_\Omega u_i \text{div} p_i dx + \int_\Omega u_i h_i dx \right\},
\] (3–22)

where $X$ is defined in (3–16) and $Y$ is the admissible set of $p_i$’s defined in (3–12).

In the discrete setting where the image $I$ consists of $N$ pixels, we can vectorize each $u_i$ into a column vector in $\mathbb{R}^N$, then its dual variable $p_i$ is a matrix in $\mathbb{R}^{N \times D}$, where $D$ is the dimension of the image (e.g. 2 or 3). Hence, the optimization problem (3–22) can be written as
\[
\min_{u \in X} \max_{p \in Y} F(u, p) := \sum_{i=1}^M \langle u_i, \alpha \nabla^T p_i + h_i \rangle,
\] (3–23)
where $\nabla : \mathbb{R}^N \to \mathbb{R}^{N \times D}$ is the discretized gradient operator, the superscript $^T$ is the conjugate operator, and $\langle \cdot, \cdot \rangle$ represents the regular inner product in $\mathbb{R}^N$. 

44
Note that both of $X$ and $Y$ are closed and convex sets. Hence a solution to the min-max problem (3–23) can be obtained by alternately solving for the primal variable $u$ and dual variable $p$

$$u^{k+1} = \Pi_X([u_i^{k} - \delta_k \nabla_u F(u^k, p^k)]_{i=1}^M) \quad (3–24)$$

$$p_i^{k+1} = \Pi_Y(p_i^k + \tau_k \nabla_p F(u^{k+1}, p^k)), \forall i \quad (3–25)$$

where $\delta_k$ and $\tau_k$ act as the step sizes of the primal and dual variables $u$ and $p$ in the $k$-th iteration, respectively, and $[u_i]_{i=1}^M$ denotes the matrix $[u_1, \ldots, u_M]$ that has $u_i$ as columns. Here $\Pi_X : \mathbb{R}^{N \times M} \to X$ and $\Pi_Y : \mathbb{R}^{N \times d} \to Y$ are projection operators onto the sets $X$ and $Y$, respectively. More precisely, $\Pi_X$ maps each row of its argument, say $z \in \mathbb{R}^M$, to the simplex $\Delta^M$ using the algorithm shown in [72] (ref. Appendix B), and $\Pi_Y$ projects each row of its argument, say $z \in \mathbb{R}^d$, to the unit ball $B^d := \{z \in \mathbb{R}^d : \|z\|_2 = 1\}$ via

$$z \mapsto \frac{z}{\max\{\|z\|_2, 1\}}.$$ 

According to the definition of $F(u, p)$ in (3–23), we know that (3–24) and (3–25) have closed forms as

$$u^{k+1} = \Pi_X([u_i^{k} - \delta_k (\alpha D^T p_i^{k+1} + h_i)]_{i=1}^M), \quad (3–26)$$

$$p_i^{k+1} = \Pi_Y(p_i^k + \tau_k Du_i^k), \quad i = 1, \ldots, M. \quad (3–27)$$

We note that the projections $\Pi_X$ and $\Pi_Y$ in (3–26)(3–27) have complexity $M \log M$ and $D$, respectively. Therefore, the main computational cost is $NM(\log M + D)$ in each iteration. Note that $M$ is the number of phases in the image and is usually less than 10, and $D$ is the dimension of the image such as 2 or 3. Moreover, these projections are applied to each of the $N$ pixels and hence the computations in both of $\Pi_X$ and $\Pi_Y$ can be carried out in parallel. On the contrary, level set function based segmentation with commonly used semi-implicit gradient descent scheme usually requires Gauss
eliminations to solve tridiagonal linear systems, and hence the computation cannot be parallelized easily.

### 3.2.3 Algorithm

In conventional settings of alternating minimizations, we need to iterate (3–26), (3–27) until convergence to get $u$ before updating the other variables $b$, $c_i$, and $\sigma_i$, $i = 1, \cdots, M$. However, we found that empirically it is more efficient to simply solve for $u$ and $p$ only once and immediately update the remaining variables.

The stopping criterion of the proposed algorithm is set to $\|u^k - u^{k-1}\|_2 / \|u^k\|_2 < \epsilon_{\text{tol}}$. Namely, the computation is automatically terminated if the relative change in the iterate $\{u^k\}$ is less than a prescribed tolerance value $\epsilon_{\text{tol}}$.

Since we relax the constraint on the function $u$, the resulting $u$ may contain values in $(0, 1)$ and hence are not characteristic functions. Therefore, we further threshold the components of $u(x)$ to $\bar{u}(x)$ by

$$
\bar{u}_i(x) = \begin{cases} 
1 & \text{if } u_i(x) = \max_{1 \leq j \leq M} \{u_j(x)\} \\
0 & \text{otherwise}
\end{cases}
$$

(3–28)

for each $x \in \Omega$. If there are several equally maximal values in $u(x)$, we just pick one randomly.

To sum up, we propose a fast segmentation (FastSEG) algorithm in Algorithm 1 below.

### 3.3 Experimental Results

In Section 3.3, we test the Algorithm 1 on a variety of images and compare with several recently proposed methods for image segmentation in the presence of noise and intensity bias.

#### 3.3.1 Experiment Settings

The proposed algorithm is implemented and all the tests are performed in MATLAB® 7.9 (R2009b) computing environment on a PC with Intel Dual Core 2 Duo
Algorithm 1: Fast Multiphase Segmentation (FastSEG)

Input $\alpha > 0$ and $\epsilon_{\text{tol}}$. Initialize $u^0$ and $p^0$, and set $b^0 = 1$, $k = 0$. 

repeat
- Update $c^k$ using (3–19) with $u^k$ and $b^k$;
- Update $\sigma^k$ using (3–20) with $u^k$, $b^k$ and $c^k$;
- Compute $h^k$ using (3–15) with $b^k$, $c^k$ and $\sigma^k$;
- Compute $u^k$ using (3–26) with $h^k$;
- Compute $p^k$ using (3–27) with $h^k$;
- Update $b^k$ using (3–18) with $u^k$, $c^k$ and $\sigma^k$;
- $k \leftarrow k + 1$.

until $\|u^k - u^{k-1}\|_2 / \|u^k\|_2 < \epsilon_{\text{tol}}$

Compute $\bar{u}$ using (3–28) and return $\{\bar{u}, b, c, \sigma\}$.

CPU at 2.4GHz (only one core is used in computation) and 3GB of memory. We expect a significant improvement in computation speed if the program is parallelized.

In this thesis, we always preselect $M$, the number of phases, for a given image. We use initialization in the same way as that of the comparison algorithms if available. In particular, as our algorithm does not employ level set functions, we set $u^0$ in Algorithm 1 to be the characteristic functions of the regions delineated by the initial contours in the comparison algorithms. For other experiments, we use K-means algorithm to the collection of pixel intensities of the given image, and assign $u^0_i(x) = 1$ if the K-means algorithm classifies $x$ to the $i$-th group and 0 otherwise for $i = 1, \cdots, M$. We remark here that, the minimizations arised from segmentation problems are nonconvex in general, and hence it is usually preferred to start from an initial guess close to the desired segmentation. The K-means algorithm can generate close approximations to the desired segmentation if there are slight noise and intensity bias. However, we can observe unsatisfactory initialization by K-means for many images tested in this paper. For instance, Figure 3-3(a) shows the initial guess obtained by K-means, which is not quite close to the final optimal segmentation shown in Figure 3-3(f). Nevertheless, we will show in 3.3.3.1 that how robust the proposed algorithm is with respect to different initializations.
We use test images and default parameter settings in the original code of comparison algorithms if available. For the proposed algorithm, the stopping criterion $\epsilon_{\text{tol}}$ is set to $10^{-3}$ throughout the experiments. The parameters $\alpha$, $\delta_k$ and $\tau_k$ are set to be $10^{-4}$, 0.2 and 0.5 respectively. The patch radius $\rho$ is set to 8, and the variance $s$ is 4 in (3–5). For all images we tested (whose intensities are scaled to [0,1]), they seem to provide good compromise between smoothness and accuracy as well as speed and stableness. We also found that moderate changes in these parameters do not yield significant difference in segmentation results.

3.3.2 Quantitative Evaluation and Comparison with Existing Methods

We use Jaccard similarity coefficient as a quantitative measure to evaluate the segmentation results. Let $\Omega_i$ be the $i$-th region obtained by the algorithm and $\bar{\Omega}_i$ be its corresponding region in the ground truth image, then the JSC between $\Omega_i$ and $\bar{\Omega}_i$ is defined as

$$J(\Omega_i, \bar{\Omega}_i) = \frac{|\Omega_i \cap \bar{\Omega}_i|}{|\Omega_i \cup \bar{\Omega}_i|},$$

where $| \cdot |$ represents the area of a region. Generally speaking, Jaccard similarity coefficients is bounded in $[0, 1]$ and larger values imply more accurate segmentation.

To demonstrate the effectiveness of the proposed model, we compare it with three recently developed methods in this field. For completeness, we give a brief summary regarding these methods in the following.

3.3.2.1 Methods Overview

The first method we are going to compare is the Weighted K-means Variational Level Set (WKVLS) method [31]. For the two-phase case, the WKVLS model can be written as

$$E_W(\phi, b, c_1, c_2) = \nu \int_{\Omega} |\nabla H(\phi)| dx + \mu \int_{\Omega} (|\nabla \phi| - 1)^2 dx + \int_{\Omega} \int_{\Omega} H(\phi)K_s(y-x)|l(y) - b(x)c_1|^2 dy dx + \int_{\Omega} \int_{\Omega} (1 - H(\phi))K_s(y-x)|l(y) - b(x)c_2|^2 dy dx,$$  \hspace{1cm} (3–29)
where $\phi$ is the level set function whose zero level set represents the partition contour, and $H$ is the Heaviside function defined by $H(z) = 1$ if $z \geq 0$ and 0 otherwise. The first two terms in (3–29) penalize the length of partition contour and force the level set function $\phi$ to be up straight (has slope 1) during evolutions, respectively. The last two terms in (3–29) are for data fitting as in the proposed algorithm, but lack the variability of noise level $\sigma_i$.

The next one is the Statistical and Variational Multiphase Level Set (SVMLS) method [33], which also utilizes level set formulation and minimizes the following energy functional

$$E_S(\Phi, b, c, \sigma) = \sum_{i=1}^{4} \int_{\Omega} \int_{\Omega} M_i(\Phi(y)) K_s(y - x) l_i(y; x) dy dx,$$

where $l_i(y; x)$ is the same as that in (3–10), $\Phi = (\phi_1, \phi_2)$, and $M_i(\Phi)$ is defined as follows:

$$\begin{align*}
M_1(\Phi) &= H(\phi_1)H(\phi_2), \\
M_2(\Phi) &= H(\phi_1)(1 - H(\phi_2)), \\
M_3(\Phi) &= (1 - H(\phi_1))H(\phi_2), \\
M_4(\Phi) &= (1 - H(\phi_1))(1 - H(\phi_2)).
\end{align*}$$

The last method we would compare is the Coherent Local Intensity Clustering (CLIC) method [29]. CLIC partitions an image by solving a constrained minimization problem,

$$E_C(b, u, c) = \sum_{i=1}^{M} \int_{\Omega} \int_{\Omega} u_i(y) K_s(y - x) |l(y) - b(x) c_i|^2 dy dx,$$

subject to $\sum_{i=1}^{M} u_i(x) = 1, \ \forall \ x \in \Omega$.

3.3.2.2 Experiment 1

In the first experiment, we compare the proposed model with the aforementioned three methods on an MR brain image with strong intensity inhomogeneity and noise. We use the default test image (shown in Figure 3-1(a)) from the source code package.
of SVMLS published online\(^1\). The initial conditions (shown in Figure 3-1(b)) is also the default setting for the SVMLS code and is used for all the tested algorithms. For this experiment, we only provide visual results in Figure 3-1 since a ground truth segmentation is not available for this data set.

The input image shown in Figure 3-1(a) contains strong intensity inhomogeneity and hence it is difficult to distinguish different tissue intensities from its histogram as shown in Figure 3-1(c). Therefore, conventional approaches based on intensity clusters cannot yield correct segmentations in this case. On the other hand, all the four tested algorithms can generate reasonable results by taking the intensity inhomogeneity into account. As the intensity bias field \(b\) is also estimated by both algorithms, we plot \(I/b\), the images after bias correction, in the second row of Figure 3-1. Here \(/\) represents pointwise division. It can be seen that the corrected images have less intensity biases compared to 3-1(a). This can also be observed in their histograms shown in the third row of Figure 3-1. The histograms of the corrected images have clear intensity peaks and hence different tissues can be distinguished more easily.

When we look into the details of the segmentation results, we can observe that those obtained by CLIC and the proposed algorithm are more accurate than those by WKVLS and SVMLS: the former two can better separate the gray and white matters as indicated by the red arrows in Figure 3-1(l), 3-1(o). One of the possible reasons is that WKVLS and SVMLS are formulated in that level set framework and hence can be easily trapped into local minimum.

\(^1\) http://www4.comp.polyu.edu.hk/~cslzhang/code.htm
3.3.2.3 Experiment 2

The test image in Experiment 2 is an MR image obtained from BrainWeb\textsuperscript{2}. As ground truth segmentation is available in this case, we use Jaccard similarity coefficient to evaluate the performance of the test algorithms quantitatively.

In Figure 3-2, we use the same initialization for all the above mentioned segmentation models as shown by the red and blue rectangles in Figure 3-2(a). The ground truth obtained from BrainWeb are presented in Figure 3-2(b), which consists of four parts: background, white matter, gray matter and cerebrospinal fluid (CSF). The segmentation results obtained by the proposed algorithm, WKVLS, SVMLS and CLIC are shown in Figure 3-2(c), 3-2(d), 3-2(e), 3-2(f), respectively. The Jaccard similarity coefficients and CPU time are summarized in Table 3-1. WKVLS and SVMLS cannot return correct segmentation as evolution of the level set functions can be easily stuck at a local minimum. The result of CLIC looks better than those got from WKVLS and SVMLS, but it contains too many superfluous points due to the lack of proper regularization in such noisy case. In general, the proposed model is more accurate and efficient than all the three methods, which can be addressed in the following aspects. Firstly, instead of using level set formulation, we directly choose characteristic functions, which can avoid the possible local minimum phenomena and reinitialization process. Next, we use total variation to regularize the characteristic functions, which is more accurate as it turns out to the length of the subregions. Finally, by applying the primal-dual formulation of TV norms and special properties of canonical simplex, the proposed algorithm involves only convolutions with Gaussian kernels and projections to the unit ball and simplex. Therefore, the proposed model is more accurate and efficient.

\textsuperscript{2} \url{http://www.bic.mni.mcgill.ca/brainweb/}
Table 3-1. Quantitative evaluation of Figure 3-2. Jaccard similarity coefficients $JSC$ of the four regions, background (b), CSF (c), grey matter (g) and white matter (w), obtained by the test algorithms and their CPU times in seconds.

<table>
<thead>
<tr>
<th>Method</th>
<th>$JSC_b$</th>
<th>$JSC_c$</th>
<th>$JSC_g$</th>
<th>$JSC_w$</th>
<th>CPU (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WKVLS</td>
<td>48.87%</td>
<td>5.51%</td>
<td>25.63%</td>
<td>6.68%</td>
<td>160.82</td>
</tr>
<tr>
<td>SVMLS</td>
<td>69.71%</td>
<td>18.79%</td>
<td>42.84%</td>
<td>61.68%</td>
<td>21.90</td>
</tr>
<tr>
<td>CLIC</td>
<td>98.75%</td>
<td>70.76%</td>
<td>60.57%</td>
<td>65.96%</td>
<td>41.07</td>
</tr>
<tr>
<td>Proposed</td>
<td>99.13%</td>
<td>80.47%</td>
<td>80.69%</td>
<td>82.54%</td>
<td>33.75</td>
</tr>
</tbody>
</table>

3.3.3 Further Evaluations of the Proposed Algorithm

It is important that an automated segmentation procedure is robust with respect to different initial segmentations, intensity bias status, noise level, and parameter settings. In Subsection 3.3.3, we further evaluate the performance of Algorithm 1 on these aspects.

3.3.3.1 Robust to initialization

As the objective functions appeared in segmentation problems are usually nonconvex, most algorithms especially those formulated using level set functions, suffer local minimums and hence are very sensitive to initializations. On the contrary, the proposed algorithm 1 appears to be robust: we test Algorithm 1 on an MR image using five different initializations as shown in Figures 3-3(a) (generated by K-means), 3-3(b), 3-3(c), 3-3(d), and 3-3(e) (generated by some seeds shown in red, green and blue squares). The final characteristic functions $\chi_i$ obtained by the proposed algorithm are shown in Figures 3-3(f), 3-3(g), 3-3(h), 3-3(i) and 3-3(j), respectively. The results imply that the proposed algorithm is quite robust with respect different initial conditions.

Table 3-2. Quantitative evaluation of Figure 3-3. Jaccard similarity coefficients $JSC$ of the four regions, background (b), CSF (c), grey matter (g) and white matter (w), obtained by the test algorithms and their CPU times in seconds.

<table>
<thead>
<tr>
<th></th>
<th>$JSC_b$</th>
<th>$JSC_c$</th>
<th>$JSC_g$</th>
<th>$JSC_w$</th>
<th>CPU (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f)</td>
<td>99.98%</td>
<td>99.58%</td>
<td>99.83%</td>
<td>99.91%</td>
<td>14.63</td>
</tr>
<tr>
<td>(g)</td>
<td>99.98%</td>
<td>99.46%</td>
<td>99.88%</td>
<td>99.97%</td>
<td>11.85</td>
</tr>
<tr>
<td>(h)</td>
<td>99.99%</td>
<td>99.58%</td>
<td>99.90%</td>
<td>99.97%</td>
<td>12.18</td>
</tr>
<tr>
<td>(i)</td>
<td>99.98%</td>
<td>99.46%</td>
<td>99.88%</td>
<td>99.96%</td>
<td>12.02</td>
</tr>
<tr>
<td>(j)</td>
<td>99.83%</td>
<td>96.75%</td>
<td>99.00%</td>
<td>99.46%</td>
<td>13.75</td>
</tr>
</tbody>
</table>
3.3.3.2 Robust to intensity inhomogeneity

We conduct more experiments on Algorithm 1 on MR images to test its capability on different intensity inhomogeneities. The results are shown in Figure 3-4. The original MR images are obtained from BrainWeb. We add synthetic intensity biases to the image (concentrated at the middle, top and bottom of the image domain, respectively), as shown in the first row of Figure 3-4. We show the corrected images $I/b$, the recovered bias fields $b$, segmentation results, histograms of the test images and corrected images under each of these three images in Figure 3-4. We can see that Algorithm 1 successfully detects the intensity biases and obtains desired segmentations regardless of bias status.

Table 3-3. Quantitative evaluation of Figure 3-4. Jaccard similarity coefficients $JSC$ of the three regions, background and CSF (bc), grey matter (g) and white matter (w) obtained by the test algorithms and their CPU times in seconds.

<table>
<thead>
<tr>
<th></th>
<th>$JSC_{bc}$</th>
<th>$JSC_{g}$</th>
<th>$JSC_{w}$</th>
<th>CPU (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>98.96%</td>
<td>91.65%</td>
<td>94.76%</td>
<td>8.34</td>
</tr>
<tr>
<td>Middle</td>
<td>98.82%</td>
<td>86.86%</td>
<td>91.78%</td>
<td>20.40</td>
</tr>
<tr>
<td>Right</td>
<td>99.48%</td>
<td>95.34%</td>
<td>95.87%</td>
<td>18.01</td>
</tr>
</tbody>
</table>

3.3.3.3 Robust to noise level

The purpose of this experiment is to test Algorithm 1 on MR images with intensity inhomogeneity and different levels of noise. The test images are generated by first multiplying a simulated bias field to the clean MR image and then adding low, medium, and strong Gaussian noise. The segmentation results are presented in Figure 3-5. From the results shown in the second row of Figure 3-5, we can see that the proposed algorithm consistently returns reasonable partitions of the image, but the accuracy can be slightly affected by the noise level.

3.3.3.4 Different parameter settings

As shown in Section 3.2, Algorithm 1 involves the penalty parameter $\alpha$, and step sizes $\tau_k, \delta_k$ for the primal and dual variables. We found that the proposed algorithm performs well for a variety of images under the same setting of these parameters.
However, the patch size $\rho$ used for local density weight calculation in (3–5) can impact the results under different level of noise. In the following experiments, we test Algorithm 1 with $\rho = 0, \cdots, 8$. In particular, we show the results with $\rho = 0, 2, 8$ in Figures 3-6(e)—3-6(g), respectively. Note that in Figure 3-6(e), there are many superfluous points in the case of $\rho = 0$, where local patch information is not utilized to estimate $I(x)$. On the other hand, the results look much better when we use the neighboring information. This suggests the importance of using local intensity density estimation (3–4), especially in the presence of strong noise.

Table 3-5. Quantitative evaluation of Figure 3-6. Jaccard similarity coefficients $JSC$ of the three regions, background and CSF (bc), grey matter (g) and white matter (w) obtained by the test algorithms and their CPU times in seconds.

<table>
<thead>
<tr>
<th>Method $\rho$</th>
<th>$JSC_{bc}$</th>
<th>$JSC_g$</th>
<th>$JSC_w$</th>
<th>CPU (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>89.06%</td>
<td>65.52%</td>
<td>82.07%</td>
<td>10.99</td>
</tr>
<tr>
<td>2</td>
<td>86.83%</td>
<td>67.14%</td>
<td>84.59%</td>
<td>12.30</td>
</tr>
<tr>
<td>8</td>
<td>89.61%</td>
<td>66.81%</td>
<td>86.15%</td>
<td>12.10</td>
</tr>
</tbody>
</table>

### 3.4 Concluding Remarks

In Chapter 3, we present a general multiphase soft segmentation framework which can deal with severe intensity inhomogeneity and noise in input images. Our model estimates the intensity distribution at a particular pixel using a multiplicative structure of distributions of all pixels in a neighborhood, and derive the minimization problem using MAP. To tackle the computational difficulty due to the highly nonsmooth and constrained formulation of the segmentation model, we apply primal-dual gradient projections to develop a fast numerical algorithm. Numerical results on various images show that
our method is more efficient and accurate in comparison with other recently proposed algorithms.
Figure 3-1. Comparison of the proposed model with WKVLS, SVMLS and CLIC on an MR brain image (a) with strong intensity inhomogeneity. The color squares in (a) depict initial contours used by all methods.
Figure 3-2. Comparison of the proposed model, WKVLS, SVMLS and CLIC on an MR brain image with severe noise and intensity inhomogeneity.
Figure 3-3. Robust to different initializations.
Figure 3-4. Robustness test of the proposed algorithm on different intensity inhomogeneity. Left, middle, and right columns show segmentation results on images with intensity bias in the middle, top, and bottom of image domain, respectively. From top to bottom: input images, segmentation results, estimated bias fields, corrected images, histograms of input images, and histograms of corrected images, respectively.
Figure 3-5. Robustness test of the proposed algorithm on different image noise levels. Left, middle, and right columns correspond to small, medium and strong noise levels, respectively. Top and bottom rows show the input images and segmentation results, respectively.
Figure 3-6. Efficiency of local intensity estimation in FastSEG when applied to images with strong noise.
CHAPTER 4
SPARSE IMAGE DEBLURRING IN THE PRESENCE OF IMPULSE NOISE

4.1 Background

Two common types of impulse noise are salt-and-pepper noise and random-valued impulse noise. If the level of impulse noise is \( p \), then the impulse noise degradation process can be summarized as: the intensity of every pixel stays the same with probability \( 1 - p \) and changes to some new value with probability \( p \). If the intensity changes, it will become the minimum or maximum value of the image intensity for salt-and-pepper noise and it will become a sample drawn from a uniform distribution on the image intensity interval for random-valued impulse noise.

In the following, we assume the clean image \( f \) is rescaled to the interval \([0, 1]\) and we consider the case when \( f \) is first degraded by blur and then the blurry image is further contaminated by impulse noise. Therefore, the observation image can be modeled by \( g = \mathcal{N}(k \ast f) \), where \( \mathcal{N} \) refers to the impulse noise operator. Mathematically speaking, the process can be modeled as

- **Salt-and-pepper noise**
  
  \[
  g_{i,j} = (\mathcal{N}(k \ast f))_{i,j} = \begin{cases} 
  0 & \text{with probability } \frac{p}{2}, \\
  1 & \text{with probability } \frac{p}{2}, \\
  (k \ast f)_{i,j} & \text{with probability } 1 - p.
  \end{cases}
  \]

- **Random-valued impulse noise**
  
  \[
  g_{i,j} = (\mathcal{N}(k \ast f))_{i,j} = \begin{cases} 
  a & \text{with probability } p, \\
  (k \ast f)_{i,j} & \text{with probability } 1 - p,
  \end{cases}
  \]

  where \( a \) is a sample drawn from a uniform distribution on \([0, 1]\).

It is not difficult to see that impulse noise is not additive, but we can view it as additive by setting

\[
 n = g - k \ast f = \mathcal{N}(k \ast f) - k \ast f, \quad (4.1)
\]

and the model \( g = k \ast f + n \) still holds. An experiment is carried out on a synthetic image to study the behavior of impulse noise. For better understanding, we plot the logarithmic
histogram of $n$ in Figure 4-1 (red). The histogram shows that the intensities of impulse noise accumulate at 0, sharply go down and fluctuate away from 0. However, there is no specific form regarding the probability distribution of impulse noise.

In [73], the authors proposed an efficient image restoration model based on EM algorithm. Later, in [62], they simplified the model as

$$E(f, u, \sigma_1^2, \sigma_2^2) = \frac{1}{2} \int_{\Omega} u \ln \sigma_1^2 dx + \frac{1}{2} \int_{\Omega} (1 - u) \ln \sigma_2^2 dx$$

$$+ \frac{1}{2} \int_{\Omega} \left( \frac{u}{\sigma_1^2} + \frac{1 - u}{\sigma_2^2} \right) (g - k * f)^2 dx + \lambda_1 \int_{\Omega} |\nabla u(x)| dx + \lambda_2 \int_{\Omega} |\nabla f(x)| dx. \quad \text{(4–2)}$$

The first four terms in (4–2) turn out to be the negative log-likelihood of two Gaussian mixture with different variances. In this thesis, we choose to use two randomly mixed Gaussian white noise to approximate the residual $n$. This is plausible as indicated in Figure 4-1. The green line refers to the logarithmic histogram of two different randomly mixed Gaussian white noise with variance 0.5 and 0.0001. This approximation

Figure 4-1. True logarithmic histogram of $n$ and its corresponding approximation.
looks quite good even if the variances are chosen by ourselves. In the following, we iteratively update the variances to get a better approximation.

### 4.2 Model Formulation

In Section 4.2, we give a different formulation about the model proposed in (4–2), which is closely related to our own work. After that, we propose a new model using sparse representation technique.

As discussed above, it is appropriate to approximate the distribution of $n$ using Gaussian mixture. Let $u : \Omega \rightarrow \{0, 1\}$ denote whether a pixel is noisy or not. More specifically, for each $x \in \Omega$, $u(x) = 1$ indicates the pixel is free of noise and $u(x) = 0$ implies $x$ is a noisy pixel. This immediately leads to a partition of the image domain $\Omega = \Omega_1 \cap \Omega_2$, with $\Omega_1 = \{x \in \Omega | u(x) = 0\}$ and $\Omega_2 = \{x \in \Omega | u(x) = 1\}$. If $x$ is a clean pixel, then we have $n(x) = 0$ by (4–1) and thus it is plausible to approximately view it as a sample drawn from the distribution $N(0, \sigma_1^2)$ with a sufficiently small $\sigma_1^2$. Nevertheless, if it is a noisy pixel, then we regard it as a sample of $N(0, \sigma_2^2)$ for a proper $\sigma_2^2$.

Our purpose is to find the the posterior probability distribution $p(\{f, u, \sigma_1^2, \sigma_2^2\}|g)$ of $\{f, u, \sigma_1^2, \sigma_2^2\}$ given the observed image $g$, and obtain an optimal recovery by the maximum-a-posteriori (MAP). Note that the Bayes’ rule implies that

$$p(\{f, u, \sigma_1^2, \sigma_2^2\}|g) \propto p(g|\{f, u, \sigma_1^2, \sigma_2^2\})p(\{f, u, \sigma_1^2, \sigma_2^2\}). \quad (4–3)$$

Therefore, we need to determine $p(\{f, u, \sigma_1^2, \sigma_2^2\})$, the prior information imposed to the segmentation $\{f, u, \sigma_1^2, \sigma_2^2\}$, and $p(l|\{f, u, \sigma_1^2, \sigma_2^2\})$, which is the joint distribution of pixel intensities in $l$ given the segmentation $\{f, u, \sigma_1^2, \sigma_2^2\}$.

Based on the above discussion, one can readily see that the random variables $\{n(x)|u(x) = 0\}$ are normally distributed as $N(n(x), \sigma_1^2)$ and the random variables $\{n(x)|u(x) = 1\}$ are distributed as $N(n(x), \sigma_2^2)$ given the values $\{f, u, \sigma_1^2, \sigma_2^2\}$. If we further assume all the random variables $\{n(x)|x \in \Omega\}$ are independent, then the joint
distribution \( p(n|\{f, u, \sigma_1^2, \sigma_2^2\}) \) can be written as

\[
p(n|\{f, u, \sigma_1^2, \sigma_2^2\}) = \prod_{i=1}^{2} \prod_{x \in \Omega_i} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left( -\frac{|n(x)|^2}{2\sigma_i^2} \right)
\]  

(4–4)

By now, we have established the conditional probability density \( p(n|\{f, u, \sigma_1^2, \sigma_2^2\}) \) in (4–4).

On the other hand, we can set the priors of \( \sigma_1^2 \) and \( \sigma_2^2 \) to (non-informative) uniform distributions. The prior of \( u \) is set to be the descriptive length of the boundaries \( \partial\Omega_1 \) (or \( \partial\Omega_2 \)) to the exponential distribution with parameter \( \alpha \). Moreover, terms in \( \{f, u, \sigma_1^2, \sigma_2^2\} \) are assumed to be independent. Regarding the prior for \( f \), let is ignore it for a moment or we can temporarily set it to be uniform distribution. Therefore, the prior \( p(\{f, u, \sigma_1^2, \sigma_2^2\}) \) can be simplified to

\[
p(\{f, u, \sigma_1^2, \sigma_2^2\}) \propto \exp(-\alpha|\partial\Omega_1|).
\]  

(4–5)

Based on (4–4), (4–5), the MAP of (4–3) is equivalent to the following minimization where we applied negative logarithm to both sides of (4–3),

\[
\min_{f, u, \sigma_1^2, \sigma_2^2} \{ |\partial\Omega_1| + L(\{f, u, \sigma_1^2, \sigma_2^2\}) \}.
\]  

(4–6)

Here \( L(\{f, u, \sigma_1^2, \sigma_2^2\}) \) is the negative log-likelihood function

\[
L(\{f, u, \sigma_1^2, \sigma_2^2\}) = -\log p(l|\{f, u, \sigma_1^2, \sigma_2^2\}) = \sum_{i=1}^{2} \int_{\Omega_i} \frac{|n(x)|^2}{2\sigma_i^2} + \frac{1}{2} \log(2\pi\sigma_i^2)dx
\]

\[
= \frac{1}{2} \int_{\Omega} u \ln \sigma_1^2 dx + \frac{1}{2} \int_{\Omega} (1 - u) \ln \sigma_2^2 dx + \frac{1}{2} \int_{\Omega} \left( \frac{u}{\sigma_1^2} + \frac{1 - u}{\sigma_2^2} \right) (g - k \ast f)^2 dx + \text{Const}.
\]  

(4–7)

Since \( u \) is the characteristic function of \( \Omega_1 \), the total variation of \( u_1 \) is then the descriptive length of \( \Omega_1 \), namely,

\[
|\partial\Omega_1| = TV(u) \triangleq \int_{\Omega} |\nabla u(x)| dx.
\]  

(4–8)
Plug \((4-8), (4-7)\) into \((4-6)\) and ignore the constant term, we obtain the image recovery model as follows,

\[
\min_{f,u,\sigma_l^2,\sigma_s^2} \left\{ \alpha TV(u) + L(\{f,u,\sigma_l^2,\sigma_s^2\}) \right\},
\]

where \(L(\{f,u,\sigma_l^2,\sigma_s^2\})\) is defined as that in \((4-7)\) subject to the constraint \(u\) only takes two values \((1\mathrm{ and }0)\). In this case, the solution set for \(u\) is not convex, so we it to the unit interval.

Finally, let us consider the prior for the recovered image \(f\). The clean images are usually smooth and allowed to have large jumps across the image boundaries. If we employ the total variation to regularize the recovered image \(f\), then we get the model in [62]. In this thesis, we propose a new model for image reconstruction by using the spare representation theory to regularize the recovered image \(f\).

4.3 Spare Representation Theory

The theory of sparse representation has gained a great deal of attention recently and has been successfully applied to the field of image processing, such as denoising [74, 75], deblurring [65, 76–78], super-resolution [79, 80], etc. To the best of our knowledge, there is little work on image deblurring in the presence of impulse noise via sparse representation theory, which is the main purpose of this paper.

Sparse representation theory is founded on the assumption that an image could be represented as a sparse linear combination of a series of atom images. The set of atoms is called a dictionary, which could either manually chosen or learned from training data set. The dictionary \(D\) is a redundant matrix in \(\mathbb{R}^{s \times t}\) for \(t > s\) \((s = 81\) and \(t = 324\) in our paper), aiming to handle image patches of size \(\sqrt{s} \times \sqrt{s}\), which are treated as column vectors in \(\mathbb{R}^s\) unless otherwise stated. Suppose the image \(f \in \mathbb{R}^{\sqrt{S} \times \sqrt{S}}\) \((S \gg s)\), we decompose it into \(J\) small patches of the size \(\sqrt{S} \times \sqrt{S}\) such that all the patches cover the entire image \(f\) and may overlap. Let \(R_j \in \mathbb{R}^{s \times S}\) be a mask with the property that \(R_j f\) extracts the \(j\)-th patch of \(f\).
For the moment, let us assume the dictionary $D$ is known. The spareland model assumes each patch $R_j f$ can be sparsely represented by the dictionary $D$. More specifically, there exist $\alpha_i \in \mathbb{R}^t$ such that

$$\| \alpha_j \|_0 < s < t, \quad \text{and} \quad D \alpha_j \approx R_j f, \quad j = 1, \ldots, J,$$

where $\| \cdot \|_0$ counts the number of nonzero components of the argument. Therefore, the sparsity of $f$ under the dictionary $D$ can be viewed as a regularization, which could be obtained by enforcing the following into our model

$$\min_{\alpha} \sum_{j=1}^{J} \left( \mu_j \| \alpha_j \|_0 + \frac{1}{2} \| D \alpha_j - R_j f \|_2^2 \right),$$

where $\alpha = (\alpha_1; \ldots, \alpha_J) \in \mathbb{R}^{t \times J}$.

Otherwise, if the dictionary is unknown, we can train a dictionary from a database and update it during the iteration process by using through the K-SVD algorithm [74, 81]. The K-SVD algorithm iteratively alternates between sparse coding of the training examples based on the current dictionary and updating the dictionary atoms to better fit the data. The outcome is a trained dictionary that can better fit the signals in the database under certain sparsity constraints and error tolerance.

### 4.4 Proposed Model and Numerical Schemes

Now it is time to present our model. We use sparse representation and total variation as regularization, and MLE (4–7) as data fidelity measure. The resulting model is formulated as a constrained minimization problem

$$\min_{f, u, 0 \leq u \leq 1, \sigma_1, \sigma_2, D, \alpha} E(f, u, \sigma_1, \sigma_2) = \alpha TV(u) + \lambda_2 \sum_j \left( \mu_j \| \alpha_j \|_0 + \frac{1}{2} \| D \alpha_j - R_j f \|_2^2 \right) + L(f, u, \sigma_1, \sigma_2),$$

(4–9)

Given the minimization problem (4–9), we apply the alternating minimization algorithm with respect to each variable to iteratively solve for the optimal solution. For
convenience, we denote

\[ H(f, u, \sigma_1^2, \sigma_2^2) = \frac{1}{2} \int_\Omega u \ln \sigma_1^2 \, dx + \frac{1}{2} \int_\Omega (1 - u) \ln \sigma_2^2 \, dx + \frac{1}{2} \int_\Omega \left( \frac{u}{\sigma_1^2} + \frac{1 - u}{\sigma_2^2} \right) (g - k \ast f)^2 \, dx. \]

First of all, for fixed \( f, u, D, \alpha \), we consider the variables \( \sigma_i^2, i = 1, 2 \), which can be obtained through their Euler-Lagrangian (E-L) equations

\[ \sigma_1^2 = \frac{\int_\Omega u (g - k \ast f)^2 \, dx}{\int_\Omega u \, dx}, \quad \sigma_2^2 = \frac{\int_\Omega (1 - u)(g - k \ast f)^2 \, dx}{\int_\Omega (1 - u) \, dx}. \] (4–10)

The \( u \) subproblem

\[ \min_{0 \leq u \leq 1} H(f, u, \sigma_1^2, \sigma_2^2) + \lambda_1 \int_\Omega |\nabla u(x)| \, dx \]

is a constrained TV problem, which is solved by the split Bregman algorithm.

\[
\begin{aligned}
(u^{n+1}, d^{n+1}) &= \arg \min_{0 \leq u \leq 1} H(f, u, \sigma_1^2, \sigma_2^2) + \lambda_1 \int_\Omega |d^n| \, dx + \gamma \int_\Omega |d^n - \nabla u - b^n|^2 \, dx, \\
b^{n+1} &= b^n + \nabla u^{n+1} - d^{n+1}.
\end{aligned}
\] (4–11)

The alternating minimization scheme is again employed to solve the problem (4–11).

The Euler-Lagrange equation of (4–11) with respect to \( u \) can be written as

\[ - \Delta u = \text{div} (b^n - d^n) - \left( \ln \sigma_1^2 - \ln \sigma_2^2 + \frac{1}{\gamma} \left( \frac{1}{\sigma_1^2} - \frac{u}{\sigma_2^2} \right) (g - k \ast f)^2 \right), \] (4–12)

which could be solved by Gauss-Seidel (GS) method. The \( d \) subproblem in (4–11) is a \( L^1 - L^2 \) minimization problem, which could be explicitly solved by the two dimensional shrinkage algorithm,

\[ d^{n+1} = \text{shrink}(\nabla u^{n+1} + b^n, \lambda_1/\gamma), \]

where the operator \text{shrink} is defined as

\[ \text{shrink}(y, \eta) = \frac{y}{|y|} \max\{|y| - \eta, 0\}. \]
Let $\omega = \frac{u}{\sigma_1^2} + \frac{1-u}{\sigma_2^2}$, then the Euler Lagrange equation of (4–9) associated with $f$ can be written as

$$\hat{k} \ast [\omega (k \ast f)] + \lambda_2 \left( \sum_j R_j^T R_j \right) f = \hat{k} \ast (\omega g) + \lambda_2 \sum_{j=1}^{J} R_j^T (D \alpha_j), \quad (4–13)$$

where $\hat{k}$ refers to the conjugated of $k$. Note that $\sum_j R_j^T R_j$ is a diagonal matrix with positive diagonal entries. So (4–13) could also be transformed to a linear matrix equation $Af = b$ with $A$ being positive symmetric definite and in this paper we choose the conjugate graduate (CG) method to solve it.

Finally, let us turn to the $(D, \alpha)$ subproblem. The $\alpha_j$ subproblem

$$\alpha_j^{n+1} = \min_{\alpha_j} \mu_j \| \alpha_j \|_0 + \frac{1}{2} \| D \alpha_j - R_j f^n \|_2^2,$$

is a sparse-coding problem, which is equivalent to

$$\alpha_j = \arg \min_{\alpha_j} \| \alpha_j \|_0, \ \text{s.t.} \ \| D \alpha_j - R_j f^n \| < \epsilon. \quad (4–14)$$

Several algorithm could be applied to solve this problem, such as thresholding, matching pursuit (MP), orthogonal matching pursuit (OMP), etc. In this paper, we choose to use OMP.

Given the coefficients $\alpha_j$, the dictionary $D$ is updated through the K-SVD algorithm [81]. According to the K-SVD algorithm, the dictionary is updated one column at a time. For a fixed column, the update can be done by performing a singular value decomposition (SVD) operation on its residual data matrix, computed only on the image patches that use this atom. In this way, the penalty term is guaranteed to drop and at the same time, the representation coefficients change as well, while keeping their sparsity structure.

To sum up, we propose an algorithm for sparse image restoration in Algorithm 2 below.
Algorithm 2: Sparse Image Deblurring and Denoising

Input dictionary $D$, $f^0 = g$, $b^0 = 0$, $d^0 = 0$, $u^0 = 0$, $(\sigma^2_i)^0$, $i = 1, 2$.

repeat
  Update $(\sigma^2_i)^{n+1}$ through (4–10), $i = 1, 2$;
  Sparse coding stage: solve (4–14) for each $\alpha_j$;
  Dictionary learning stage: update $D$ if necessary;
  Update $f^{n+1}$ through (4–13) using CG method;
  Update $u^{n+1}$ through (4–12) using GS method;
  $d^{n+1} = \text{shrink}(\nabla u^{n+1} + b^n, \lambda_1/\gamma)$;
  $b^{n+1} = b^n + \nabla f^{n+1} - d^{n+1}$;
  $n \leftarrow n + 1$.
until $\|f^n - f^{n-1}\|_\infty < \epsilon_{tol}$
Return $f^n$.

Remark: By saying update $D$ if necessary in the second step means there is no need to update $D$ each time during the iteration process. For one thing, it wastes time and is computationally expensive. For another, the dictionary varies a little if there is no significant enhancement about the recovered image $f$.

4.5 Experimental Results

In Section 4.5, a number of experiments are carried out to evaluate the performance of the proposed algorithm. Before going to the details, we would like to mention selection of the dictionary at the very beginning. As can be seen in the following experiments, the observed noisy blurry images contain less information and thus it is not suitable to train a initial dictionary based on the contaminated images. Like the approach in [65], the initial dictionary is trained from a series of clean images (excluding the test ones). In [65], for each test image, they train a specific dictionary based on the same category of images and fix the dictionary throughout the experiment. Instead, we train a single initial dictionary from a series of images for all the experiments and try to update the dictionary based on the recovered images during the iteration process. The reason is that during the iteration process, more information would be recovered and thus the dictionary based on the recovered images would be more plausible than the initial one.

The set of training images for the initial dictionary contain four different categories:
“people”, “building”, “animal” and “flower”, taken from the Berkeley Segmentation Database [82]. Part of the training images are shown in Figure 4-2.

Figure 4-2. Samples of training images.

Let \( f_{\text{ori}} \) be the original clean image and \( f \) be the restored image. We use peak signal to noise ratio (PSNR) to measure the goodness of the restored image,

\[
\text{PSNR} = 10 \log_{10} \frac{m_1 m_2}{\| f - f_{\text{ori}} \|_2^2},
\]

where \([m_1, m_2]\) refers to the size of the image.

All the test images are rescaled to the unit interval \([0, 1]\). The clean test images are corrupted by \(9 \times 9\) Gaussian kernel with \(\sigma = 15\) and further contaminated by impulsive noise. Part of the original images and blurry ones are shown in Figure 4-3.

We compare our results with two different image restoration models. The first one is the TVL1 model,

\[
\min_f \int_{\Omega} |\nabla f| + \lambda \| k * f - g \|_1,
\]

proposed in [59], which is solved by the FTVd algorithm and the code is published online\(^1\). The second one is model (4–2) proposed in [62]. For each algorithm, we test different parameters and choose the ones with the best recovery results.

4.5.1 Test with sever blur and 30\% impulse noise

In Section 4.5.1, we compare the proposed model with TVL1 and the model (4–2) in recovering images from blurry images contaminated by 30\% impulse noise. The results

\(^1\) http://www.caam.rice.edu/~optimization/L1/ftvd/
Figure 4-3. Part of the clean test images and their corresponding blurry ones.

for salt-and-pepper noise is presented in Figure 4-4 and the results for random-valued impulse are shown in Figure 4-5. The corresponding PSNR values are presented in Table 4-1.

It is easy to see that the results got from the FTVd algorithm (second row) suffer strong stair-casing effect, which has been efficiently suppressed in the results by the model (4–2) (third row) and the proposed model (forth row). Moreover, the proposed model is capable to recover more detailed information as shown in the texture of the image Barbara and also the coat of the bear. In addition, high PSNR values of the proposed model also reveals that the results of the proposed model are the best among the three.

4.5.2 Test with sever blur and different levels of impulse noise

The purpose of the second experiment (Figure 4-6) is to test the robustness of the proposed in recovering images with different levels of noise. The test images are shown in the first row of Figure 4-6, which are generated by first convolving the original clean image with a $9 \times 9$ Gaussian kernel with variance $\sigma = 15$. The resulted images are
Figure 4-4. Test on salt-and-pepper noise.
blurry noisy images (impulse noise)

FTVd1

Model (4–2)

Proposed Model (4–9)

Figure 4-5. Test on random-valued impulse noise.
further contaminated by 10%, 50% salt-and-pepper noise and 10%, 50% random-valued impulse noise, respectively. The results got from FTVdL1, model (4–2) and the proposed model are presented in the second, third and forth row. The corresponding PSNR values are shown in Table 4-2.

From Figure 4-6, we can see that when the results got from the FTVdL1 are satisfactory for the low noise level case, but the stair-casing effect becomes very strong when we further increase the noise level. On the other hand, both model (4–2) and the proposed one are not affected by the high noise level. In addition, the proposed model is very effective in recovering tiny detailed information as shown in the hair of Lena and its PSNR values are always the largest among all the three models.

Table 4-1. PSNR values for Figure 4-4, 4-5.

<table>
<thead>
<tr>
<th>Figures</th>
<th>Salt-and-pepper</th>
<th>Random-valued</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FTVdL1 Model (4–2)</td>
<td>Proposed</td>
</tr>
<tr>
<td>Barbara</td>
<td>27.5274</td>
<td>28.9324</td>
</tr>
<tr>
<td>Bear</td>
<td>33.3419</td>
<td>35.7737</td>
</tr>
<tr>
<td>Flower</td>
<td>32.0752</td>
<td>33.4399</td>
</tr>
</tbody>
</table>

Table 4-2. PSNR values for Figure 4-6.

<table>
<thead>
<tr>
<th>Figures</th>
<th>Salt-and-pepper</th>
<th>Random-valued</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FTVdL1 Model (4–2)</td>
<td>Proposed</td>
</tr>
<tr>
<td>Lena(10%)</td>
<td>28.1276</td>
<td>29.0068</td>
</tr>
<tr>
<td>Lena(50%)</td>
<td>25.4332</td>
<td>27.2297</td>
</tr>
</tbody>
</table>

4.6 Conclusion

In Chapter 4, we propose a spareland model for deblurring images in the presence of impulse noise. Impulse noise is non-additive. Our model uses mixture gaussian to approximate the residual information. Moreover, to avoid the stair-casing effects brought by total variation, we utilize the theory of sparse representation to regularize the recovered image. Experimental results on different images have demonstrate the accuracy of the proposed model.
Figure 4-6. Test on different level of noise.
APPENDIX A
REPRODUCING KERNEL HILBERT SPACE

Let $E$ be an arbitrary set and $H$ be a Hilbert space of real functions on $E$. We say that $H$ is a RKHS if the linear map $F_x : f \rightarrow f(x)$ is a bounded functional for any $x \in E$.

By this definition, $F_x \in H^*$, which is the dual space of $H$. Therefore, Reisz representation theorem shows that there exists a unique $K_x \in H$, such that

$$f(x) = \langle F_x, f \rangle = \langle K_x, f \rangle, \quad \forall f \in H.$$ 

Define $K : E \times E \rightarrow \mathbb{R}$ as $K(x, y) = \langle K(x, \cdot), K(y, \cdot) \rangle$. It is easy to see that $K$ has the following properties:

1. $K$ is symmetric: $K(x, y) = K(y, x)$.
2. Reproducing property: $f(x) = \langle K(x, \cdot), f(\cdot) \rangle$.
3. $K$ is positive definite: $\sum_{i,j} a_i a_j K(x_i, x_j) \geq 0$ holds for all $x_1, x_2, \cdots, x_n \in E$, $a_1, a_2, \cdots, a_n \in \mathbb{R}$ and the equality holds if and only if $a_i = 0$, $i = 1, 2, \cdots, n$.

We call such an $K$ the reproducing kernel for the Hilbert space $H$.

On the other hand, suppose $K : E \times E \rightarrow \mathbb{R}$ is symmetric and positive definite, then according to the Moore-Aronszajn theorem ([83]), there is a unique Hilbert space of functions on $E$ for which $K$ is a reproducing kernel. In fact, let $H_0(E)$ be the linear span of the functions $\{K(x, \cdot) | x \in E\}$ and define the inner product in $H_0(E)$ to be

$$(\sum_{i=1}^n a_i K(x_i, \cdot), \sum_{j=1}^m b_j K(y_j, \cdot)) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j K(x_i, y_j).$$

Let $H(E)$ be the completion of $H_0(E)$ with respect to this inner product. It is not difficult to check that $H(E)$ is the unique RKHS with reproducing kernel $K$.

For the particular case $E = \mathbb{R}$. Let $C_0(\mathbb{R})$ be the space of real valued continuous functions vanishing at infinity with the supremum norm. Then we have the following result

$$\sup_{f, g \in V(\mathbb{R})} CC(f(X), g(Y)) = \sup_{f, g \in H_0(\mathbb{R})} CC(f(X), g(Y)), \quad (A-1)$$
where $V(\mathbb{R})$ is the space of all real Borel measurable functions with finite positive variance. This is the main result we have used in this paper and (??)(??) follows directly from this result.

The proof of this result can be obtained through the following three steps (We omit the details here).

1. $H_0(\mathbb{R})$ is dense in $C_0(\mathbb{R})$.

2. Let $V(B)$ be the space of all real bounded Borel measurable functions, then

$$\sup_{f,g \in V(B)} CC(f(X), g(Y)) = \sup_{f,g \in C_0(\mathbb{R})} CC(f(X), g(Y)).$$

3. $$\sup_{f,g \in V(\mathbb{R})} CC(f(X), g(Y)) = \sup_{f,g \in V(B)} CC(f(X), g(Y)).$$
APPENDIX B
SIMPLEX PROJECTION ALGORITHM

The projection onto simplex $\Delta^M$ can be accomplished easily as shown in Algorithm 3 [84]. As can be seen, the main computation is in the sorting step, and hence complexity is $M \log M$. An inspiring derivation with detailed analysis of computational complexity can be found in [72].

Algorithm 3 Projection onto the simplex $\Delta^M$

Input $z = (z_1, \cdots, z_M)^T \in \mathbb{R}^M$;
Sort $z$ in the ascending order as $z_{(1)} \leq \cdots \leq z_{(M)}$, and set $i = M - 1$;
repeat
   Compute $t_i = (\sum_{j=i+1}^{M} z_{(j)} - 1)/(M - i)$.
   if $t_i \geq z_{(i)}$ then
      Set $\hat{t} = t_i$ and break
   else
      Set $i \leftarrow i - 1$
   end if
until $i = 0$
if $i = 0$ then
   Set $\hat{t} = (\sum_{j=1}^{M} z_j - 1)/M$
end if
Return $(z - \hat{t})_+$ as the projection of $z$ onto $\Delta^M$. 
REFERENCES


BIOGRAPHICAL SKETCH

Haili Zhang was born in Luoyang, Henan Province, P. R. China. In 2005, she received her Bachelor of Science in mathematics from Henan Normal University, P.R. China. After that, she was admitted to Beijing Normal University graduate program under the guidance of Dr. Zhongdan Huan and completed her Master of Science in applied mathematics in 2008.

In fall 2008, she joined the math graduate program at the University of Florida and started to work with Dr. Yunmei Chen. She received her Ph.D. in applied mathematics from the University of Florida in the summer of 2013. Her research interests are mathematical modeling, numerical analysis and optimization theories with applications to computer vision and imaging science using tools based on partial differential equations, calculus of variations and statistical methods.