MULTIPHASE IMAGE SEGMENTATION BASED ON INTENSITY STATISTICS: MODELING AND APPLICATIONS

By

FUHUA CHEN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

To my family

ACKNOWLEDGMENTS

I am very grateful to all my committee members. I want to thank them for spending time to attend my oral qualifying exam and final defense. In particular, I want to thank Dr. Rao, who helped me to better understand some stochastic concepts and principles. I want to thank Dr. Groisser who gave a detailed check of my research proposal and helped to revise my research statement. I want to thank Dr. Hager who was not my committee member in the beginning. When he knew that one of my previous committee members moved out from University of Florida, he kindly accepted my invitation to fill the place. I want to thank Dr. Song. Although he is very busy in his own teaching and research, he still accepted my invitation to be my committee member for academic commitment as well as friendship.

I want to especially express my gratitude to my advisor Dr. Yunmei Chen for her patient guidance, frank suggestions and sincere help. During my current PhD study, she has given me a lot of help from study, research to living. Her instruction in research methods will benefit me for my whole life. In preparing for this final defense, she spent lots of time helping me. So, I want to say thanks once more to her.

During my study, Dr. Lei Zhang also gave me a lot of academic help. Meanwhile, many staff members such as Conie Doby, Margaret Somers, Sandy Gagnon and Kristen Cason also provided me with much assistance. I want to express my sincere appreciation to them here. I also want to give my thanks to all my group members. In my previous work, they also gave me a lot of help.

TABLE OF CONTENTS

page

LIST OF TABLES

Table page

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MULTIPHASE IMAGE SEGMENTATION BASED ON INTENSITY STATISTICS: MODELING AND APPLICATIONS

By

Fuhua Chen

August 2012

Chair: Yunmei Chen
Major: Mathematics

Image segmentation is the first stage of image processing and analysis. The result of image segmentation is used for image post-processing. In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments, called phases (or classes). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

An important task of image segmentation is to distinguish objects from background without (or least) affected by noise, intensity-inhomogeneity and artifacts. Based on the classification policy that each pixel is classified to only one phase exclusively, or each pixel can partly belong to more than one phase, image segmentation methods are divided into two categories: hard segmentations and soft segmentations. Variational methods are powerful in hard segmentation. Roughly speaking, variational models for hard segmentations contain region-based methods and edge-based methods. The level-set technique is extensively used in the implementation of these models. However, the non-convexity of the energy functional in the level-set formulation is an inherent drawback of the level-set method. As a result, many level-set-based variational segmentation models are sensitive to initial values. This problem is more difficult to deal with for multiphase image segmentation. Another problem of hard segmentation is its applications to some real world problems. For example, in MRI brain image

segmentation, each voxel (3-D pixel) may contain more than one type of matter due to limited spatial resolution of imaging equipment. The effect is called partial volume. Instead of labeling each image voxel with a unique tissue type [31], partial volume segmentation aims to estimate all the percentage of each voxel that belongs to each tissue-type.

The second problem mentioned above produces the need directly to develop soft segmentation methods, while for the problem of non-convexity of variational hard segmentation models, it has been found that for some of them, the problems can be solved by relaxing the characteristic functions of interested regions to membership functions (also called ownership functions).

In this research, after reviewing the existing image segmentation methods in Chapter 1, we developed three models based mainly on stochastic theory. In all these models, we assume that the intensity of the image at each pixel is a random variable with Gaussian distribution (or mixed Gaussian distribution). Chapter 2 - 5 describe these models. In Chapter 2, we extend the Sine-Sinc model to Gaussian-distribution-like image. Moreover, we choose a normalization of the original image as an initialization of the iterations so that it helps converge to the "true segmentation". Furthermore, we replaced the sinc function by the exponential function. With this change, the new model is more adaptable, and can still be implemented using convex-concave procedure (CCCP) which is guaranteed to converge to a local minimum or saddle point. In Chapter 3, we define a piecewise function $h(x) \in C^1$ to replace the exponential function in the first model and the Sinc function in Sine-Sinc model(discussed in Chapter 2). The advantage of this change lies in the fact that the constructed function has a sum of 1 at each point over all phases. This makes the set of composition functions $\{h_k(x) = h(z(x) - k)\}_{k=1^K}$ be essentially a set of membership functions. Another advantage of this function is that only the nearest neighbor branches can have an overlap of their supports. This property is similar to the partial volume effect in MRI

partial volume segmentation where, approximately, different types of matter, called white matter, gray matter and CSF overlap only at their border. This similarity motivated us to apply our model to partial volume segmentation for MRI brain images. In Chapter 4, we start from considering the piecewise constant Mumford-Shah model for images with intensity-inhomogeneity and develop a stochastic variational soft-segmentation model with mixed Gaussian distribution. The model is more robust to noise and robust to intensity-inhomogeneity too. The problem is formulated as a minimization problem to estimate the mixture coefficients, spatially varying means and variances in the Gaussian mixture. The optimized mixture coefficients lead to a desirable soft segmentation, as well as a hard segmentation. We apply the primal-dual-hybrid-gradient (PDHG) algorithm to our model for iterations of membership functions and use a novel algorithm for explicitly computing the projection from R^K to simplex Δ_{K-1} for any dimension K using dual theory. Our algorithm is more efficient in both coding and implementation than existing projection methods.

Unsupervised image segmentation models are usually efficient only for a specific kind of image. For example, intensity-based unsupervised models usually assumes images to be smooth. It usually fails to work on textured images. Another example is in medical images. When the interested part of some tissue in the image has the same or similar intensity as other tissues, the segmentation will lead to an incorrect result. On the other hand, supervised image-segmentation methods take a learning procedure with a labeled training set to form a classifier. Although supervised methods are likely to give a better result than unsupervised methods, marking the training set is very time-consuming. Semi-supervised segmentation can save the time of machine learning while still utilizing the advantage of unsupervised methods. So far, most of the semi-supervised segmentation methods are developed for two-phase case. Only a few papers have dressed this topic for multiphase segmentation. In Chapter 5, we develop a framework for semi-supervised image segmentations based on the model in Chapter

4. The frame work can be implemented interactively, and can actually be applied to many static image-segmentation models. By using semi-supervised and interactive image segmentation framework developed in this chapter, people can expected more meaningful segmentation results.

All the three models and the semi-supervised frame work are demonstrated with experimental results. In Chapter 6, we give a short prospectus of our anticipated future work related to the aforesaid models and methods.

CHAPTER 1
INTRODUCTION TO IMAGE SEGMENTATION

Image segmentation is at the first stage of image processing and analysis. The result of image segmentation will be used for image post-processing such as feature extraction, pattern recognition, volume analysis and some other statistics and analysis. In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels). Each of the segment or set is called a phase or class. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze [85]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics or some reasonable meaning. In this case each pixel belongs exclusively to exactly one phase. The segmentation is called hard segmentation. In some applications, such as medical image processing, hard segmentation is not enough for post-processing. For example, in Magnetic Resonance Imaging (MRI), a voxel (3D pixel) may contain more than one tissues (white matter, gray matter or cerebrospinal fluid)due to limited resolution, i.e., a voxel partly belongs to more than one phases. The goal of soft segmentation is to find all the probabilities that each pixel belongs to all phases. This probability is also called membership (or ownership) in the literatures. Since soft segmentation allows each pixel to belong to several phases with certain probabilities, it provides a more flexible mechanism, and thereby keeps more options available for post-processing steps.

For convenience of statement, in this whole work, we suppose that the image used for segmentation is always two dimensional one. However, most of the methods we developed can also applied to three dimensional case or can be easily extended to three dimension spaces. Let $\Omega \subset R^2$ be an open and bounded domain with Lipschitz domain. A digital image on Ω is a bounded function $I : \Omega \to R$. Each point $x \in \Omega$ is called a

pixel for two dimensional case and voxel for three dimensional case. The segmentation problem in computer vision (as formulated by Mumford and Shah (1989)) can be defined as follows: given an observed image I, find a decomposition $\Omega_k, (1 \leq k \leq K)$ of Ω and an optimal piecewise smooth approximation f of I, such that the features of f varies smoothly within each Ω_k , and rapidly or discontinuously across the boundaries of Ω_k, where Ω_k is called a segment, class or phase and K is the number of segments of the image we want to find. When $K = 2$, the segmentation means to distinguish or extract an object (one phase) from the image background (another phase), called two phase segmentation. When $K > 2$, the segmentation is called multiphase segmentation.

Based on the notations above, a hard segmentation is to find a partition of Ω, i.e., to find connect sub-domains $\Omega_k, (1 \leq k \leq K)$ of Ω such that (1) $\Omega = \bigcup_{k=1}^{K} \Omega_k$ and $\Omega_i \bigcap \Omega_j = \emptyset$ for $i \neq j$ and (2) in each Ω_k the pixels share some similar features. So each pixel belongs exclusively to one phase, and related to one label k. Different from hard segmentation, in soft segmentation, a pixel $x \in \Omega$ may partly belongs to more than one phases, and is therefore associated to a set of ownerships $p_k(x), 1 \leq k \leq K$(or called memberships) such that $p_k(x) \geq 0$ for all $1 \leq k \leq K$ and $\sum_{k=1}^{K} p_k(x) = 1$ for all $x \in \Omega$. Through all the thesis, except for especially stated, Ω denotes image domain, $p_k(x)$ denotes k-th membership function, and K denotes the number of all phases.

1.1 Some Topics in Image Segmentation

In image segmentation, there are a lot of interesting topics that are still open or need to be further investigated. For instance, how well can we expect to segment images without recognizing objects? What determines a segment? How can we pose the problem mathematically? How do we solve the specified problem(s)? How can we evaluate the results? These questions are actually too large to be answered in detail. In this section, we only list some specific topics that are closely related to this thesis.

1. **Presence of Noise.** One of the most basic topics in image segmentation is how to avoid or minimize the interference of noise during segmentation, namely denoising. Early image segmentation employed data clustering techniques

in the field of pattern recognition which are under discrete setting. One of the most famous data-clustering methods is c-means clustering algorithm using least-square-method. The original c-mean clustering algorithm has no immunity to noise. Later on, the method was extended by adding neighborhood information into the model[9, 39], which is much more robust to noise compared with the original algorithm. Under continuous setting, noises can be easily restrained by introducing variation[83]. One typical example is the well-known Mumford-Shah model [74] which assumes images to be piecewise smooth. ROF model is a representative of variational denoising model [83]. Compared with discrete setting, variational models under continuous setting are theoretically more strict and more beautiful.

2. **Presence of Intensity-Inhomogeneity.** In order to be simple for implementation, many image segmentation models have been developed for piecewise constant images. However, even if in the very ideal case where the actual image is piecewise constant, there may still be intensity inhomogeneity and artifacts presented in the image due to non-uniform illumination, or non-uniform imaging procedure, which can mislead the intensity/brightness-based image segmentation. One way to overcome the intensity-inhomogeneity interference is to utilize spatially non-local information of images [14, 15], called non-local (NL) segmentation. The NL model is capable of integrating semi-local and global image information simultaneously through a specific graph, and thus improve the original model that does not work efficiently with images having local intensity inhomogeneities. Most other approaches to dealing with intensity inhomogeneity focus on bias correction [1, 53, 63, 81, 102]. Such methods are usually integrated into soft segmentation schemes.

3. **Hard-Segmentation versus Soft-Segmentation.** Traditional image segmentation methods are always hard segmentation. Good examples are level-set methods [29, 77] under continuous setting and graph-cut methods [13, 14, 48, 56] under discrete setting (Recently, J. Yuan et al. developed a framework that discussed graph-cut methods under continuous setting [6]). These methods always assume that there exists a boundary separating two spatially neighboring phases. Each point belongs exclusively to only one phase. So, the segmentation is equivalent to find the boundary between different phases. Different from hard segmentation, the soft segmentation [34, 87, 92] is to find all the ownerships (or memberships) of all pixels. Compared with hard segmentation, soft segmentation has many advantages. For example, each pixel in an image may not belong to only one class due to limited resolution. A typical application is the partial volume segmentation for MRI brain images [42, 60], where near the boundary between white matter and gray matter, each pixel usually contains part of white matter and part of gray matter. Simply assigning a pixel to one pure matter may cause a significant accumulated error. In addition, even if in some applications, soft segmentation has to be converted to hard segmentation at the final stage by using the maximum membership classification (In this case, x belongs to i-th phase if

$i = argmax_j\{p_j(x)\}$), it provides a more flexible mechanism, thereby keeping more options available for post-processing steps.

4. **Globally Optimal Solution.** Given a segmentation model, the solution is usually non-unique. Among all the solutions, globally optimal solutions are usually more attractive. From theoretical perspective, a globally optimal solution makes a model more beautiful. From the perspective of implementation, globally optimal solution is always preferred because of their potentially better stability. Under discrete settings, graph-cut methods can guarantee to achieve a global solution[13, 14, 47]. Under continuous setting, it is usually hard to find global solutions for hard-segmentation methods. For instance, it is impossible to obtain global solutions directly from a level-set based method due to the non-convexity of the model. However, by relaxing a hard-segmentation model to a soft-segmentation model, it is possible to find a global solutions for the original hard-segmentation model. For instance, in two-phase case, the active contour model or snake model can be solved globally [16]. For multi-phase case, there have been several papers addressing such topic [5, 17, 18, 52]. However, compared with two-phase case, the implementation is much more complicated and the cost is much more expensive.

5. **Multi-Phase Segmentation versus Two-Phase Segmentation.** The relation between two-phase image segmentation and multi-phase image segmentation is similar to one-variable analysis and multi-variable analysis. As we mentioned at the very beginning, the segmentation of an image $I(x)$ defined in Ω is a process of partitioning the image domain Ω under some rule. In two-phase case, one phase is the object or foreground Ω_F while another phase is the background Ω_B. As long as the foreground is found, the background is also found. Therefore, there is only one variable Ω_F (or Ω_B) in the problem that intrinsically needs to be distinguished. On the other hand, when the problem is a multi-phase segmentation, there are at least two different objects to be determined except for background. Therefore, multiphase image segmentation is more complicated than two-phase image segmentation. Some methods that work very efficient for two-phase image segmentation may not be efficient for multiphase segmentation and some of them cannot be simply extended to multiphase image segmentation. For instance, in level-set based two-phase image segmentation [29], the initial zero-level-set can be randomly chosen. When the level-set method was extended to multiphase image segmentation, however, the initial zero-level-set must be chosen properly in order for the iterations to converge to a feasible solution. Another example is the relaxation of hard segmentation. As mentioned above, the two-phase active contour model can be relaxed to a soft segmentation model so that a global solution of original hard segmentation can be exactly recovered from a global solution of the soft segmentation model. However, when this procedure is applied to multiphase segmentation, it is hard to achieve such a similar result. This is because there are more than one membership functions in the functional. Although

the functional is convex with respect to each membership function $p_i(x)$, it may not be convex with respect to the vectorial function $\vec{p}(x) = (p_1(x), \ldots, p_K(x))$.

6. **Supervised Segmentation versus Unsupervised Segmentation.** Among all image segmentation methods, one can simply divide them into two categories. One is unsupervised methods, another is supervised methods. Fully supervised segmentation models use assigned features as reference for classification. The assigned features are obtained by training a large set of samples. Unsupervised segmentations usually don't assign the features for each class. Only after segmentation, the features of each class can be found. By a trade-off, semi-supervised segmentation models classify pixels by assigning some pixels to each class before or at the very beginning of each implementation of the model. Those assigned pixels will take the role of seed and will be used partly as a kind of reference for the segmentation. Unsupervised methods are theoretically beautiful. However, an unsupervised method cannot deal with all kinds of situations due to the complexity of images. Some prior knowledge of images is sometime necessary for obtaining an ideal segmentation. In some extreme cases, repeated human interactions are needed to take in order to achieve an ideal segmentation.

1.2 Introduction to Existing Techniques

1.2.1 Level-Set Based Segmentations

The task of image segmentation is either to find subregions corresponding to different objects or to find the boundaries of those objects. In recent two decades, one of the most advanced segmentation models is called active contour model or snake model, which is first introduced by M. Kass, A.Witkin, and D. Terzopoulos [51]. The motivation of this model is to find an optimal contour of the object in an image by evolving an initially given contour. Figure 1-1 shows how this model works, where (A) is the original object, (B) shows that the contour moves (shrinks) in when it is out of the object, and (C) shows that the contour moves out (expands) when it is inside the object.

The existing active contour models can be broadly classified as either parametric active contour models or geometric active contour models according to their representation and implementation. In particular, the parametric active contours[51, 104] are represented explicitly as parameterized curves in a Lagrangian framework, while the geometric active contours [22, 23, 66] are represented implicitly as level sets of a two-dimensional function that evolves in an Eulerian framework.

Geometric active contours are independently introduced by Caselles et al. [22] and Malladi et al. [66], respectively. These models are based on curve evolution theory [58] and level-set method [78]. Geometric active contours present several advantages [86] over the traditional parametric active contours. First, the contours represented by the level-set function may break or merge naturally during the evolution, and the topological changes are thus automatically handled. Second, the level-set function always remains a function on a fixed grid, which allows efficient numerical schemes.

Early geometric active contour models [22, 23, 66] are typically derived using a Lagrangian formulation that yields a certain evolution PDE of a parametrized curve. This PDE is then converted to an evolution PDE for a level-set function using the related Eulerian formulation from level-set methods. As an alternative, the evolution PDE of the level-set function can be directly derived from the problem of minimizing a certain energy functional defined on the level-set function. This type of variational methods are known as variational level-set methods [29, 97, 106].

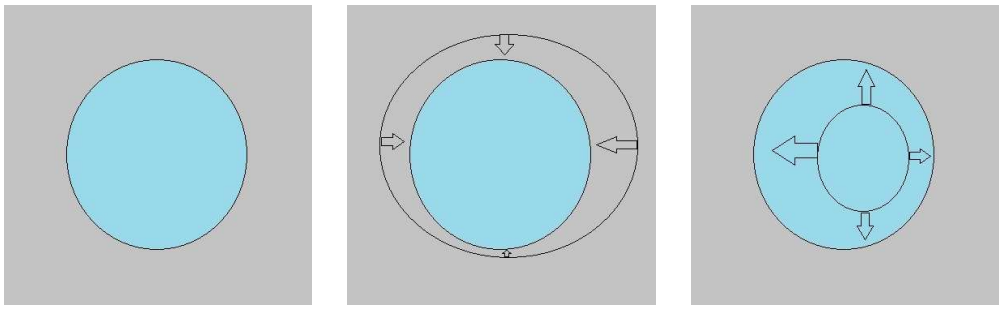


Figure 1-1. Level-set evolution.

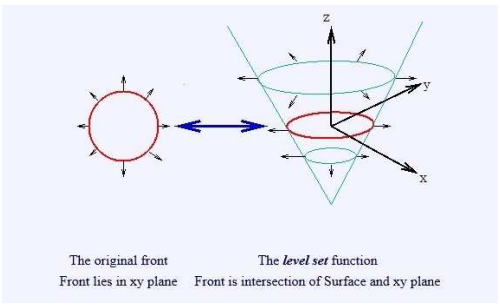


Figure 1-2. Comparison between 2-D evolvement and 3-D evolvement with level-set.

The level-set approach was introduced by Osher and Sethian[76, 77] which is used to model curve evolution implicitly by introducing a higher dimensional function and made the original contour (for example, 2-D contour) a zero-level-set of the higher dimensional function (3-D function). The evolvement of the 2-D contour is then implemented by the evolvement of the higher dimensional surface. As shown in Figure 1-2, the left one shows the evolvement of the curve in 2-D case while the right one shows the evolvement of the surface as well as the evolvement of zero-level-set. The 2-D topological restriction can be therefore solved by such a dimension lifting.

Compared with pure PDE driven level-set methods, the variational level set methods are more convenient and natural for incorporating additional information, such as region-based information [29] and shape-prior information [97], into energy functionals that are directly formulated in the level-set domain, and therefore produce more robust results. For examples, Chan and Vese [29] proposed an active contour model using a variational level-set formulation. By incorporating region-based information into their energy functional as an additional constraint, their model has much larger convergence range and flexible initialization. Vemuri and Chen [97] proposed another variational level-set formulation. By incorporating shape-prior information, their model is able to perform joint image registration and segmentation.

In implementing the traditional level-set methods, it is numerically necessary to keep the evolving level set function close to a signed distance function [78, 80]. Re-initialization, a technique for periodically re-initializing the level set function to a signed distance function during the evolution, was extensively used as a numerical remedy for maintaining stable curve evolution and ensuring usable results. C. Li et al. presented another variational formulation for geometric active contours that forces the level-set function to be close to a signed distance function, and therefore completely eliminates the need of the costly re-initialization procedure [62].

All above mentioned active contour methods are explored for two-phase image segmentations. The extension of level-set method from two-phase to multiphase is not trivial. L. Vese and T. Chan proposed a multiphase level set framework for image segmentation using the Mumford-Shah model, for piecewise constant and piecewise smooth optimal approximations [98]. The proposed method is also a generalization of [29] for two-phase segmentation. The method used $log\ n$ level set functions for n phases in the piecewise constant case. G. Chung and L. Vese applied level set method to multiphase image segmentation in another way for piecewise constant segmentation of images [36]. They represented the set of boundaries of the segmentation implicitly using a multilayer of level lines of a continuous function. In the standard approach of front propagation, only one level line is used to represent the boundary. Later on, X.-C. Tai, et al., proposed a variant of the level set formulation for identifying curves separating regions into different phases. Compared with aforesaid multiphase level-set methods, the novelty in this method is to introduce a piecewise-constant-level-set function and use each constant value to represent an unique phase. If 2^n phases should be identified, the level-set function must approach 2^n predetermined constants. The method also only needs one level set function to represent 2^n unique phases. However, the re-initialization procedure required in classical level set methods is superfluous using such approach. The minimization functional is convex and differentiable and thus avoid some of the problems with the non-differentiability of the heaviside functions.

1.2.2 Graph Based Methods

Recall that a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a finite set $\mathcal{V}$ of vertices (or nodes) and a set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ of edges. An edge $e := (u, v) \in \mathcal{E}$ is said to be from vertex u to vertex v. An undirected graph is one in which all edges go both ways: $(u, v) \in \mathcal{E}$ iff $(v, u) \in \mathcal{E}$.

The idea of graph based image segmentation is to represent an image I by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ being the set of pixels. For each edge $e = (u, v) \in \mathcal{E}$, define a cost $c(e)$ to reflect the similarity between the linked vertices. The more similar

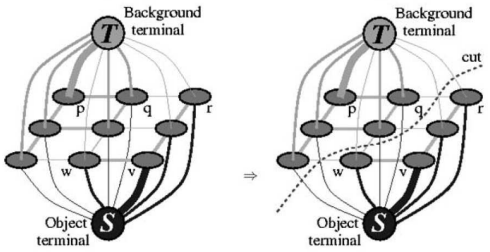


Figure 1-3. Minimum cut segmentation.

two pixels are, the larger the cost is. Moreover, in order to apply graph theory to image segmentation, two special nodes are added to the graph, i.e. s (the source: object) and t (the sink: background), called terminal nodes, and an edge from each terminal to each node in $\mathcal{V}$ is also added. The new formulated graph is denoted by a triple $(\mathcal{G}, s, t)$ and called an s-t graph. An s-t cut is a subset of edges $C \subset \mathcal{E}$ such that the terminals S and T become completely separated on the induced graph $\mathcal{G}(C) = (\mathcal{V}, \mathcal{E} \setminus C)$. Then for a given cut, the total cost of the cut is defined as

$$\sum_{e \in C} c(e) \tag{1–1}$$

The task of an image segmentation is therefore to find a cut with minimum total cost, called minimum cut. See Figure 1-3, where a pixel having a thick edge with S means that the pixel most probably belongs to the object, and a pixel having a thick edge with T means that the pixel most probably belongs to the background.

Consider a graph $G = (\mathcal{V}, \mathcal{E})$ with a function c defined on $\mathcal{V} \times \mathcal{V}$ such that $c(u, v) = 0$ if $(u, v) \notin \mathcal{E}$. Such a function is called a capacity function of graph $\mathcal{G}$. Let the triple (G, s, t) be the s-t graph, still denoted by $(\mathcal{V}, \mathcal{E})$. A **flow** with respect to the s-t graph is a function $f : \mathcal{V} \times \mathcal{V} \to R$ such that:

- $f(u, v) \leq c(u, v)$ for all $u, v \in \mathcal{V}$;
- $f(u, v) = -f(v, u)$ for all $u, v \in \mathcal{V}$;
- $\sum_{v \in V} f(u, v) = 0 \; \forall \; u \in \mathcal{V} \setminus \{s, t\}$

The value $|f|$ of a flow f is defined as $|f| = \sum_{v \in \mathcal{V}} f(s, v)$. A max-flow problem is to find a flow f with maximum value $|f|$.

In 1993, Z. Wu and R. Leahy applied graph-theory based data-clustering method to image segmentation [103]. Clustering is achieved by removing edges of G to form mutually exclusive subgraphs such that the largest inter-subgraph maximum flow is minimized. The method was implemented using combinatorial optimization algorithms. A globally optimal segmentation can be computed efficiently in low-order polynomial time using max-flow/min-cut algorithms on graphs [37, 44, 46]. Such a method was then extensively studied by many other people [43, 55, 96]. J. Shi and J. Malik used a different way called normalized cuts for image segmentation [89]. The key idea of such methods is a completely automatic high-level grouping of image pixels. Typically, this means that they divide an image into blobs or clusters using only generic cues of coherence or affinity between pixels. Y. Boykov and M.-P. Jolly [12] are first ones that demonstrated how to use binary graph cuts to build efficient object extraction tools for N-dimensional applications for any positive integer N based on a wide range of model-specific (boundary and region-based) visual cues, contextual information, and useful topological constraints. In 2006, Boykov et al. showed a very strong connection between graph-cuts and level-sets [14]. In particular, they developed a novel integral approach to solving surface propagation PDEs based on combinatorial graph-cut algorithms.

The max-flow min-cut theorem (by Ford and Fulkerson) states that a maximum flow from s to t saturates a set of edges in the graph dividing the nodes into two disjoint parts $\{S, T\}$ corresponding to a minimum cut. Thus, min-cut and max-flow problems are equivalent. In fact, the maximum flow value is equal to the cost of the minimum cut.

In 2003, H. Ishikawa extended graph-cut methods to multiphase image segmentation using Markov Random Field (MRF) [56]. He introduced a method to solve exactly a first order Markov Random Field optimization problem with the MRF having a prior term

that is convex in terms of a linearly ordered label set. The method maps the problem into a min-cut problem for a directed graph, for which a globally optimal solution can be found in polynomial time. The convexity of the prior function in the energy is shown to be necessary and sufficient for the applicability of the method.

In 2010, E. Bae et al. extended the graph-cut method under continuous setting using dual method. They proposed a continuous analogue of Ishikawa's graph construction [56] by formulating continuous max-flow and min-cut models over a specially designed domain. These max-flow and min-cut models are equivalent under a primal-dual perspective, which can be seen as exact convex relaxations of the original problem and can be used to compute global solutions.

1.2.3 Supervised Segmentations

Unsupervised methods explore the intrinsic data structure to segment an image into regions with different statistics. The segmentation procedure can be implemented using some assigned algorithm automatically without human beings' interaction or interfering. Different from unsupervised segmentation methods, supervised image segmentation is a technique that classifies images using some assigned features for each class. These features usually obtained by machine learning due to the complexity of images. When image features are simple and be able to be distinguished easily, supervised methods are not really necessary. However, when the image is much complicated and hard to be classified, unsupervised methods often fail to achieve the desired result, especially if a desired segmentation includes regions with very different characteristics. On the other hand, supervised image segmentation methods take a learning procedure with a labeled training set to form a classifier. Although supervised methods are likely to give a better result than unsupervised methods, marking the training set is very time consuming.

The terms ”supervised” or ”unsupervised” comes from machine learning in computer vision. One typical example of unsupervised method is k-mean clustering. By only using image statistics, clustering algorithms partition the image in coherent

groups without using labeled information. Therefore, they are appealing because they avoid users' supervision and ground surveys to define the classes of interest. However, clustering algorithms are much more difficult to calibrate than supervised methods, especially because the final aggregation is still need to be decided by the user, either by finding an appropriate number of clusters or by deciding the level of pruning of a hierarchical similarity tree [59]. For both cases, few robust criteria are available and the user typically turns to heuristic solutions. Moreover, the outputs of clustering methods are often difficult to interpret since there is no explicit link between the clusters found by the model and the classes desired by the user. The former describes data similarities and the latter are semantic interpretations of the objects of interest. Thus, the resulting clusters may represent mixed semantic properties of the scene, harming the labeling of the final classification map. Most of the model based segmentation methods mentioned above belong to unsupervised methods.

Semi-supervised methods take a trade-off between supervised methods and unsupervised methods by inferring the classification from partially labeled data. The key difference between supervised learning and semi-supervised learning is that semi-supervised methods utilize the data structure in both the labeled and unlabeled data points [27]. Hence, the main advantage of semi-supervised image segmentation methods is that they take advantage of the user markings to direct the segmentation, while minimizing the need for user labeling. There are several general approaches towards semi-supervised learning, but recent developments have focused on graph-based methods [27], probably because the graph-based representation naturally copes with nonlinear data manifolds. In this formulation, data are represented by nodes in a graph, and the edge weights are given by some measure of distance or affinity between the data. Then, the labels for the unlabeled points are found by propagating the labels of labeled points through the graph. Based on this methodology, a number of methods have been proposed [7, 50, 100, 107, 110, 111]. However, these

methods are all for general data classification. There is no one of these papers being applied to general image segmentations, probably because of their discrete settings.

Although in some papers, authors didn't strictly distinguish semi-supervised methods and supervised methods, strictly supervised segmentation methods are actually quite different from semi-supervised segmentation methods. Generally speaking, the supervised segmentation means to set up a learning machine before segmentation is carried out. The learning process is performed at a large training set of the similar kind of data. Therefore, a strict supervised segmentation model is usually designed for a specific kind of image segmentation, such as cell segmentations, spine-segmentations, prostate segmentation, and so on. The learning procedure is usually carried out before all such kind of segmentations are performed. As soon as the learning procedure is finished, the features obtained will be used for all such kind of segmentations. Different from supervised segmentation, semi-supervised segmentation methods or interactive segmentation methods are carried out by interfering segmentation each time before an automatic segmentation procedure is performed.

Except for discrete settings, supervised image segmentation technique has also been embedded in continuous models. G. Gilboa and S. Osher [45] proposed a supervised segmentation model based on non-local information. N. Houhou et al. proposed a semi-supervised image segmentation method that relies on a non-local continuous version of the min-cut algorithm and labels or seeds provided by a user [50]. The segmentation process is performed via energy minimization. The proposed energy is composed of three terms. The first term defines labels or seed points assigned to objects that the user wants to identify. The second term carries out the diffusion of object and background labels and stops the diffusion when the interface between the object and the background is reached. The diffusion process is performed on a graph defined from image intensity patches. The graph of intensity patches is known to better

deal with textures because this graph uses semi-local and non-local image information. The last term is the standard total variation (TV) term that regularizes the geometry of the interface.

Image matting is another class of supervised segmentation. In an image matting, a foreground element is extracted from an image by estimating a color and opacity for the foreground element at each pixel. The opacity value at each pixel is typically called its alpha, and the opacity image, taken as a whole, is referred to as the alpha matte (between 0 and 1) or key. Matting is used in order to composite the foreground element into a new scene. Matting and composition were originally developed for film and video production. The classic image matting methods contain Poisson Matting [90], Bayesian Matting [35] and Spectral Matting [61]. More precisely, image matting is a semi-supervised soft image segmentation. The opacity alpha is actually the membership of the foreground. The method assigns some pixels to be foreground and some other pixels to be background by heuristic. The classification is then propagated by some algorithm.

Among these semi-supervised segmentations mentioned above, all are for two-phase image segmentation except for [45].

1.2.4 Soft Segmentations

Soft segmentation is motivated by insufficiency of hard segmentation. It is well known that MRI brain images contain white matter, gray matter and cerebrospinal fluid (CSF). Accurately estimating the volumes of different matters and finding their change can help diagnose some brain diseases. However, due to limited resolution of imaging equipments, a voxel may contain more than one matters, especially near the border of two different matters. Simply classifying a voxel into one matter may cause a large accumulated error. The better way for MRI brain image segmentation is to find all the percentages of a voxel belonging to all matters. Corresponding to soft segmentation,

this means to find all the membership functions. This kind of brain image segmentation is called partial volume (PV) segmentation.

Soft segmentation is also motivated by non-convexity of some hard segmentation models. One of the extensively studied approaches for hard segmentation is the variational method. Many effective variational models have been developed, for instance, the Mumford-Shah model [74], geodesic active contour [23], geodesic active region [79], and region competition [109]. Level-set technique [76] has been proved to be powerful in the implementation of variational models. In two-phase segmentation the composition of the heaviside function with the level set function is used to represent the regions of the object and background. In [98, 106], the authors extended the level-set method to multiphase segmentation by using multiple level-set functions, while in [36, 65], the authors proposed another way to extend the level-set method by using multiple layers for each level-set function. With carefully choosing the initial values, these methods can work very well. However, the non-convexity of the energy functional in the level-set formulation is an inherent drawback. As a result, many level-set based variational segmentation models are sensitive to initial values and may converge to an undesirable local minimum. This problem is more difficult to deal with for multiphase segmentation.

To overcome the non-convexity problem mentioned above, one approach is to replace the composition of the heaviside function with the level-set function in level-set formulation by a weight/membership function (or more generally replace the characteristic functions with membership functions in region-based models). This relaxation provides a possibility to make the energy convex with respect to membership functions and so convenient to find a global minimizer. For example, Chan et al [30] and Bresson et al [10] restated certain non-convex minimization problems for image segmentation and denosing as equivalent convex minimization problems by using membership functions to replace characteristic functions. These new models allow to

find global minimizers via standard convex minimization schemes. In particular in [10] efficient and fast numerical schemes to globally minimize the variational segmentation models were proposed. These algorithms are based on a dual formulation of the TV (total variation) norm proposed and developed in [4, 21, 25, 26, 28].

Some approaches of soft segmentation are directly extended from the corresponding hard segmentation methods. For example, fuzzy region-competition method [71, 72] is extended from region-competition method [109]; fuzzy c-mean and adaptive fuzzy c-mean method are developed from c-mean clustering method; soft Mumford-Shah model[87] is developed from Mumford-Shah model [74], et al. There are also other approaches of soft segmentation that are not simply derived from hard segmentation methods, such as maximum-likelihood (ML) [38], maximum-a-posteriori probability method (MAP) [64, 102], Markov-random-field (MRF) [64] method and other stochastic methods [57, 87].

Mory and Ardon extended the original region-competition model to a fuzzy region-competition method [71, 72]. The technique generalizes some existing supervised and unsupervised region-based model. The proposed functional is convex, which guarantees a global solution in the supervised case. Unfortunately, this method only applies to two-phase segmentation and is hard to be extended to multiphase segmentation. Fuzzy C-mean (FCM) [31, 64, 81] is a method developed for pattern classification and pattern recognition. It is also applicable to image segmentation. The standard FCM model partitions a data set $\{x_k\}_{k=1}^N \subset R^d$ into M clusters by the following objective function [9, 39]

$$J_{FCM} = \sum_{i=1}^{N}\sum_{k=1}^{M} u_{ik}\|x_i - v_k\|_2^2 \tag{1–2}$$

where u_{ik} is the membership value of datum x_i for class k with $\sum_{k=1}^{M} u_{ik} = 1$, and v_k stands for the cluster centers. The original FCM method is very sensitive to noise. An adaptive fuzzy c-means (AFCM) was proposed by Pham et.al [81], where the constant cluster centers v_k used in the FCM model are substituted by spatially varying functions.

The energy functional can be written as

$$J_{AFCM} = \sum_{i=1}^{N}\sum_{k=1}^{M} u_{ik}\|x_i - b_i v_k\|_2^2 + R(b) \tag{1–3}$$

where b is the bias field and $R(b)$ is the regularity term for the bias field b. AFCM is more robust to noise than the standard FCM. The soft segmentation model developed in [31] used a different similarity measure from that in [81]. Their objective functional reads as

$$J_{AFCM} = \sum_{i=1}^{N}\sum_{k=1}^{M} u_{ik}(1 - K_\sigma(x_i - v_k)) + \frac{\alpha}{|N_k|}\sum_{i=1}^{N}\sum_{k=1}^{M} u_{ik}\sum_{r\in N_k}(1 - K_\sigma(x_r - v_k)) \tag{1–4}$$

where

$$K_\sigma(x, y) = exp(\frac{-\|x - y\|}{\sigma^2}). \tag{1–5}$$

Another class of soft segmentations are based on stochastic approaches [64, 87]. In these approaches the intensity of pixels are assumed to be random variables which are either independent or somehow dependent. Let $I(x)$ be an image defined on an open bounded domain Ω containing N classes. Let w be the class label variable, $w = 1, \cdots, N$. At each pixel $x \in \Omega$, both $w(x) \in \{1, ..., N\}$ and $I(x)$ are viewed as random variables indexed by x. The probability that x belongs to the i-th phase is represented by the ownership functions $p_i(x)$, $1 \leq i \leq N$. Denote by $Prob(I(x)|w(x) = i)$ the probability density function (pdf) of the random variable $I(x)$ belonging to the i-th pattern. Then the pdf of the image $I(x)$ at each $x \in \Omega$ is a mixed distribution given by

$$\sum_{i=1}^{N} Prob(I(x)|w(x) = i)p_i(x). \tag{1–6}$$

The optimal segmentation of the image is deduced to maximize the likelihood, the joint pdf of $\{I(x) : x \in \Omega\}$. With these analysis, maximum-likelihood (ML) [38] and maximum-a-posteriori (MAP) [64, 102] principle based techniques have been widely used in soft segmentation. The ML methods find the optimal parameters in the joint pdf such that the likelihood function is maximized [38]. However, simply using likelihood to

model an image is not enough since it ignored the prior knowledge of an image. In [64], a segmentation framework based on MAP principle was proposed for partial volume (PV) segmentation of MRI brain images. A mixture of the probability density functions is considered to address the PV effect. A Markov Random Field (MRF) model is used to define the prior distribution of the mixture coefficient field imposing a smoothness on the mixture coefficients (ownerships). The fuzzy c-mean model is extended to define the likelihood function of the observed image. J. Shen [87] proposed a general multiphase stochastic variational fuzzy segmentation model combining stochastic principle. The intensity of images was modeled as a mixed Gaussian distribution. The assumption in the model that membership functions should be either close to 1 or close to 0 simplified the model itself but limited its application. For example, it's not reasonable to apply the model to partial volume segmentation since in that case the membership functions usually evaluated neither close to 1 nor close to 0 at the border of different matters.

Another feature of Shen's paper is that it utilized Modica-Mortola's phase-transition theory. The similarities between image segmentation and phase transition theory in material sciences and fluid mechanics have inspired people to borrow some ideas in contemporary material sciences, e.g., the diffuse interface model of Cahn-Hilliard [20], and its rigorous mathematical analysis in the framework of Γ-convergence approximation by Modica and Mortola [70, 73]. The authors in [57] presented a model for image segmentation based upon the phase transition theory of Modica and Mortola and discussed its connections to the Mumford-Shah segmentation model and some related works.

Another way to improve the implementation efficiency of variational models is to reduce a model to a piecewise constant one [29]. Under piecewise constant assumption, the implementation can be simplified since the total variation of an object is exactly its perimeter. Piecewise constant assumption is reasonable in many cases. However, due to non-uniform illumination, or non-uniform imaging

procedure, intensity-inhomogeneity and artifacts can be produced, which will mislead the intensity based segmentation. One way to overcome the intensity-inhomogeneity is to utilize spatially non-local information of the image [14, 15], called non-local (NL) segmentation. The NL model is able to integrate simultaneously semi-local and global image information through a specific graph, and thus improve the original model that does not work with images with local intensity inhomogeneities.

Most other approaches dealing with intensity inhomogeneity focus on bias correction [1, 53, 63, 81, 102]. These methods are usually integrated into soft segmentation schemes. For example, Wells et al. proposed an expectation-maximization (EM) algorithm to solve the bias correction problem and the tissue classification problem [102]. The EM algorithm was used to iteratively estimate the posterior tissue class probabilities when the bias field is known, and to estimate the MAP of the bias field when tissue class probabilities are known. The disadvantage of this method is that the directly computed bias field is not smooth which will lead to a poor bias correction and segmentation results. Pham et al. in their AFCM method replaced the constant cluster centers by spatially varying functions [81], which are product of bias field and the constant clustering centers. Smoothness of the bias field is ensured by penalizing its first and second order derivatives, which leads to a computationally expensive procedure for the smoothing of the bias field. Ahemd et al. proposed to add a neighborhood term that enabled the class membership of a pixel to be influenced by its neighbors [1]. The neighborhood effect acts as a regularizer and forces the solution toward a piecewise homogeneous labeling. This approach proved tolerant to salt and pepper noise, resulting in smoother segmentation. Li et al. proposed a variational level-set based method for medical image segmentation and bias correction [53], the smoothness of the bias field is intrinsically ensured by the data term in the variational formulation, but the scheme is computationally expensive. F. Li et al. proposed a variational fuzzy Mumford-Shah model for multi-phase segmentation [63]. The model is based on the

assumption that an image can be approximated by the product of a smooth function and a piecewise constant function. The energy functional is as follow.

$$E(p, b, c) = \sum_{i=1}^{K} \lambda \int_{\Omega} \left((I(x) - b(x)c_i)^2 + w \int_{\Omega} (I(x) - c_i)^2 \right) p_i dx + \mu R(b) + R(p) \quad (1\text{–}7)$$

where I is the given image, b is the bias field, c_i is the ideal mean of i-th class, p_i is the i-th class membership function, $R(b)$ and $R(p)$ are regularization terms for bias field b and membership functions respectively, and λ and μ are parameters used to balance the weight of fitting term penalty and regularity term penalty. Since the model is convex in membership functions, the iterations are guaranteed to converge for arbitrary initial value. In [64], a segmentation framework based on MAP principle was proposed for partial volume (PV) segmentation of MRI brain images. The model is developed from a stochastic point of view. The final energy functional is in fact equivalent to Paper [63] with a discrete form.

1.3 Our work

In this research, after reviewing the existing image segmentation methods in Chapter 1, we developed three models based mainly on stochastic theory. In all these models, we assume that the intensity of the image at each pixel is a random variable with Gaussian distribution (or mixed Gaussian distribution). In Chapter 2, we extend the Sine-Sinc model to Gaussian-distribution-like image. Moreover, we choose a normalization of the original image as an initialization of the iterations so that it helps converge to the “true segmentation”. Furthermore, we replaced the sinc function by the exponential function. With this change, the new model is more adaptable, and can still be implemented using convex-concave procedure (CCCP) which is guaranteed to converge to a local minimum or saddle point. In Chapter 3, we defined a piecewise function $h(x) \in C^1$ to replace the exponential function in the first model and the Sinc function in Sine-Sinc model(discussed in Chapter 2). The advantage of this

change lies in the fact that the constructed function has a sum of 1 at each point over all phases. This makes the set of composition functions $\{h_k(x) = h(z(x) - k)\}_{k=1^K}$ be essentially membership functions. Another advantage of this function is that only the nearest neighbor branches can have an overlap of their supports. This property is similar to the partial volume effect in MRI partial volume segmentation where approximately, different matters only overlap at their border, called white-gray matter and gray-CSF matter. This similarity motivated us to apply the model to partial volume segmentation for MRI brain images. In Chapter 4, we started from considering the piecewise constant Mumford-Shah model for images with intensity-inhomogeneity and developed a stochastic variational soft segmentation model with mixed Gaussian distribution. The model is more roust to noise and robust to intensity-inhomogeneity too. The problem was formulated as a minimization problem to estimate the mixture coefficients, spatially varying means and variances in the Gaussian mixture. The optimized mixture coefficients lead to a desirable soft segmentation, as well as a hard segmentation. We applied the primal-dual-hybrid-gradient (PDHG) algorithm to our model for iterations of membership functions and used a novel algorithm for explicitly computing the projection from R^K to simplex Δ_{K-1} for any dimension K using dual theory, which is more efficient in both coding and implementation than existing projection methods.

Unsupervised image segmentation models are usually only efficient for a specific kind of images. For example, intensity-based unsupervised models usually assumes images to be smooth. It usually fails to work on texture images. Another example is in medical images. When the expected part of some tissue in the image has a same or similar intensity with other tissues, the segmentation will lead to an incorrect result. On the other hand, supervised image segmentation methods take a learning procedure with a labeled training set to form a classifier. Although supervised methods are likely to give a better result than unsupervised methods, marking the training set is very time

consuming. By trade-off, semi-supervised segmentation can save the time of machine learning while still utilizing the advantage of unsupervised methods. So far, most of the semi-segmentation methods are developed for two-phase case. There are few papers dressed this topic for multiphase segmentation. In Chapter 5, we developed a frame work for semi-supervised image segmentations based on the model we developed in Chapter 4. The frame work can be implemented interactively, and can actually be applied to many static image segmentation models. By using semi-supervised and interactive image segmentation frame work developed in this chapter, people can obtain any desired segmentation results.

All the three models and the semi-supervised frame work are demonstrated with experimental results. In Chapter 6, we give a short prospect on our future work related to the aforesaid models and methods.

CHAPTER 2
MULTI-PHASE IMAGE SEGMENTATION BASED ON PHASE-TRANSITION THEORY

The similarities between image segmentation and phase transition theory in material sciences and fluid mechanics have inspired people to borrow some ideas in contemporary material sciences, e.g., the diffuse interface model of Cahn-Hilliard [20], and its rigorous mathematical analysis in the framework of Γ-convergence approximation by Modica and Mortola [70, 73]. The phase-field relaxation consists in approximating the perimeter using a Cahn-Hilliard type penalization functional [20], with the form

$$P_\epsilon(\upsilon) = \frac{\epsilon}{2}\int_\Omega |\nabla \upsilon|^2 dx + \frac{1}{\epsilon}\int_\Omega W(\upsilon)dx, \tag{2–1}$$

where $W : \mathbb{R}^p \to \mathbb{R} \cup +\infty$ is a scalar function with exactly two minimizers at 0 and 1 satisfying $W(0) = W(1) = 0$. The second term of the penalty functional ensures that the values of the material density υ converges to 0 or 1 as $\epsilon \to 0$, while the first term controls the perimeter. The parameter ϵ can be interpreted as the width of the diffused edge representation in υ. The phase-field approach has been used in topological optimization problems [11, 19, 101]. In [84], the authors used the phase field to approximate sharp edges and a variational phase field model is derived to compute a shape average of a given number of shapes. In [19], the authors used the phase transition theory in a Cahn-Hilliard impainting model.

In paper [87], J. Shen proposed a general multiphase stochastic variational fuzzy segmentation model combining stochastic principle and the Modica-Mortola's phase-transition theory. Applying phase-transition theory into the model can enhance pattern separation and make boundaries smooth. The stochastic variables are used to represent the ownerships of all classes. The regularization is made using a double well potential borrowed from the phase-transition theory. By assuming that all patterns are Gaussian distributions with mean fields u_i $(i = 1, \ldots N)$, and a fixed variance σ^2, the pdf

of the mixed Gaussian is given by

$$Prob(I(x)|P(x), U(x)) = \sum_{i=1}^{N} g(I|u_i(x), \sigma)p_i(x), \tag{2–2}$$

where

$$g(I|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(I-\mu)^2}{2\sigma^2}) \tag{2–3}$$

defines the Gaussian probability density function. The model is to solve the following minimization problem:

$$\begin{aligned} \min E_S(P, U) =& \lambda \sum_{i=1}^{N} \int_{\Omega} (u_i - I)^2 p_i + \alpha \sum_{i=1}^{N} \int_{\Omega} |\nabla u_i|^2 \\ &+ \sum_{i=1}^{N} \int_{\Omega} \left(9\epsilon |\nabla p_i|^2 + \frac{(p_i\,(1-p_i))^2}{\varepsilon} \right) \end{aligned} \tag{2–4}$$

with constraints

$$0 \leq p_i \leq 1 \ \text{ and } \ \sum_{i=1}^{N} p_i(x) = 1, \tag{2–5}$$

where $p_i(x)$ are the ownerships and $u_i(x)$ are called patterns. Unlike the original Mumford Shah model, the energy of each channel is defined on the entire domain Ω instead of on a specific subregion Ω_i.

The authors in [57] presented another model for image segmentation based upon phase-transition theory and discussed its connections to the Mumford-Shah segmentation model and some related works. The method uses a sine-sinc model which can also be thought to be motivated by the Mumford-Shah model. However, instead of using a target image $u(x)$, the new model uses a signature function $z(x)$ that is supposed to take value $0, 1, \ldots, K-1$ (suppose there are totally K phases in the image). With the theory of Γ-convergence and the convex-concave procedure (CCCP) [8, 20, 105], the iteration scheme can guarantee converge to a local minimum or a saddle point.

Despite of the success of the method, there are still some shortcomings in it. First, it supposed the image to be piecewise constant. This is not the general case in real applications. For instance, objects and background may not behavior to be intensity constant but show some distribution property. Second, based on the theorem on convex-concave procedure (CCCP), although it can guarantee the iteration scheme converge to a local minimum or saddle point, it can not guarantee the iteration to converge to a required segmentation. For example, during an implementation, it may sometime converge to two phases when three or more phases are expected.

On the first point, we extended the model to be suitable for those images having Gaussian distribution intensities, which is somehow motivated by the idea of soft-segmentation [88]. On the second point, unlike in paper [87] where the author recommend to use random value as the initialization of the target function, we choose the initialization of the signature function $z(x)$ based on the original image(see Section 3). The advantage of our method is that in most cases, it can push the iterations to converge exactly to the required number of phases. The rest of this chapter is organized as follows. In Section 2.1, we first review on the sine-sinc model. Then develop our improved model. Section 2.2 is the implementation and some considerations, where we choose a normalization of the original image as the initialization of the signature function $z(x)$. In Section 2.3, we exhibit experiment results for different kind of images, synthetic image and human brain image. We especially take a comparison between our model and the sine-sinc model. Finally, the chapter is closed with a short conclusion.

2.1 Model Development

In this section, we first take a short review on the sine-sinc model. Then we develop our new model in two steps. At the first step, we replace $sinc(x)$ by $e^{-\alpha x^2}$ $(\alpha > 0)$. At the second step, we extend the model to be suitable for those images obeying Gaussian distribution.

2.1.1 Review on Sine-Sinc Model

Let Ω be a bounded Lipschitz domain, and $u_0 : \Omega \to \mathcal{R}$ be a given image. Suppose the image contains K phases which correspond to regions $\Omega_0, \Omega_1, ...\Omega_{K-1}$. Define a signature function z by

$$z(x) = k, \text{if } x \in \Omega_k, k = 0, 1, ..., K-1. \tag{2–6}$$

The model is to minimize the following energy function

$$E_\epsilon[z|u_0] = \int_\Omega [\epsilon|\nabla z|^2 + \frac{1}{\epsilon} sin^2 \pi z] dx + \lambda \sum_{k=0}^{K-1} \int_\Omega (c_k - u_0)^2 sinc^2(z-k)\, dx, \tag{2–7}$$

where ϵ is a parameter and sinc function is defined by

$$sinc(x) = \frac{sin\pi x}{\pi x}. \tag{2–8}$$

To be consistent with the paper, we also use $F_\epsilon[z]$ to denote the first part of (2–7), and use $G[z|u_0]$ to denote the second part of (2–7), i.e.,

$$F_\epsilon[z] = \int_\Omega [\epsilon|\nabla z|^2 + \frac{1}{\epsilon} sin^2 \pi z] dx \tag{2–9}$$

and

$$G[z|u_0] = \sum_{k=0}^{K-1} \int_\Omega |c_k - u_0|^2 sinc^2(z-k)\, dx. \tag{2–10}$$

Now, we have

$$E_\epsilon[z|u_0] = F_\epsilon[z] + \lambda G[z|u_0]. \tag{2–11}$$

In paper [69], Modica and Mortola established that

$$\int_\Omega [\epsilon|\nabla z|^2 + \frac{1}{\epsilon} sin^2 \pi z] dx \tag{2–12}$$

γ-converges to $\frac{4}{\pi}\int_\Omega |Dz(x)|$ in $L^1(\mathcal{R})$ for phase fields that ultimately only take integer values. It is easy to see that

$$\sum_{k=0}^{K-1} |\partial\Omega_k| \leq \int_\Omega |Dz(x)| \leq K \sum_{k=0}^{K-1} |\partial\Omega_k|. \tag{2–13}$$

Thus, $F_\epsilon[z]$ takes a role similar to the length of the edge of all phases when ϵ being small enough. On the other hand, if we replace $sinc(z-k)$ by 1_{Ω_k}, the characteristic function of Ω_k, then $G[z|u_0]$ is a difference of the original image $u_0(x)$ and the piecewise constant image $u(x)$ defined by

$$u(x) = c_k \text{ if } x \in \Omega_k. \tag{2–14}$$

So, $G[z|u_0]$ in Equation (2–7) is an approximation of the difference of the original image $u_0(x)$ and the piecewise constant image $u(x)$. Since the energy function is non-convex in z, the author adopted the convex-concave procedure (CCCP) [105]. We have a short review on this procedure in Section 3.

2.1.2 New Model

To be convenient for statement, from now on, we call the aforesaid model Model 1. As we talked in Section 1, Model 1 assumed the image to be piecewise constant. In this section, we extend the model in two steps. At the first step, we replace $sinc(x)$ by $e^{-\alpha x^2}$, where α is a positive parameter that can be chosen elastically. We do so for two considerations. First, we think that the exponential function $e^{-\alpha x}$ would be better in approximating a characteristic function than sinc function when we adjust the parameter α; second, as we will see in the implementation, for our new model, it is then easier to estimate the upper bound of $G(z|u_0)$ so that we can still apply the CCCP algorithm (see Section 3). After this change, the energy function becomes

$$E_\epsilon[z|u_0] = \int_\Omega [\epsilon|\nabla z|^2 + \frac{1}{\epsilon} sin^2 \pi z] dx + \lambda \sum_{k=0}^{K-1} \int_\Omega |c_k - u_0|^2 e^{-\alpha(z-k)^2} dx. \tag{2–15}$$

We call this model Model 2.

At the second step, we assume that, the intensity $u_0(x)$ at each pixel x is a random variable which obeys Gaussian distribution, and for each area Ω_k $(k = 1, 2, ..., K-1)$, the set of random variables $\{u_0(x)|x \in \Omega_k\}$ share the same mean c_k and the same variance σ_k^2. Then the pdf of $u_0(x)$ for fixed x in area Ω_k is denoted by

$$P(u_0(x)|c_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(u_0(x)-c_k)^2}{2\sigma_k^2}} (k = 0, 1, ..., K-1). \tag{2–16}$$

In further, let $c = (c_0, c_1, ..., c_{K-1})$ and $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{K-1})$. We want to maximize the likelihood, joint pdf of $\{u_0(x), x \in \Omega\}$

$$\begin{aligned} L(c, \sigma) = P(\{u_0(x), x \in \Omega\}|c, \sigma) &= \prod_{k=0}^{K-1} P(\{u_0(x)|x \in \Omega_k\}|c_k, \sigma_k) \\ = \prod_{k=0}^{K-1} \prod_{x \in \Omega_k} P(u_0(x)|c_k, \sigma_k) &= \prod_{k=0}^{K-1} \prod_{x \in \Omega_k} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(u_0(x)-c_k)^2}{2\sigma_k^2}} \end{aligned} \tag{2–17}$$

or equivalently, to minimize the negative log-likelihood

$$\begin{aligned} -logL(c, \sigma) &= \sum_{k=0}^{K-1} \int_{\Omega_k} (log(\sqrt{2\pi}\sigma_k) + \frac{(c_k - u_0)^2}{2\sigma_k^2}) \\ &= \sum_{k=0}^{K-1} \int_{\Omega} (log(\sqrt{2\pi}\sigma_k) + \frac{(c_k - u_0)^2}{2\sigma_k^2}) 1_{\Omega_k} \end{aligned} \tag{2–18}$$

where 1_{Ω_k} is the characteristic function of Ω_k. Similar to step 1, we now use $e^{-\alpha(z(x)-k)^2}$ to approximate the characteristic function 1_{Ω_k}. Then Equation (2–18) becomes the following form

$$\sum_{k=0}^{K-1} \int_{\Omega} (log(\sqrt{2\pi}\sigma_k) + \frac{(c_k - u_0)^2}{2\sigma_k^2}) e^{-\alpha(z(x)-k)^2} \tag{2–19}$$

We use this expression to replace the second part $G[z|u_0]$ of $E_\epsilon[z|u_0]$ in Model 2. The new energy function is as follows.

$$E_\epsilon[z|u_0] = \int_{\Omega} [\epsilon|\nabla z|^2 + \frac{1}{\epsilon} sin^2 \pi z] dx + \lambda \sum_{k=0}^{K-1} \int_{\Omega} [log(\sqrt{2\pi}\sigma_k) + \frac{(c_k - u_0)^2}{2\sigma_k^2}] e^{-\alpha(z-k)^2} dx \tag{2–20}$$

We call this final model Model 3. In case that no confusion will be caused, we still use $G[z|u_0]$ to denote the second part of (2–20). In this paper, we can simply denote the

energy function by

$$E_\epsilon[z|u_0] = F_\epsilon[z|u_0] + \lambda G[z|u_0]. \tag{2–21}$$

Note that if we take all σ_k $(k = 0, 1, ...K-1)$ to be $\frac{1}{\sqrt{2\pi}}$, and change $e^{-\alpha x^2}$ back to $sinc(x)$, it is exactly Model 1.

2.1.3 Existence of Solution

In the following, Ω always denotes a bounded open set with smooth boundary $\partial\Omega$. $W^{1,q}(\Omega)$ $(1 < q < \infty)$ denotes the Sobolev space. K is a positive integer. $\mathbf{c} = (c_0, c_1, ..., c_{K-1}) \in R^K$ and $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{K-1}) \in R_+^K$. $\epsilon > 0$ and $\lambda \in R$ are parameters.

Definition 1. *We say that a functional $J[\cdot]$ is weakly lower semicontinuous on $W^{1,q}(\Omega)$, provided*

$$J[u] \leq \liminf_{k\to\infty} J[u_k] \tag{2–22}$$

whenever

$$u_k \rightharpoonup u \text{ weakly in } W^{1,q}(\Omega). \tag{2–23}$$

Theorem 2.1. *Assume that L is smooth, bounded below, and in addition, the mapping*

$$p \mapsto L(p, z, x)$$

is convex with respect to p for each $z \in R$ and $x \in \Omega$. Then the functional

$$J[w] := \int_\Omega L(Dw(x), w(x), x)\, dx \tag{2–24}$$

is weakly lower semicontinuous on $W^{1,q}(\Omega)$, $1 < q < \infty$.

To be convenient for statement, we denote the energy functional in (2–20) by

$$\begin{aligned} I[z] &= \int_\Omega [\epsilon|\nabla z|^2 + \frac{1}{\epsilon} sin^2 \pi z] dx \\ &+ \lambda \sum_{k=0}^{K-1} \int_\Omega [log\sqrt{2\pi}\sigma_k + \frac{(c_k - u_0)^2}{2\sigma_k^2}] e^{-\alpha(z-k)^2} dx. \end{aligned} \tag{2–25}$$

and denote by $\mathcal{A}$ the following family of functions

$$\mathcal{A} := \{z \in W^{1,2}(\Omega) |\ z = g \text{ on } \partial\Omega \text{ in the trace sense}\}. \tag{2–26}$$

Theorem 2.2. *Suppose* $min_{\{0\le i\le K-1\}}\sigma_i \ge a$ *for some constant* $a > 0$*. Then there exists a function* $z_0(x) \in \mathcal{A}$ *solving*

$$I[z_0] = min_{z\in\mathcal{A}} I[z]. \tag{2–27}$$

Proof.

Step 1: *Show that there exists a sequence* z_l *converging to* $inf_{z\in A} I[z]$.

It is easy to see that

$$I[z] \ge \lambda \sum_{k=0}^{K-1} \int_\Omega [log(\sqrt{2\pi}\sigma_k)] e^{-\alpha(z-k)^2} dx. \tag{2–28}$$

If $\sigma_k \ge \frac{1}{\sqrt{2\pi}}$, then $log(\sqrt{2\pi}\sigma_k) \ge 0$, and so

$$\int_\Omega [log(\sqrt{2\pi}\sigma_k)] e^{-\alpha(z-k)^2} dx \ge 0. \tag{2–29}$$

If $\sigma_k < \frac{1}{\sqrt{2\pi}}$, then $log\sqrt{2\pi}\sigma_k < 0$, and so

$$\int_\Omega [log(\sqrt{2\pi}\sigma_k)] e^{-\alpha(z-k)^2} dx \ge |\Omega| log(\sqrt{2\pi}\sigma_k) \tag{2–30}$$

since $e^{-\alpha(z-k)^2} \le 1$. To sum up, we have in general,

$$\int_\Omega [log(\sqrt{2\pi}\sigma_k)] e^{-\alpha(z-k)^2} dx \ge min\{0, |\Omega| log(\sqrt{2\pi}\sigma_k)\} \tag{2–31}$$

It follows that

$$I[z] \ge \lambda \sum_{k=0}^{K-1} min\{0, |\Omega| log\sqrt{2\pi}\sigma_k\} \ge c > -\infty \tag{2–32}$$

since $min_{\{0\le k\le K-1\}}\sigma_k \ge a > 0$. So, $I(z)$ is bounded below. Set $m := inf_{z\in\mathcal{A}} I[z]$. Then m is finite. Select a minimizing sequence $\{z_l\}_{l=1}^\infty$ satisfying

$$I[z_l] \to m. \tag{2–33}$$

Step 2: *Show that there is a subsequence z_{k_j} converging to $z_0 \in A$.*

By the same argument as in Step 1, we have

$$\begin{aligned} I(z) &\geq \epsilon \int_\Omega |\nabla z|^2 dx + \lambda \sum_{k=0}^{K-1} \int_\Omega [log(\sqrt{2\pi}\sigma_k)] e^{-\alpha(z-k)^2} dx \\ &\geq \epsilon \int_\Omega |\nabla z|^2 dx + \lambda \sum_{k=0}^{K-1} |\Omega| log(\sqrt{2\pi}\sigma_k) \end{aligned} \tag{2–34}$$

So, the fact that $\{I[z_l]\}_{l=1}^{\infty}$ is bounded implies that $\{\int_\Omega |\nabla z_l|^2 dx\}_{l=1}^{\infty}$ is also bounded, i.e., $\{||Dz_l||_2\}_{l=1}^{\infty}$ is bounded. Choose fixed $z \in \mathcal{A}$. Then $u_l = z_l - z \in W_0^{1,2}$, and $\{\|Du_l\|_2\} = \{||Dz_l - Dz||_2\}$ is bounded in $W_0^{1,2}(\Omega)$. By Poincare's inequality, we have

$$\|z_l - z\| \leq c\|Dz_l - Dz\| \leq C \tag{2–35}$$

which means that $u_l = z_l - z$ is bounded in $L^2(\Omega)$, and thus bounded in $W^{1,2}(\Omega)$. Since both z_l and z are in $W^{1,2}(\Omega)$, it follows that $u_l \in W_0^{1,2}(\Omega)$. Now by weak compactness of reflexive Banach spaces, there is a subsequence of $\{u_l = z_l - z\}$, denoted by $u_{l_j} = z_{l_j} - z$ such that

$$u_{l_j} = z_{l_j} - z \rightharpoonup v \tag{2–36}$$

weakly for some $v \in W_0^{1,2}(\Omega)$, or equivalently,

$$z_{l_j} \rightharpoonup z_0 \tag{2–37}$$

for some $z_0 \in \mathcal{A}$.

Step 3: *Show that $I[z_0] = m = min_{z \in A} I[z]$.*

Note that if we define

$$\begin{aligned} &L(p, z, x) = [\epsilon p^2 + \frac{1}{\epsilon} sin^2 \pi z] + \\ &\lambda \sum_{k=0}^{K-1} [log(\sqrt{2\pi}\sigma_k) + \frac{(c_k - u_0(x))^2}{2\sigma_k^2}] e^{-\alpha(z-k)^2} \end{aligned} \tag{2–38}$$

then we have $I[z] = \int_\Omega L(Dz, z, x)dx$. It is easy to see that L is smooth and bounded below. In fact, we have

$$L(p, z, x) \geq \lambda log\sqrt{2\pi}a > -\infty. \tag{2–39}$$

Moreover, $L(p, z, x)$ is convex in p. By Theorem 2.1,

$$I[z_0] \leq liminf_{j\to\infty} I[z_{l_j}] = m \tag{2–40}$$

But since $z_0 \in \mathcal{A}$, we also have $I[z_0] \geq m$. Therefore,

$$I[z_0] = m. \qquad \blacksquare$$

2.2 Implementation and Considerations

Like Model 1, the main challenge for our model comes from the non-convexity of the energy function $E_\epsilon[z|u_0]$ with respect to z. So, we also adopt the CCCP method. Before presenting our implementation scheme, we first give a short review on CCCP. We only describe the basic result that is necessary to understand this paper. Details can be found in [88].

2.2.1 Convex-Concave Procedure (CCCP)

The convex-concave procedure is a convex splitting method in optimization which was explored by Yuille and Rangarajan [105].

Theorem 2.3. *Consider an energy function which is bounded below and is an addition of convex and concave functions:*

$$E(\vec{x}) = E_{convex}(\vec{x}) + E_{concave}(\vec{x}).$$

Then, the discrete iterative CCCP algorithm given by

$$\nabla E_{convex}(\vec{x}^{n+1}) = -\nabla E_{concave}(\vec{x}^n),\ n = 0, 1, \ldots \tag{2–41}$$

is guaranteed to monically decrease the energy $E(\vec{x})$ as a function of time and to converge to a local minimum or a saddle point of $E(\vec{x})$.

Proposition 2.1. *Let $F(u) = \int_\Omega f(u(x))dx$, where $f \in C^2(R)$.*

Case 1. $f''' \leq \gamma$ for some $\gamma \geq 0$. We define the splitting

$$F(u) = \int_\Omega \frac{\gamma}{2}u^2 dx - \int_\Omega (\frac{\gamma}{2}u^2 - f(u))dx := F^1(u) - F^2(u); \tag{2–42}$$

Case 2. $f''' \geq -\gamma$ for some $\gamma \geq 0$. We define the splitting

$$F(u) = \int_\Omega (\frac{\gamma}{2}u^2 + f(u))dx - \int_\Omega \frac{\gamma}{2}u^2 dx := F^1(u) - F^2(u). \tag{2–43}$$

Then in either case, both $F^1(u)$ and $F^2(u)$ are convex.

The proof is trivial since in Case 1,

$$(F^2)'''(u) = \int_\Omega (\gamma - f'''(u))dx \geq 0$$

if $f''' \leq \gamma$ for some $\gamma \geq 0$. And in Case 2,

$$(F^1)'''(u) = \int_\Omega (\gamma + f'''(u))dx \geq 0$$

if $f''' \geq -\gamma$ for some $\gamma \geq 0$.

2.2.2 Iteration Scheme

In our scheme, there are three group of stuffs to be determined, the means $\mathbf{c} = (c_0, c_1, ..., c_{K-1})$, the variances $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{K-1})$, and the signature function $z(x)$. We compute $E_\epsilon[z, \mathbf{c}, \sigma|u_0]$ regarding z, $\mathbf{c}$, and σ as independent variables. This allows the application of the alternating minimization (AM) scheme, i.e., to alternatingly optimize the three conditional energies $E_\epsilon[z|\mathbf{c}, \sigma, u_0]$, $E_\epsilon[\mathbf{c}|\sigma, z, u_0]$, and $E_\epsilon[\sigma|\mathbf{c}, z, u_0]$, under the iterations of $z^n \to \mathbf{c}^n \to \sigma^n \to z^{n+1}$ given by

$$\mathbf{c}^n = argmin E_\epsilon[\mathbf{c}|\sigma^n, z^n, u_0] \tag{2–44}$$

$$\sigma^n = argmin E_\epsilon[\sigma|\mathbf{c}^n, z^n, u_0] \tag{2–45}$$

$$z^{n+1} = argmin E_\epsilon[z|\mathbf{c}^n, \sigma^n, u_0]. \tag{2–46}$$

It is well known (see Vogel [99]) that the AM scheme is monotone:

$$E_\epsilon[z^{n+1}, \mathbf{c}^{n+1}, \sigma^{n+1}|u_0] \leq E_\epsilon[z^n, \mathbf{c}^n, \sigma^n|u_0].$$

For equation (2–44) and (2–45), one can simply have at pixel level,

$$c_k = \frac{\sum_i \sum_j u_0^{i,j} e^{-\alpha(z_{i,j}-k)^2}}{\sum_i \sum_j e^{-\alpha(z_{i,j}-k)^2}}, \; k = 0, 1, ..., K-1 \tag{2–47}$$

and

$$\sigma_k^2 = \frac{\sum_i \sum_j (u_0^{i,j} - c_k)^2 e^{-\alpha(z_{i,j}-k)^2}}{\sum_i \sum_j e^{-\alpha(z_{i,j}-k)^2}}, \; k = 0, 1, ..., K-1, \tag{2–48}$$

where $z_{i,j}^n$ denotes computational phase field on the Cartesian image domain. To find $argminE_\epsilon[z|c^n, \sigma^n, u_0]$, we now apply the CCCP algorithm. For functional $F_\epsilon[z]$, since $\frac{d^2}{dz^2} sin^2 \pi z \geq -2\pi^2$, we split $F_\epsilon[z]$ as

$$F_\epsilon[z] = (F_\epsilon[z] + \frac{\pi^2}{\epsilon}\int_\Omega z^2 dx) - \frac{\pi^2}{\epsilon}\int_\Omega z^2 dx := F_\epsilon^1[z] - F_\epsilon^2[z]. \tag{2–49}$$

Similarly, since $\frac{d^2}{dz^2} e^{-\alpha(z-k)^2} \leq 4\alpha e^{-1.5}$, we split $G[z|u_0]$ as

$$\begin{aligned} G[z|u_0] = &\sum_{k=0}^{K-1} \int_\Omega 2\alpha e^{-1.5} [log\sqrt{2\pi}\sigma_k + \frac{(c_k - u_0)^2}{2\sigma_k^2}](z-k)^2 dx \\ -(&\sum_{k=0}^{K-1} \int_\Omega 2\alpha e^{-1.5} [log\sqrt{2\pi}\sigma_k + \frac{(c_k - u_0)^2}{2\sigma_k^2}](z-k)^2 dx - G[z|u_0]). \end{aligned} \tag{2–50}$$

We simply denote it by

$$G[z|u_0] := G^1[z|u_0] - G^2[z|u_0]. \tag{2–51}$$

By Proposition 1, $F_\epsilon^1[z]$, $F_\epsilon^2[z]$, $G^1[z|u_0]$ and $G^2[z|u_0]$ are all convex. By Theorem 1, we can then use the CCCP iteration scheme via Frechet derivative, i.e.,

$$(F^1 + \lambda G^1)\ddot{(}z^{n+1}) = (F^2 + \lambda G^2)\ddot{(}z^n). \tag{2–52}$$

Under integration by parts, the above equation is equivalent to the following PDE.

$$\begin{aligned}
&[-2\epsilon\triangle z^{n+1} + \frac{\pi}{\epsilon} sin2\pi z^{n+1}] + \frac{2\pi^2}{\epsilon} z^{n+1} \\
&+ \lambda \sum_{k=0}^{K-1} 4\alpha e^{-1.5} (log\sqrt{2\pi}\sigma_k + \frac{|c_k - u_0|^2}{2\sigma_k^2})(z^{n+1} - k) \\
&= \frac{2\pi^2}{\epsilon} z^n + [\lambda \sum_{k=0}^{K-1} 2\alpha e^{-\alpha(z^n-k)^2} (log\sqrt{2\pi}\sigma_k + \frac{|c_k - u_0|^2}{2\sigma_k^2})(z^n - k)] \\
&+ \lambda \sum_{k=0}^{K-1} 4\alpha e^{-1.5} (log\sqrt{2\pi}\sigma_k + \frac{|c_k - u_0|^2}{2\sigma_k^2})(z^n - k).
\end{aligned} \tag{2–53}$$

Here the terms in the square brackets come from the Euler-Lagrange equation of E_ϵ. We use the following numerical approximation:

$$\triangle z^{n+1} \sim z^n_{i-1,j} + z^n_{i,j-1} + z^n_{i,j+1} + z^n_{i+1,j} - 4z^{n+1}_{i,j} \tag{2–54}$$

and

$$sin2\pi z^{n+1} \sim \frac{sin2\pi z^n}{z^n} z^{n+1} \tag{2–55}$$

The corresponding iteration scheme is as follows.

$$\begin{aligned}
&\left(8\epsilon + \frac{\pi}{\epsilon}(\frac{sin2\pi z^n_{i,j}}{z^n_{i,j}} + 2\pi)\right) z^{n+1}_{i,j} \\
&+ \left(4e^{-1.5}\lambda\alpha \sum_{k=0}^{K-1} (log\sqrt{2\pi}\sigma_k + \frac{|u_{i,j} - C_k|^2}{2\sigma_k^2})\right) z^{n+1}_{i,j} \\
&= 2\epsilon(z^n_{i-1,j} + z^n_{i,j-1} + z^n_{i+1,j} + z^n_{i,j+1}) + \frac{2\pi^2}{\epsilon} z^n_{i,j} \\
&+ \lambda\alpha \sum_{k=0}^{K-1} (log\sqrt{2\pi}\sigma_k + \frac{|u_{i,j} - C_k|^2}{2\sigma_k^2})(4e^{-1.5} z^n_{i,j} + 2(z^n_{i,j} - k)e^{-\alpha(z^n_{i,j}-k)^2}).
\end{aligned} \tag{2–56}$$

Now we have all the three minimizations (2–44), (2–45) and (2–46). We can use the alternating minimization scheme. Although one can treat c_k and σ_k as an independent variable in image segmentation [92], we update c_k and σ_k in every alternating step, as in any usual AM scheme.

2.2.3 Initialization

Since the energy functional may not be convex we may be only able to find a local minimum or saddle point, which is guaranteed by Theorem 1. Therefore, the choice of the initial function z_0 can affect the segmentation result greatly. Sometimes, it would be even worse since the local minimum may be two phases when it is expected to be three or more phases. To avoid this phenomenon, we would like to choose an initial z_0 that is not "too far" from the expected segmentation results. In this paper, we choose $z_0(x) = \frac{\widetilde{u_0}(x)-m}{M-m}K$ as the initialization of $z(x)$, where $\widetilde{u_0}(x)$ is a smooth version of $u_0(x)$, K is the number of phases, and $M = maximum(\widetilde{u_0}(x))$ and $m = minimum(\widetilde{u_0}(x))$, respectively. We do so for two reasons. First, we want to set the range of $z_0(x)$ to be $[0, K]$ if we want to partition the image into K regions, since the optimal $z(x)$ is a smooth version of a piecewise constant function ranging from zero to K. The smaller range of $z_0(x)$ can lead to a segmentation with less numbers of phases as desired. Second, note that $\frac{\widetilde{u_0}(x)-m}{M-m}K$ is just a shift and rescaling of the smoothed image u_0, which fully reflects the feature of the original image u_0 for segmentation. Hence, our choice of z_0 is somehow close to the expected optimal solution.

2.2.4 Segmentation Decision

Finally, once the iteration is stable, we apply the hard thresholding decision rule: $k - \frac{1}{2} \leq z < k + \frac{1}{2}$ for each individual k-th phase. A simple morphological transformation can also be employed to remove any spurious dots due to the hard thresholding.

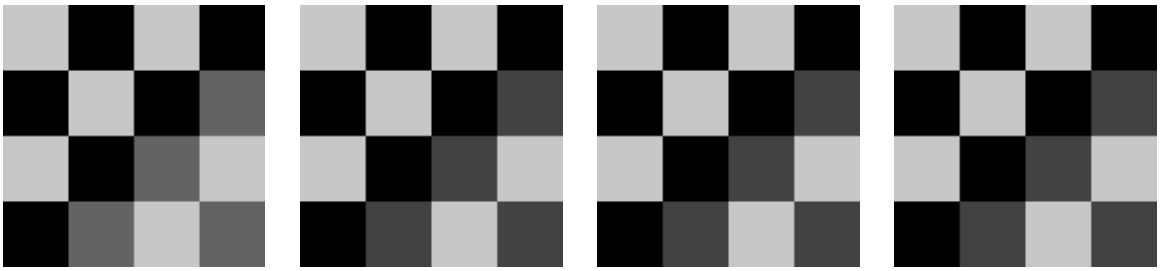
2.3 Experiment and Discussions

In our experiments, we use two different kind of images: an artificial synthetic image and a human brain image. We exhibit our experiment results in a comparison way in two groups. Group 1 are "clean" images, while Group 2 are the same images as in Group 1 but added with Gaussian noise. In order to compare the three models as precise as we can, we did not apply any post-processing in our experiments after hard thresholding, such as morphological transform.

Since for different models, the iterations approximate their "best" results at different parameters, the results shown in our experiments are chosen as good as we can for all models. From the segmentation results, we see that, for "clean" images, all three models are segmented almost same good (See Fig. 2-1, where (A) is the original artificial synthetic image, (B) is the segmentation using Model 1($\lambda = 150$), (C) is the segmentation using Model 2($\lambda = 150, \alpha = 5$), and (D) is the segmentation using Model 3 ($\lambda = 150, \alpha = 5$)). However, for noisy images, both Model 2 and Model 3 are better than Model 1 (See Fig. 2-2 and Fig. 2-4). In Fig. 2-2, (A) is the original Artificial synthetic image with Gaussian noise ($m = 0, \sigma = 0.02$), (B) is the segmentation using Model 1($\lambda = 150$), (C) is segmentation using Model 2($\lambda = 150, \alpha = 5$), and (D) is the segmentation using Model 3 ($\lambda = 150, \alpha = 5$). From the result, we can see that there are more spurious dots in Fig. 2-2(B) than in Fig. 2-2(C) and in Fig. 2-2(D). Meanwhile, comparing Fig. 2-2(C) and Fig. 2-2(D), the edge of Fig. 2-2(D) is damaged less than Fig. 2-2(C). Fig. 2-3 is the segmentation for a real MRI brain image, where (A) is the original human brain image without noise, (B) is segmentation using Model 1 ($\lambda = 150$), (C) is the segmentation using Model 2($\lambda = 150, \alpha = 5$), and (D) is the segmentation using Model 3 ($\lambda = 5000, \alpha = 15$). Fig. 2-4 shows the result for the image in Fig. 2-3 added some Gaussian noise. In Fig. 2-4 where $\epsilon = 2, K = 3$, we can also see that Model 3 is better than Model 1 and Model 2 in that spurious dots can hardly be found from the white matters in Fig. 2-4(D), while in Fig. 2-4(B) and Fig. 2-4(C), there are still many spurious dots in the white matters.

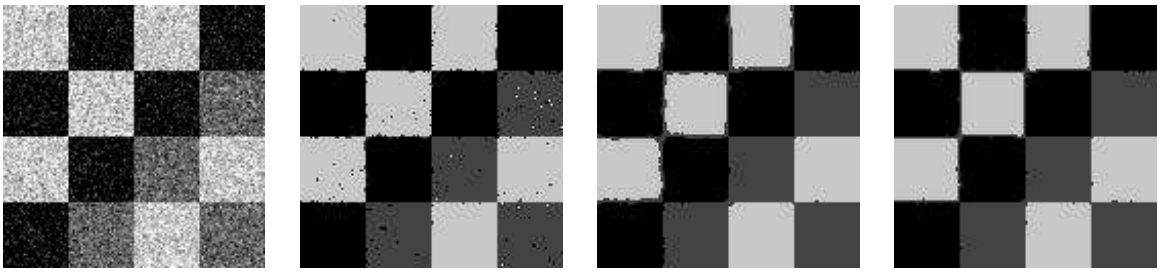
2.4 Conclusions

In this chapter, we improved the sine-sinc model in two aspects. First, we replace sinc function by exponential function. Second, we extend the model to fit those images not necessarily piecewise constant but close to Gaussian distribution with different parameters. The experiments show that our new model is more efficient when the image



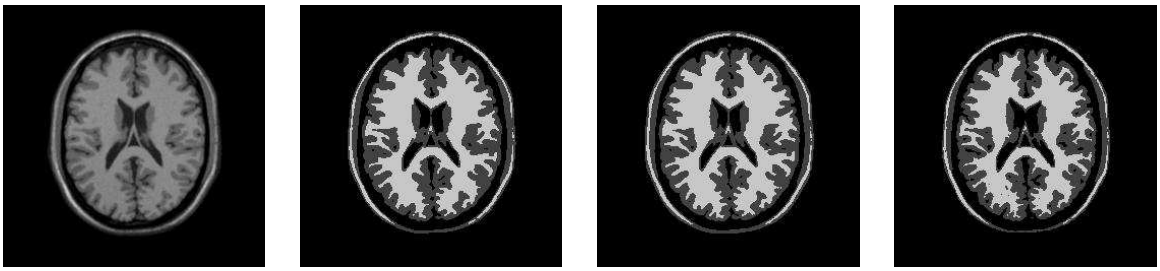
(A) (B) (C) (D)

Figure 2-1. Segmentations for clean image.



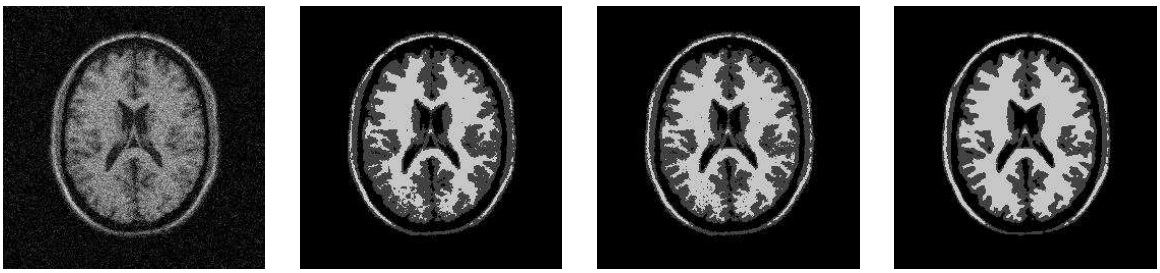
(A) (B) (C) (D)

Figure 2-2. Segmentations for noisy image.



(A) (B) (C) (D)

Figure 2-3. Segmentations of real MRI brain image.



(A) (B) (C) (D)

Figure 2-4. Segmentations of real MRI brain image with noise.

is noisy. We also recommended a normalized initialization of the signature function, which helps the iteration converge to the “true” result.

CHAPTER 3
MULTIPHASE SOFT SEGMENTATION USING CONSTRUCTED MEMBERSHIP FUNCTIONS

There is a common point for piecewise constant variational models. They all contain a regularity term and a fidelity term, and assume that the image is piecewise constant or smooth enough, which makes them vulnerable to noise. Although we can choose large weight for the fidelity term to restrain noise, the segmentation result is sensitive to the choice of the weights between the regularity term and the fidelity term. If the weight of the regularity term is much bigger than that of the fidelity term, then the edge and fine structure can be preserved very well but the noise may not be removed ideally. On the other hand, if we choose the contrary, i.e., the weight of fidelity term is much bigger than that of the regularity term, noise will be removed very well but edge may be damaged and some fine structure may be lost. To avoid this dilema, there have been some studies in recent years aiming at employing non-local information of images, such as graph-cut based method (discrete case) [14] and non-local variation based method (continuous case) [15]. In these methods, whether a point is an edge depends not only on the local intensity difference, but also on finding how often the similar features of the point have been repeated in the whole domain. By taking non-local information, the edge can be well preserved while the noise is smoothed.

Different from using non-local information, the framework in this paper uses stochastic theory to restrain noise and improve segmentation. It can be thought an extension of piecewise constant Mumford-Shah kind models mentioned above. More precisely, we assume that the intensity of each point is a Gaussian distributed random sample. In each phase Ω_k, the points follow a same Gaussian distribution with mean c_k and variance σ_k. We assume that the clean true image $u(x)$ is still piecewise constant (i.e., inside each phase, the intensities are always a constant equal to c_k) but contaminated by a Gaussian noise $n(x)$, i.e., $I(x) = u(x) + n(x)$. As a result, the intensities of points in a same phase will not be a constant, but a family of samples from

a same Gaussian distribution. By maximizing the likelihood (joint PDF) of all random samples, the fidelity term becomes the following form (see detail in Section 3).

$$\sum_{k=0}^{K-1} \int_{\Omega} (log\sigma_k + \frac{(c_k - I(x))^2}{2\sigma_k^2}) p_k(x)\, dx, \tag{3–1}$$

where $0 \leq p_k(x) \leq 1$ is a smooth approximation of characteristic function of k-th phase, and c_k and σ_k are mean and standard variation respectively. In application, $p_k(x)$ are usually chosen in such a way that $\sum_{k=0}^{K-1} p_k(x) = 1$ holds in the whole domain. So, the model becomes a standard soft segmentation with $p_k(x)$ as membership functions. Compared with those piecewise constant models [32, 33, 36, 40, 65, 87, 88, 98, 106] where variants are not involved, for point x where $(c_k - I(x))^2$ is relatively larger than $(c_i - I(x))^2$ for some $i \neq k$ due to noise, the model with fidelity term (3–1) can still classify it to k-th phase (the correct phase) and the noise can therefore be restrained. The reason is that the effect of large $(c_k - I(x))^2$ will be partly counteracted by the variance. On the other hand, based on probability theory, we know that the probability that $|I_k(x) - c_k| > \epsilon$ for some $\epsilon > 0$ and all $x \in S$ in a connected area $S \subset \Omega_k$ is much smaller than the probability that $|I_k(x) - c_k| > \epsilon$ for one isolated point x. This fact can guarantee that the model based on (3–1) can preserve small structure while removing isolated noise.

In this chapter, we do not introduce membership functions as an approximation of characteristic function. Instead, we introduce a constructed function $h(x)$ so that the composite function $h \circ (z(x) - k)$ has the property of membership function $p_k(x)$. As a result, the model itself is still a soft segmentation. The advantage of applying constructed function in the model lies in the fact that there will be less variables introduced in the model which makes the discussion and the implementation easier. For example, as long as we know $z(x)$, the probability $p_k(x)$ is followed by $p_k(x) = h \circ (z(x) - k)$.

In MRI brain image, there are three tissues, i.e., white matter, gray matter and cerebrospinal fluid (c.s.f). Due to limited resolution and non-regularity of the boundaries of pure matters, there are some partial volumes formed by overlaps of pure matters. Those partial volumes are hard to be classified to pure matters. We assume each partial volume follows a mixed Gaussian distribution generated by two Gaussian distributions which correspond to pure matters. The proposed frame work can then be applied to partial volume segmentation. The rest of the paper is organized as follows. In Section 3.1, we develop the framework of multiphase segmentation by combining phase transition theory and Gaussian distribution. Section 3.2 is the implementation and some considerations. Then in Section 3.3, we apply the frame work to partial volume segmentation. Experiments are carried out in Section 3.4. We show by examples the advantage of the proposed model by comparing with other multiphase segmentation models. Finally, the chapter is closed with a short conclusion.

3.1 Framework Development

In this section, we develop a framework of multiphase soft segmentation with constructed functions and phase-transition theory. In order to be self-included, something discussed in last chapter is still stated here. Let $I \in L^2(\Omega)$ be an image defined in a bounded, smooth and open domain $\Omega \subset R^2$. Suppose the image contains K phases and we take the image as a random field with the following assumptions:

(a) At each point $x \in \Omega$, the intensity $I(x)$ is a random variable;

(b) All the random variables $\{I(x)|x \in \Omega\}$ are independent;

(c) In each phase $I|_{\Omega_k}, 0 \leq k \leq K-1$, all the random variables $\{I(x) : x \in \Omega_k\}$ are identically distributed as a Gaussian distribution with same mean c_k and same variance σ_k^2 (which are to be determined).

We want to maximize the likelihood, joint pdf of $\{I(x), x \in \Omega\}$, which is equivalent to minimize the following energy (the detail can be found in [34]):

$$-logL(c,\sigma) = \sum_{k=0}^{K-1} \int_{\Omega_k} (log\sigma_k + \frac{(c_k - I)^2}{2\sigma_k^2}) = \sum_{k=0}^{K-1} \int_{\Omega} (log\sigma_k + \frac{(c_k - I)^2}{2\sigma_k^2}) 1_{\Omega_k}, \tag{3–2}$$

Second, we borrow the length term from [88] which is based on phase transition theory. Let I be the image defined above. The *signature function* $z(x)$ is defined by

$$z(x) = k, \text{if } x \in \Omega_k, k = 0, 1, ..., K-1. \tag{3–3}$$

Then the total variation of $z(x)$ is $\int_\Omega |Dz|$, where Dz denotes the vectorial Radon measure of the total variation (TV) of z. We have the following relation:

$$\sum_{k=0}^{K-1} |\partial\Omega_k| \leq \int_\Omega |Dz(x)| \leq K \sum_{k=0}^{K-1} |\partial\Omega_k|, \tag{3–4}$$

where $\partial\Omega_k$ is the boundary of Ω_k. Thus, $\int_\Omega |Dz(x)|$ works as the length of the edges of all phases. The ideal model is to minimize the sum $E[z, c, \sigma|I]$ of the length term and the negative log-likelihood (3–2).

$$E[z, c, \sigma|I] = \int_\Omega |Dz(x)| + \lambda \sum_{k=0}^{K-1} \int_\Omega (log\sigma_k + \frac{(c_k - I)^2}{2\sigma_k^2}) 1_{z=k}. \tag{3–5}$$

Note that when $\sigma_0 = \sigma_1 = ... = \sigma_{K-1} \neq 0$, the model is equivalent to piecewise constant Mumford-Shah model. However, this model has intrinsic drawback due to the discreteness of the signature function $z(x)$ and the characteristic function $1_{z=k}$, which will impede the application of PDE based method. So, we need to use some relaxed version. For the signature function, we introduce its relaxed version via the celebrated model of Modica and Mortola on phase transitions in material science and fluid mechanics. We refer the authors to papers [20, 69, 70, 88] for further understanding to phase transition theory.

Let $\tilde{z}(x)$ be a smoothed version of the signature function $z(x)$, which is called phase fields. To be simple, we still use the same notation $z(x)$ to denote the phase field.

Mortola [69] established that

$$\int_{\Omega}[\epsilon|\nabla z|^2+\frac{1}{\epsilon}sin^2\pi z]dx \tag{3–6}$$

γ-converges to $\frac{4}{\pi}\int_{\Omega}|Dz(x)|$ in $L^1(\mathcal{R})$ for phase fields $z(x)$ that ultimately only take integer values. Now we replace the length term in (3–5) by (3–6) when ϵ is small enough. For the characteristic function $1_{z=k}$, we can use any smooth function $h_k(x)$ as an approximation if only $0\leq h_k(x)\leq 1$ and $\sum_{k=0}^{K-1}=1$ for all $x\in\Omega$. The ideal smooth function $h_k(x)$ should have the following properties.

(a) At each point $x\in\Omega$ where $z(x)=k$, $h_k(x)$ is close to 1;

(b) At each point $x\in\Omega$ where $|z(x)-k|>0.5$, $h_k(x)$ is close to 0.

If we can choose $h(x)$ satisfying that $h(x)$ is close to one at small neighborhood of 0 and close to zero elsewhere, then $h_k(x)$ can be denoted as $h(z(x)-k)$ since $z(x)$ is almost integer. Then, the fidelity term (3–2) becomes

$$\sum_{k=0}^{K-1}\int_{\Omega}(log\sigma_k+\frac{(c_k-I)^2}{2\sigma_k^2})h(z(x)-k). \tag{3–7}$$

Now, as the final step, we integrate the relaxed length term (3–6) and the relaxed fidelity term (3–7). The new energy functional is

$$\begin{aligned}E_{\epsilon}[z,c,\sigma|I]&=\int_{\Omega}[\epsilon|\nabla z|^2+\frac{1}{\epsilon}sin^2\pi z]dx+\lambda\sum_{k=0}^{K-1}\int_{\Omega}(log\sigma_k+\frac{(c_k-I)^2}{2\sigma_k^2})h(z-k)\\&:=F_{\epsilon}[z]+G[z,\mathbf{c},\sigma|I].\end{aligned} \tag{3–8}$$

This is the proposed framework of multiphase segmentation. Compared with most existing models, the proposed model is robust to noise and more applicable; Moreover, since the phase function $z(x)$ is involved in the model, segmentation can be directly derived from the phase function.

3.2 Implementation and Considerations

In our framework (3–8), there are three groups of parameters to be determined, the means $\mathbf{c} = (c_0, c_1, \ldots, c_{K-1})$, the variances $\sigma^2 = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{K-1}^2)$, and the phase field $z(x)$. We compute $E_\epsilon[z, \mathbf{c}, \sigma|I]$ regarding z, $\mathbf{c}$, and σ as independent variables. This allows the application of the alternating minimization (AM) scheme, i.e., to alternatingly optimize the three conditional energies.

$E_\epsilon[z|\mathbf{c}, \sigma, I]$, $E_\epsilon[\mathbf{c}|\sigma, z, I]$, and $E_\epsilon[\sigma|\mathbf{c}, z, I]$, under the iterations of $z^{(n)} \to \mathbf{c}^{(n)} \to \sigma^{(n)} \to z^{(n+1)}$ given by

$$\mathbf{c}^{(n)} = argminE_\epsilon[\mathbf{c}|\sigma^{(n)}, z^{(n)}, I] \tag{3–9}$$

$$\sigma^{(n)} = argminE_\epsilon[\sigma|\mathbf{c}^{(n)}, z^{(n)}, I] \tag{3–10}$$

$$z^{(n+1)} = argminE_\epsilon[z|\mathbf{c}^{(n)}, \sigma^{(n)}, I]. \tag{3–11}$$

It is well known (see Vogel [99]) that the AM scheme is monotone:

$$E_\epsilon[z^{(n+1)}, \mathbf{c}^{(n+1)}, \sigma^{(n+1)}|I] \leq E_\epsilon[z^{(n)}, \mathbf{c}^{(n)}, \sigma^{(n)}|I].$$

For equation (3–9) and (3–10), one can simply have at pixel level,

$$c_k^{(n)} = \frac{\sum_i \sum_j I^{i,j} h(z_{i,j}^{(n)} - k)}{\sum_i \sum_j h(z_{i,j}^{(n)} - k)}, \qquad k = 0, 1, \ldots, K-1 \tag{3–12}$$

and

$$(\sigma_k^{(n)})^2 = \frac{\sum_i \sum_j (I^{i,j} - c_k^{(n)})^2 h(z_{i,j}^{(n)} - k)}{\sum_i \sum_j h(z_{i,j}^{(n)} - k)}, \qquad k = 0, 1, \ldots, K-1, \tag{3–13}$$

where $z_{i,j}^{(n)}$ denotes the computational phase field on the Cartesian image domain at step n.

For $argminE_\epsilon[z|c^{(n)}, \sigma^{(n)}, I]$, since the model is not an active contour model (the integral of the fidelity term can not be separated by boundaries), it is not proper to

use level set method for implementation. Besides, our purpose is not to find the rough segmentation of $z(x)$ but the exact value of $z(x)$ so that we can estimate pure matters form the partial volume in MRI brain images using the composition $h \circ (z(x) - k)$ for $0 \leq k \leq K-1$. Note that the Euler-Lagrange equation of $E_\epsilon[z|c^{(n)}, \sigma^{(n)}, I]$ with respect to z is

$$-2\epsilon\Delta z(x) + \frac{\pi}{\epsilon} sin(2\pi z(x)) + \lambda \sum_{k=0}^{K-1} \ddot{h}(z(x) - k)(log\sigma_k + \frac{|c_k - I|^2}{2\sigma_k^2}) = 0. \tag{3–14}$$

As for most multiphase segmentation models, the energy functional $E_\epsilon[z, c, \sigma|I]$ is not convex with respect to z. Thus, simple iteration scheme obtained directly from the Euler-Lagrange equation may not converge. In our application, we adopt the convex-concave procedure (CCCP). Before apply CCCP to our model, we first give a short review on CCCP. We only describe the basic result that is necessary to understand this paper. For more details, we refer the readers to [87, 105]. We also recommend the reader to use selected initial value to help converge (but not guaranteed).

We recall the following theorem on CCCP.

Theorem 3.1. *Consider an energy function which is bounded below and is an addition of convex and concave functions:*

$$E(\vec{x}) = E_{convex}(\vec{x}) + E_{concave}(\vec{x}).$$

Then, the discrete iterative CCCP algorithm given by

$$\nabla E_{convex}(\vec{x}^{(n+1)}) = -\nabla E_{concave}(\vec{x}^{(n)}), n = 0, 1, \ldots \tag{3–15}$$

is guaranteed to monotonically decrease the energy $E(\vec{x})$ as a function of time and to converge to a local minimum or a saddle point of $E(\vec{x})$.

Proposition 3.1. *Let $F(u) = \int_\Omega f(u(x))dx$, where $f \in C^2(R)$. If $f\ddot{\Gamma}\Gamma \geq -\gamma$ for some $\gamma \geq 0$. We define the splitting*

$$F(u) = \int_\Omega (\frac{\gamma}{2}u^2 + f(u))dx - \int_\Omega \frac{\gamma}{2}u^2 dx := F^1(u) - F^2(u). \tag{3–16}$$

Then both $F(u)$ and $F^2(u)$ are convex.

The proof is trivial since

$$(F^1)\ddot{\Gamma}\Gamma(u) = \int_\Omega (\gamma + f\ddot{\Gamma}\Gamma(u))dx \geq 0$$

if $f\ddot{\Gamma}\Gamma \geq -\gamma$ for some $\gamma \geq 0$.

3.2.1 Iteration Scheme for phase function $z(x)$

To find $argminE_\epsilon[z|c^{(n)}, \sigma^{(n)}, I]$, we now apply the CCCP algorithm. In this section, we will temporally suppose the parameters $c_k, \sigma_k, 1 \leq k \leq K$ are all known for the purpose of statement.

Note that $\frac{d^2}{dz^2}sin^2\pi z \geq -2\pi^2$, we split $F_\epsilon[z]$ as

$$F_\epsilon[z] = \left\{F_\epsilon[z] + \frac{\pi^2}{\epsilon}\int_\Omega z^2 dx\right\} - \frac{\pi^2}{\epsilon}\int_\Omega z^2 dx := F_\epsilon^1[z] - F_\epsilon^2[z]. \tag{3–17}$$

Similarly, suppose $\frac{d^2}{dz^2}h(z-k) \geq -\gamma$, where γ depends on function $h(.)$. We split $G[z|I]$ as

$$G[z|I] := G^1[z|I] - G^2[z|I]. \tag{3–18}$$

where

$$G^1[z|I] = \sum_{k=0}^{K-1}\int_\Omega [log\sigma_k + \frac{(c_k - I)^2}{2\sigma_k^2}]h(z-k)dx + \sum_{k=0}^{K-1}\int_\Omega \frac{\gamma}{2}[log\sigma_k + \frac{(c_k - I)^2}{2\sigma_k^2}](z-k)^2 dx \tag{3–19}$$

$$G^2[z|I] = \sum_{k=0}^{K-1}\int_\Omega \frac{\gamma}{2}[log\sigma_k + \frac{(c_k - I)^2}{2\sigma_k^2}](z-k)^2 dx. \tag{3–20}$$

We simply denote it by

$$E_\epsilon[z|I] = \{F_\epsilon^1[z] + G^1[z|I]\} - \{F_\epsilon^2[z] + G^2[z|I]\}. \tag{3–21}$$

By Proposition 3.1, $F_\epsilon^1[z]$, $F_\epsilon^2[z]$, $G^1[z|I]$ and $G^2[z|I]$ are all convex. In order to apply CCCP, we still need the the functional to be bounded below. To do that, we must assume that the image is not constant for any phase. Or in detail, we suppose each variance $\sigma_k^2 \neq 0$. Then

$$min\{\sigma_k, 0 \leq k \leq K-1\} > 0. \tag{3–22}$$

Thus, the functional $E_\epsilon[z|c^{(n)}, \sigma^{(n)}, I]$ is lower bounded.

Remark This assumption is true for most applications since if the image has been piecewise constant, we do not need to do any segmentation. Even if some phases are exactly constant, the assumption can still be true by adding some noise or spurious dots in the phases.

By Theorem 3.1, we can now use the CCCP iteration scheme via Frechet derivative, i.e.,

$$(F^1 + \lambda G^1)\ddot{(}z^{(n+1)}) = (F^2 + \lambda G^2)\ddot{(}z^{(n)}). \tag{3–23}$$

Under integration by parts (see (3–14)), the above equation is equivalent to the following PDE.

$$\begin{aligned}
&[-2\epsilon\triangle z^{(n+1)} + \frac{\pi}{\epsilon} sin 2\pi z^{(n+1)}] + \frac{2\pi^2}{\epsilon} z^{(n+1)} + \lambda[\sum_{k=0}^{K-1}(log\sigma_k + \frac{|c_k - I|^2}{2\sigma_k^2})\ddot{h}(z^{(n+1)} - k)] \\
&+\lambda\gamma \sum_{k=0}^{K-1}(log\sigma_k + \frac{|c_k - I|^2}{2\sigma_k^2})(z^{(n+1)} - k) \\
&= \frac{2\pi^2}{\epsilon} z^{(n)} + \lambda\gamma \sum_{k=0}^{K-1}(log\sigma_k + \frac{|c_k - I|^2}{2\sigma_k^2})(z^{(n)} - k),
\end{aligned} \tag{3–24}$$

where the terms in the square brackets come from the Euler-Lagrange equation of E_ϵ. Then we can use any existing method to solve z_{n+1} such as Gauss-Seidel method. Now we have all the three minimizations (3–9), (3–10) and (3–11). We can use the alternating minimization scheme as discussed at the beginning of this section.

3.2.2 Construct approximation function $h(x)$

We construct $h_\delta(x)$ in such a way that $h_\delta(z(x)-k)$ forms an ownership function of k-th phase, where $h_\delta(x)$ is defined as follows.

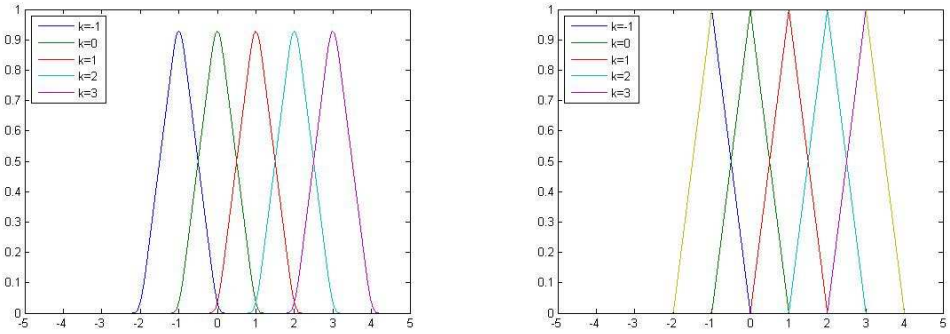
$$h_\delta(x)=\begin{cases}\frac{1+\delta}{2}+\frac{x}{2}-\frac{\delta}{\pi}cos\frac{\pi(x+1)}{2\delta}, & if|x+1|\le\delta\\ 1+x, & if-1+\delta<x<-\delta\\ \frac{2\delta}{\pi}cos\frac{x\pi}{2\delta}+1-\delta, & if|x|\le\delta\\ 1-x, & if\delta<x<1-\delta\\ \frac{1+\delta}{2}-\frac{x}{2}-\frac{\delta}{\pi}cos\frac{\pi(x-1)}{2\delta}, & if|x-1|\le\delta\\ 0, & otherwise\end{cases} \tag{3–25}$$

where δ is a parameter theoretically satisfying $0<\delta<0.5$ and should be small enough. Fig. 3-1 shows different branches of $h_\delta\circ(z(x)-k)$ for $k=-1,0,1,2,3$, where $\delta=0.2$ in (A) and $\delta=0.02$ in (B). In Fig. 3-2, a three phase example shows that

$$h_\delta(x+1)+h_\delta(x)+h_\delta(x-1)=1 \tag{3–26}$$

holds only for $x\in[-1-\delta,1+\delta]$. In application, we define $h_\delta(x):=1$ for $x\in(-\infty,-\delta)\cup(K-1+\delta,\infty)$. Then $\sum_{k=0}^{K-1}h_\delta(x-k)\equiv 1$, and so our model is a standard soft segmentation model.

Note that although the function $h_\delta(x)$ itself is not a good approximation of $1_{z=k}$ based on its graph, the composition $h_\delta\circ(z(x)-k)$ is actually a very good approximation of $1_{z=k}$ for the reason that, with the length term (3–6), the phase function $z(x)$ will ultimately only takes integer values which makes the composition $h_\delta\circ z(x)$ mostly evaluated only based on the piece $\frac{2\delta}{\pi}cos\frac{x\pi}{2\delta}+1-\delta$ since $|z(x)-k|\le\delta$ for x belonging to phase k. Thus, $h_\delta\circ(z(x)-k)$ will ultimately takes the value $\frac{2\delta}{\pi}+1-\delta$ which approximate 1 very well when δ is small enough. At the end, we want to mention that as $h_\delta(x)$ is



(A) (B)

Figure 3-1. Different branches of the constructed function.

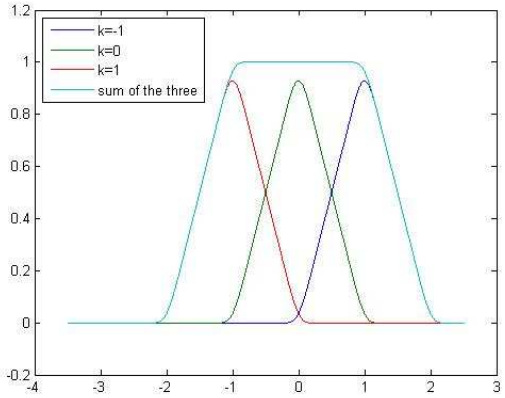


Figure 3-2. Branches of the construction functions and their sum.

defined in an explicit way, its derivative can be calculated easily. So we can finally use the iteration scheme CCCP developed in above section.

3.2.3 Segmentation Decision

Finally, once the iterations are stable, we have two ways to determine the segmentation. One way is hard segmentation, i.e., we apply to $z(x)$ the following hard thresholding decision rule:

$$x \in \begin{cases} \Omega_0, \texttt{if } \ z(x) < 0.5 \\ \Omega_{K-1}, \texttt{if } \ z(x) \geq K - 1.5 \\ \Omega_{k-1}, \texttt{if } \ k - 1.5 \leq z(x) < k - 0.5, \textit{otherwise} \end{cases} \tag{3–27}$$

When the choice of $h(x)$ makes $h_\delta \circ (z(x) - k)$ an ownership function, the model is a soft segmentation. Which scheme (hard segmentation or soft segmentation) should be chosen depends on the purpose of application. If we only want to know the pieces of segments, we should use the hard segmentation. However, in the application to partial volume estimation, we should choose the soft segmentation scheme.

3.3 Applied to Partial Volume Segmentation

Since our model is a soft segmentation and the membership functions can be viewed as ownership, we can apply our framework to partial volume estimation.

3.3.1 Introduction to Partial Volume Segmentation

One important application of multiphase segmentation is MRI brain image segmentation. There are three different tissues in human brains, i.e., the white matter, the gray matter and the cerebrospinal fluid (c.s.f). It is well known that the volume of gray matter has a close relation to some intelligence diseases. Precisely computing the volume of white matter and gray matter can help to diagnose those disease earlier. On the other hand, due to the limited spatial resolution of imaging equipment, not all voxels in the image contain a same type of tissue, especially the voxels near the tissue borders, which are highly likely contain more than one tissue types, called partial volume (PV). Fig.(3-3) shows the principle of partial volume formulation (which is originally used by [60]). The left image contains two phases with higher resolution. Due to lower resolution (half of the left image in each dimension), four squares in the left image contribute to one square in the right image. As a result, the right image contains more phases (for this example, it contains four different phases). When all four subsquares with a same phase in the left image contribute to one square in the right image, the resulted square with lower resolution will be still the same phase as original one, called *pure matter*. However, when the four subsquares contain different phases, the resulted square will present a phase looks like between the original two phases. In this case, the resulted square is called *partial volume*. The intensity of the partial volume is a weighted average

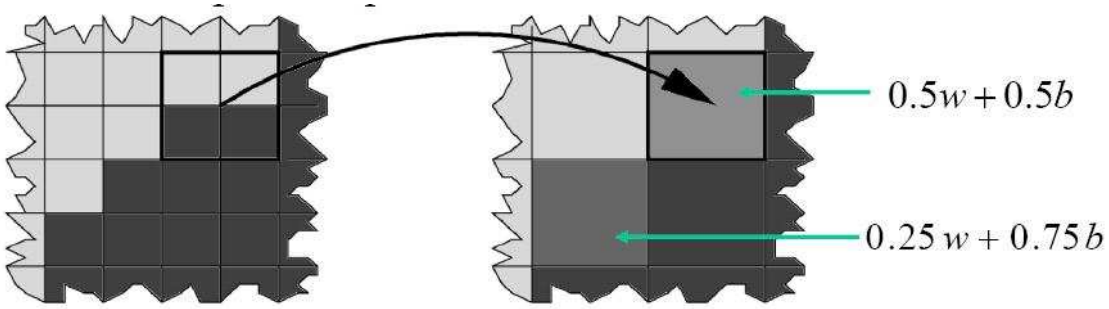


Figure 3-3. Formulation of Partial volume.

of original pure matters with combination ratios depending on the number of subsquares of each pure matter in the original image forming the partial volume.

This kind of partial volume may cause a big error in the estimation of pure tissue volumes. The error is sometimes as big as 40-60 percent [75]. Thus, partial volume segmentation of MRI images has now received considerable attention. The ideal partial volume segmentation should contain two aspects: finding partial volume and deciding its combination ratio of different pure matters. Most recent work on partial volume segmentation are based on statistical principal, e.g., the expectation-maximization (EM) method [42, 60]. These methods improved the precision of pure matter estimation. However, they do not contain length term, which makes them sensitive to noise. In [93], the author applied a reparameterized level set algorithm to partial volume segmentation. The method does include the length term. It takes the partial volume part as separated classes that are composition of pure tissues. The drawback of the paper is to use fixed ratios (e.g., 50%) of combinations for partial volumes.

3.3.2 Apply the framework to Partial Volume Segmentation

We can now apply the proposed framework to partial volume segmentation for brain MR images with mixed Gaussian distribution. In order to apply the proposed framework to partial volume segmentation, we treat the brain image as three different phases: white matter, gray matter and pure c.s.f. We calculate the pure matter volumes in a natural

way:

$$Volume(phase(k)) = \int_{\Omega} p_k(x)dx = \int_{\Omega} h_\delta(z(x) - k))dx \tag{3–28}$$

3.4 Experiments and Discussion

Since this paper focus on soft segmentation and applications to partial volume segmentation, our experiments are mostly applied to MRI brain images although a natural image is also presented which is not very close to piecewise constant. Because this work is related to the Sine-Sinc model [88] (where the author proposed a multiphase segmentation model based on phase-transition theory but it is not a soft model), we first give a comparison between our model and the Sine-Sinc model in Fig. 3-4 with synthetic image and MRI brain image, where (A) is the original artificial synthetic image with Gaussian noise ($m = 0, \sigma = 0.02$), (B) is the segmentation using the Sine-Sinc model ($\lambda = 150$), (C) is segmentation using the proposed model ($\lambda = 150, \delta = 0.2$), (D) is the original brain MR image with Gaussian noise($m = 0, \sigma = 0.005$), (E) is the segmentation using the Sine-Sinc model ($\lambda = 150$), and (F) is the segmentation using the proposed model ($\lambda = 150, \delta = 0.2$). Then in Fig. 3-5, we show three membership functions of a natural image, where (A) is the original image and (B)-(D) are membership functions, whose value is between 0 and 1.

The rest of the examples give a comparison between adaptive fuzzy c-mean method [81] and the proposed model. Fig. 3-6 is the segmentations for a synthetic biased image which is first used by X. Bresson and T. Chan[15]. The contour of segmentations in the third column is obtained by thresholding the membership functions (same thing is true for next two figures). First line is of AFCM model; second line is of the proposed framework. Middle column shows the soft segmentations. Right column shows the hard segmentations. Fig. 3-7 and Fig. 3-8 show different results for MRI brain images in a similar way to Fig.(3-5).

Finally, we apply our model to partial volume segmentation using simulated brain images. Then compare the ground truth of pure matters and our segmentation results.

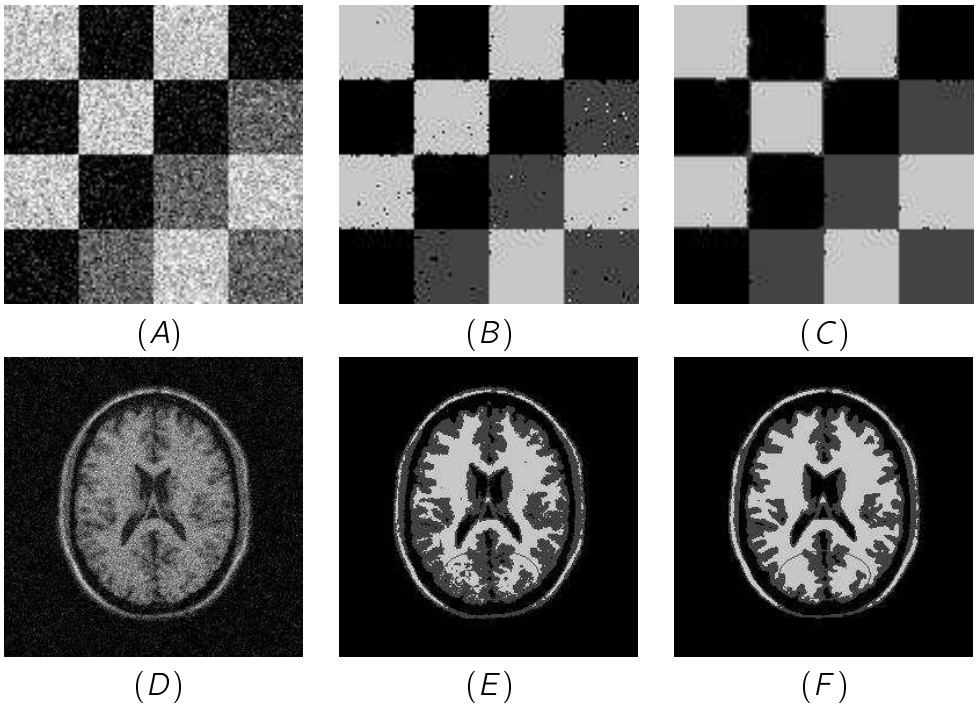


Figure 3-4. Comparison with noisy image.

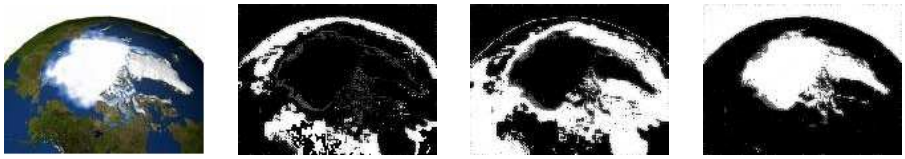


Figure 3-5. Soft segmentation.

We calculate the errors of partial volume estimation in two ways. One way is based on hard segmentation. Another way is based on soft segmentation.

Fig. 3-9 is a comparison between the ground truth of the original simulated brain MR image and the membership functions obtained using the proposed framework. Fig. 3-9(A) is the original simulated noised image. The corresponding ground truth of white matter, gray matter, and C.S.F are shown in Fig. 3-9(B), Fig. 3-9(C), and Fig. 3-9(D), respectively. Phase membership functions are shown in Fig. 3-9(E)-(G),

We carried out the experiment with 35 consecutive 2D slices of a 3D simulated brain MR image. Then compare the errors between the Sine-Sinc model and the proposed model. As an average, the errors are shown in Table 1.

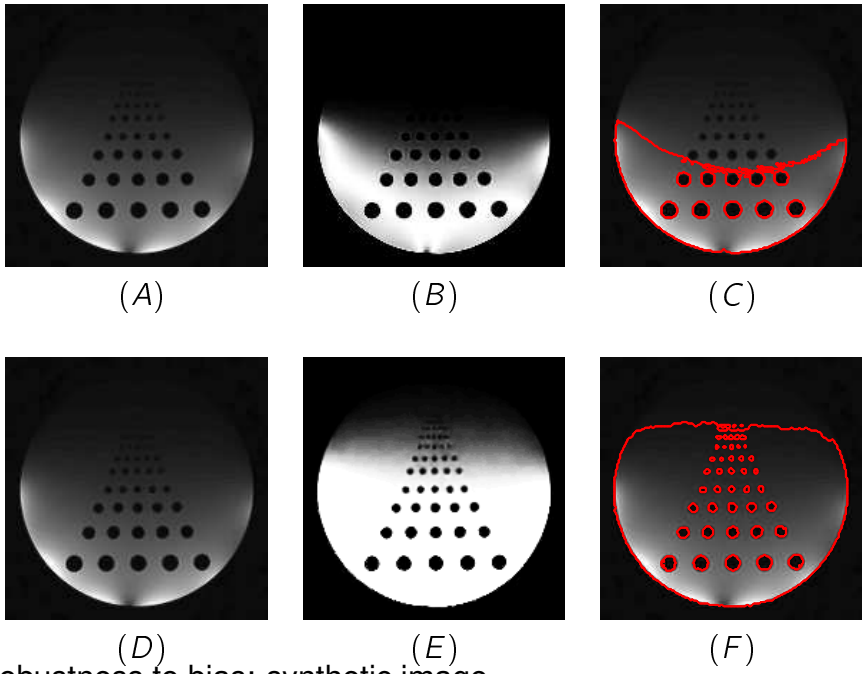


Figure 3-6. Robustness to bias: synthetic image.

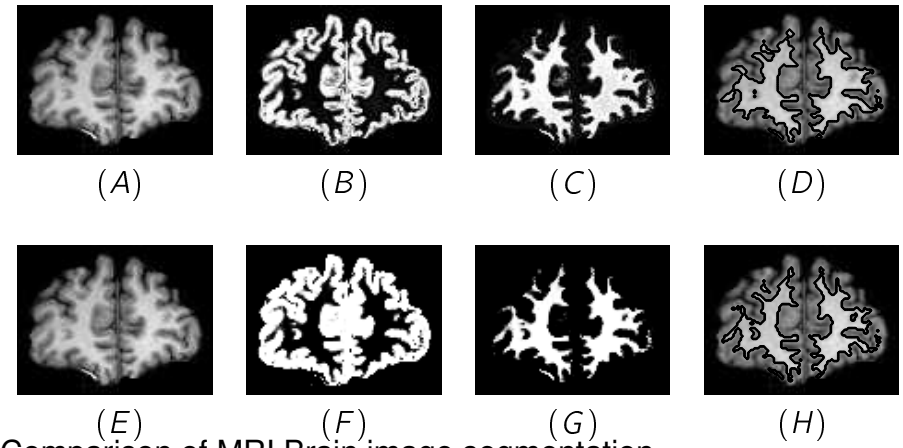


Figure 3-7. Comparison of MRI Brain image segmentation.

Table 3-1. Error Comparison

methods	white matter	gray matter
.Old Model	6.83%	7.22%
.New Model	4.68%	2.73%

Finally, we add series of Gaussian noises to the images with zero mean and different variances. Then compare their errors among the AFCM model, the proposed model with hard segmentation by thresholding, and the proposed model with soft segmentation. The errors are shown in Fig. 3-10(A). From the graph, we can see that as

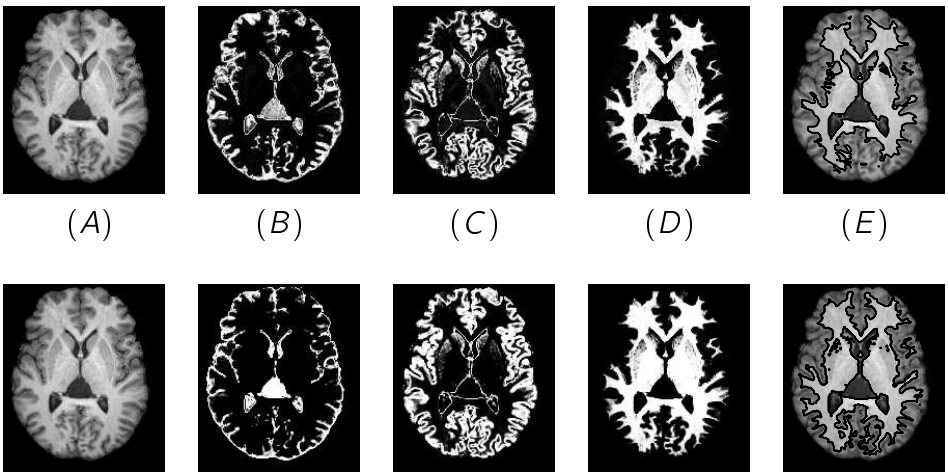


Figure 3-8. Comparison of MRI Brain image segmentation.

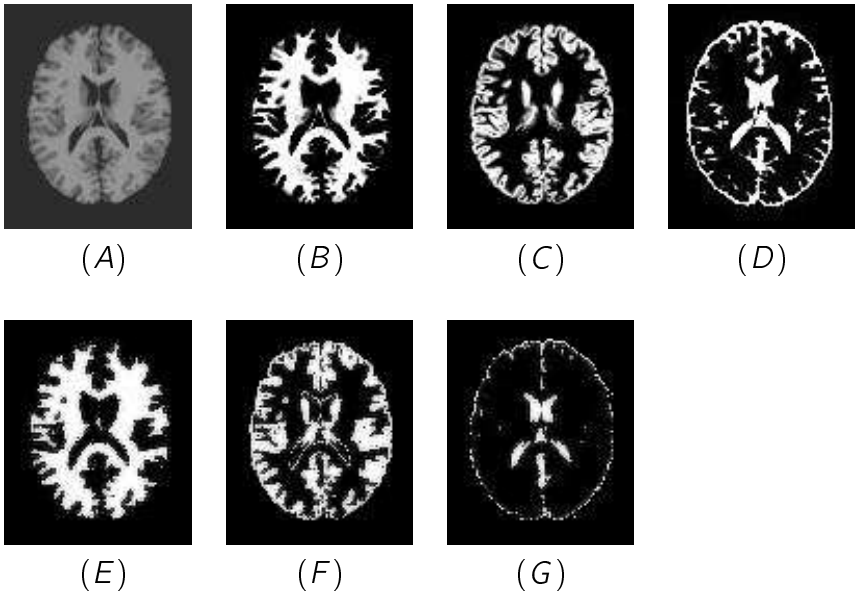
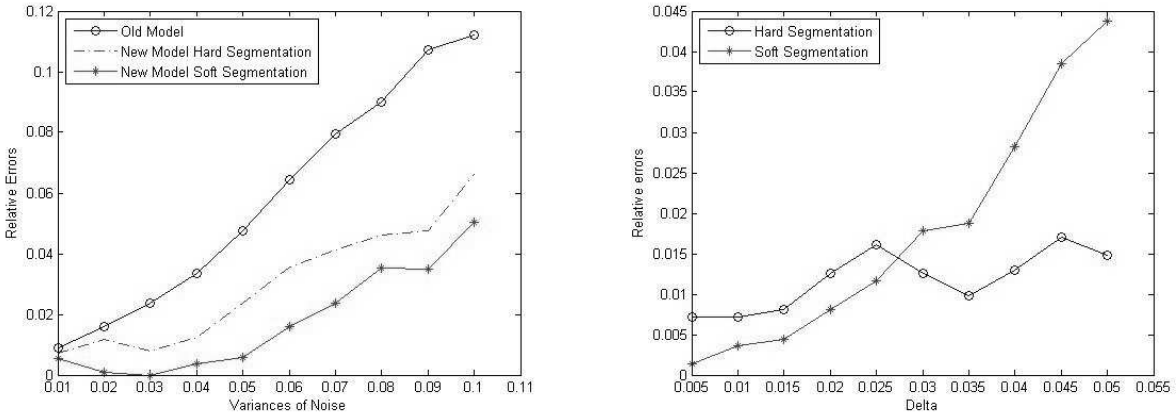


Figure 3-9. Comparison with ground truth image.

the variance of the noise rises, the error will also rise for all the three cases. However, compared with the AFCM model, the errors using the proposed model rises much more slowly as the variance of the noise rises.

We also compared the influence on the errors as the parameter δ in function $h_\delta(x)$ changes. This is shown in Fig. 3-10(B). From the graph, pure matter estimation based on phase function $z(x)$ looks a little better than the estimation based on the ownership function $h_\delta(z(x) - k)$ when δ is bigger, while the pure matter estimation based on the



(A) (B)

Figure 3-10. Error comparison.

ownership function $h_\delta(z(x) - k)$ is a little better than the estimation based on the phase function $z(x)$ when δ is smaller.

3.5 Conclusions

A critical problem in energy based multiphase segmentation is the non-convexity of the energy functional. The problem is more difficult to handle than two phase case. Level set based method is proved very successful in multiphase segmentation based on hard segmentation, but they don't work for soft segmentation since the area of different phases may be overlapped and so no clear boundaries between different phases. The soft segmentation is more useful in partial volume segmentation for brain MR images. This paper borrowed the ideas from phase transition based methods [33, 87, 88]. The key point of this paper is to construct an approximation function so that the membership functions can be obtained by its composition with phase function. With this constructed function, we can avoid to add new variables for membership functions and so it saves memory space and promotes efficiency. Moreover, since the composition of the constructed function and phase functions forms membership functions, we also avoid the general constraint problem for soft segmentation in implementation. The framework is then applied to partial volume segmentation. The future contains choosing better constructed function $h(x)$, and better discretization scheme and iteration scheme.

CHAPTER 4
A STOCHASTIC VARIATIONAL MODEL FOR MULTI-PHASE SOFT SEGMENTATION IN THE PRESENCE OF INTENSITY INHOMOGENEITY

There have been many soft segmentation methods [31, 64, 71, 72, 81, 87]. Mory and Ardon extended the original region competition model [109] to a fuzzy region competition method [71, 72]. The technique generalizes some existing supervised and unsupervised region-based models. The proposed functional is convex, which guarantees the global solution in the supervised case. Unfortunately, this method only applies to two phase segmentation and is hard to be extended to multiphase segmentation. Fuzzy C-mean (FCM) is a method developed for pattern classification and recognition, and has been applied to image segmentation [31, 64, 81]. The standard FCM model partitions a data set $\{x_k\}_{k=1}^N \subset R^d$ into M clusters by the following objective function

$$J_{FCM} = \sum_{i=1}^{N}\sum_{k=1}^{M} u_{ik}\|x_i - v_k\|_2^2 \tag{4–1}$$

where $u_{ik} \geq 0$ is the membership value of datum x_i for class k with $\sum_{k=1}^{M} u_{ik} = 1$, and v_k stands for the cluster centers [9, 39]. The original FCM method is very sensitive to noise. Pham et.al proposed an adaptive fuzzy C mean (AFCM) model [81] which is more robust to noise than the standard FCM, where the constant cluster centers v_k used in the FCM model (4–1) are substituted by functions that are smooth enough and close to the corresponding cluster centers.

Another class of soft segmentations are based on stochastic approaches [34, 64, 87]. In these approaches, pixel intensities are considered as samples of one or several random variables. The advantage of stochastic method is its stronger ability to deal with random noise. In most stochastic segmentation models, the likelihood functions are used to represent the fitting term in an energy functional. It starts from the assumption that reasonable segmentation should maximize the likelihood. The method is called Maximum Likelihood (ML) method [38]. An expectation-maximization (EM) algorithm is

usually employed to solve it when data is incomplete. However, simply using likelihood to model an image is not enough since it ignored the prior knowledge of an image. In [64], a segmentation framework based on *maximum a posteriori* principle (MAP) was proposed for partial volume (PV) segmentation of MRI brain images, which is a classic application of soft segmentation.

J. Shen proposed a general multiphase stochastic variational fuzzy segmentation model combining stochastic principle and Modica-Mortola's phase-transition theory [87]. The intensity of images was modeled as a mixed Gaussian distribution. The model assumed that membership functions should be either close to 1 or close to 0, which simplified the model but limited its application. For example, it's not reasonable to apply the model to partial volume segmentation since in that case the membership functions are usually neither close to 1 nor close to 0 at the boundary of different matters.

Bias correction is an important mean in soft segmentation to deal with intensity inhomogeneity [1, 53, 81, 102]. For example, Wells et al proposed an expectation-maximization (EM) algorithm to solve the bias correction problem and the tissue classification problem [102]. The EM algorithm was used to iteratively estimate the posterior tissue class probabilities when the bias field is known, and to estimate the MAP of the bias field when tissue class probabilities are known. The disadvantage of this method is that the directly computed bias field may not be smooth which will lead to a poor bias correction and segmentation results. Pham and Prince proposed an adaptive fuzzy C-means algorithm which is formulated by modifying the objective function in the fuzzy C-means algorithm to include a multiplicative bias field, which allows the centroids of each class to vary across the image. Smoothness of the bias field is ensured by penalizing its first and second order derivatives, which leads to a computationally expensive procedure for the smoothing of the bias field. Ahmed et al proposed to add a neighborhood term that enabled the class membership of a pixel to be influenced by its neighbors [1]. The neighborhood term acts as a regularizer and forces the solution toward a piecewise

homogeneous labeling. Li et al. proposed a variational level set based method for medical image segmentation and bias correction [53], the smoothness of the bias field is intrinsically ensured by the data term in the variational formulation.

In this chapter, we proposed a stochastic variational model for multi-phase soft segmentation in the presence of noise and intensity inhomogeneity, where the image intensity at each point is modeled as a mixed Gaussian distribution with means and variances to be optimized. Different from J. Shen's work [87], our model does not set the assumption that membership functions must be close to either 1 or 0. So, our model are more suitable for soft segmentation and application to partial volume analysis. Since our model is developed based on the assumption that the image intensity is a mixed Gaussian distribution with possibly different variances for different phases, it is also different from [64, 81] in that our model adaptively corrects bias of intensities and removes noise by finding optimized means and variances. It is demonstrated by experiments that our model is not only robust to noise, but also robust to bias. The rest of the chapter is organized as follows. The new model is developed in Section 4.1. The numerical implementation scheme is presented in Section 4.2. In Section 4.3, we show some experiment results and also give some explanation and analysis. Both synthetic images and authentic images are used.

4.1 Model Development

Let $I(x)$ be a 2-D image defined on an open bounded domain Ω containing K phases. Let w be phase label variable (i.e., $w(x) \in \{1, \cdots, K\}$ for all $x \in \Omega$). At each pixel x, both $w(x)$ and $I(x)$ are viewed as random variables indexed by x. The probability that x belongs to the i-th phase is represented by the ownership functions $p_i(x)$, $1 \leq i \leq K$. If we denote the probability density function (PDF) of the random variable $I(x)$ given that x belongs to the i-th phase by $Prob(I(x)|w(x) = i)$, then the

random variable $I(x)$ follows a mixed distribution with the following PDF:

$$\sum_{i=1}^{K} Prob(I(x)|w(x)=i)p_i(x). \tag{4–2}$$

Suppose in further that $Prob(I(x)|w(x)=i)$ is a Gaussian PDF for each $i=1,\ldots,K$ and all random variables $\{I(x):x\in\Omega\}$ are independent. Then the likelihood (joint PDF) is

$$\prod_{x\in\Omega}\sum_{i=1}^{K} g(I|u_i(x),\sigma_i)p_i(x). \tag{4–3}$$

where

$$g(x|\mu,\sigma)=\frac{1}{(\sqrt{2\pi}\sigma)}exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{4–4}$$

The negative log-likelihood is

$$E[I|P,U,\sigma]=-\int_{\Omega} log\left(\sum_{i=1}^{K} g(I|u_i(x),\sigma_i)p_i(x)\right), \tag{4–5}$$

where $P=[p_1,p_2,\ldots,p_K]$, $U=[u_1,u_2,\ldots,u_K]$, and $\sigma=[\sigma_1,\sigma_2\ldots,\sigma_K]$. We assume that is a constant inside each phase, i.e., $u_i(x)=c_i$, which leads to the following energy functional

$$E_F(p,c,\sigma)=\lambda\int_{\Omega}-log[\sum_{i=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_i}exp(-\frac{(I(x)-c_i)^2}{2\sigma_i^2})p_i]dx. \tag{4–6}$$

By adding the total variation of $p_i(x)$ to $E_F(p,c,\sigma)$ to enforce smoothness of membership functions $p_i(x)$, we have

$$\begin{aligned} E_{FR}(p,c,\sigma) =& \lambda\int_{\Omega}-log[\sum_{i=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_i}exp(-\frac{(I(x)-c_i)^2}{2\sigma_i^2})p_i]dx+\sum_{i=1}^{K}\int_{\Omega}|\nabla p_i|dx \\ \triangleq& -\lambda\int_{\Omega} log(\sum_{i=1}^{K} f_i(x)p_i(x))dx+\sum_{i=1}^{K}\int_{\Omega}|\nabla p_i|dx \end{aligned} \tag{4–7}$$

Remark. We want to mention that our model is not the first time to use Gaussian distribution. On the contrary, the Gaussian distribution has been introduced to many

segmentation models, such as graph cut [14] and soft Mumford-Shah model [87]. The difference between the proposed model and other models lies in the fact that all other models assume different Gaussian distributions sharing a same variance which is usually fixed. However, in our model, different Gaussian distributions can be different, which makes the model more flexible.

4.2 Numerical Implementation

Note that the energy functional is convex with respect to all its variables except for variances. For fixed variances, global minimization can be achieved for any initialization. With gradient descend method, the means and variances' iteration scheme is as follows.

$$c_i^{(n+1)} = c_i^{(n)} + t_c \left[\frac{\lambda}{\sigma_i^2} \int_\Omega \frac{f_i p_i (I - c_i)}{\sum f_i p_i} dx \right]^{(n)} \tag{4–8}$$

$$\sigma_i^{(n+1)} = \sigma_i^{(n)} + t_\sigma \lambda \left[\int_\Omega \frac{f_i p_i ((I - c_i)^2 - \sigma_i^2)}{2\sigma_i^4 \sum_{i=1}^K f_i p_i} dx \right]^{(n)} \tag{4–9}$$

The challenge in the implementation is the optimization of membership functions $p_i(x)$ because of the constraints

$$1 \geq p_i(x) \geq 0 \; and \; \sum_{i=1}^K p_i(x) = 1 \tag{4–10}$$

which requires $p = (p_1, p_2, \ldots, p_K)$ lies in the simplex Δ_{K-1}. There have been two ways to deal with the simplex constraint. One is to use Lagrangian multiplier method (or augmented Lagrangian multiplier method) for $\sum_{i=1}^K p_i(x) = 1$, and add an exact penalty term for each $0 \leq p_i(x) \leq 1$ (see [10], [30] and [95]). The drawback of Lagrangian multiplier method is its low convergence rate. The so-called *exact penalty term* is exact only under some constraint and is not differentiable at end points, and must be replaced by a smoothed version for approximation which will affect the exactness. Another way to deal with the simplex constraint is to use the Euler-Lagrangian equation of the unconstraint problem for iterations and then project the result to simplex Δ_{K-1} at each

iteration step [87]. The drawback of this method is that no general analytic expression can be written for all dimensions. For different dimensions, the projection functions are different, and need to be written in a different way. Especially, when the dimension is greater than three, the projection function becomes complicated, which leads to a low efficiency in both coding and implementation. In this paper, we give a novel way of projection using dual method. The projection can be expressed explicitly and uniformly for all dimensions, and the analytic property is guaranteed due to dual theory.

Dual method has been extensively studied to deal with total variation which is not differentiable at points where the first order variation is zero. One of the popular example is Chambolle dual method [25]. Recently, M. Zhu and T. F. Chan developed a new algorithm combining the gradient decent method and dual method, called Primal-Dual-Hybrid-Gradient method (PDHG) (see [108] for details). The method integrates the advantages of both gradient method and dual method, and thus faster than using either method. It is proved to be faster than using dual method only and its modified iteration form is guaranteed to converge when step size satisfies some condition (see [18], [41] and [82]). In our application, we adopted the ideal of PDHG and apply it to our model with constraint on simplex Δ_{K-1}.

4.2.1 Primal Dual Hybrid Gradient Algorithm

In order to make the paper self-contained and easy to read, we give a review on the PDHG algorithm. Let f be the given image defined in a rectangular domain Ω with size $m \times n$, and u be the restored image. Then the ROF model to find an optimal restored image is to solve the following minimization problem

$$\min_u \int_\Omega |\nabla u| + \frac{\lambda}{2}\|u - f\|^2 = TV(u) + \frac{\lambda}{2}\|u - f\|^2 \tag{4–11}$$

where $TV(u)$ is the total variation of u. The discrete form of the TV(u) can be written as

$$TV(u) = \sum_{i=1}^{m}\sum_{j=1}^{n} \|\nabla u(i,j)\|_2 \tag{4–12}$$

where $\|u(i,j)\|_2 = \sqrt{(\nabla u^1(i,j))^2 + (\nabla^2 u(i,j))^2}$ and

$$\nabla u^1(i,j) = \begin{cases} u(i+1,j) - u(i,j) & if\ i < m \\ 0 & if\ i = m \end{cases} \tag{4–13}$$

$$\nabla u^2(i,j) = \begin{cases} u(i,j+1) - u(i,j) & if\ j < n \\ 0 & if\ j = n \end{cases} \tag{4–14}$$

Equation (4–11) is called primal ROF model. Reordering the image matrix u (resp. f) row-wisely into a vector y (resp. z) and using the matrix A_l for the gradient operator at element l, we get the following discrete form of the primal ROF model

$$\min_{y \in R^{mn}} \sum_{l=1}^{mn} \|A_l^T y\| + \frac{\lambda}{2}\|y - z\|^2 \tag{4–15}$$

Using the dual form of the TV norm

$$\sum_{l=1}^{mn} \|A_l y\| = max_{x \in X} \sum_{l=1}^{mn} \langle y, A_l x \rangle \tag{4–16}$$

Then the primal-dual ROF model reads as

$$\begin{aligned} \min_{y \in R^{mn}} \max_{x \in X} \Phi(y,x) :=& y^T A x + \frac{\lambda}{2}\|y - z\|^2 \\ =& \max_{x \in X} \min_{y \in R^{mn}} y^T A x + \frac{\lambda}{2}\|y - z\|^2 \end{aligned} \tag{4–17}$$

where $x = [x_1, \ldots, x_{mn}]^T$, $Ax = \langle A_1 x, A_2 x, \ldots, A_{mn} x \rangle$. The equality holds because of min-max theorem. The primal dual hybrid gradient algorithm can be described as follows.

1. Dual Step

Fix $y = y^k$, apply one step of (projected) gradient ascent method to the maximization problem

$$\max_{x \in X} \Phi(y^k, x). \tag{4–18}$$

The ascent direction is $\nabla_x \Phi(y^k, x) = A^T y^k$, so we can update x as

$$x^{k+1} = P_X(x^k + \tau_k \lambda A^T y^k) \tag{4–19}$$

where τ_k is the (dual) stepsize and P_X denotes the projection onto the set X:

$$P_X(z) = arg \min_{x \in X} \|x - z\| \tag{4–20}$$

In this case, it can be calculated simply as (for each l)

$$(P_X(z))_l = \frac{z_l}{max\{\|z_l\|, 1\}} \tag{4–21}$$

2. Primal Step

Fix $x = x^{k+1}$, apply one step of gradient descent method to the minimization problem

$$y^{k+1} = \min_{y \in R^{mn}} \Phi(y, x^{k+1}) \tag{4–22}$$

The descent direction is $-\nabla_y \Phi(y, x^{k+1}) = -(Ax^{k+1} + \lambda(y^k - z))$ and therefore the update is

$$y^{k+1} = y^k - \theta_k(\frac{1}{\lambda} Ax^{k+1} + \lambda(y^k - z)) \tag{4–23}$$

where θ_k is the (primal) step size.

4.2.2 Optimize membership functions using PDHG

By the principle of PDHG, to minimize (4–7) with respect to membership functions p_i ($i = 1, \ldots, K$) under constraint (4–10), it is equivalent to solve the following discrete min-max problem

$$\max_{q \in X^K} \min_{p \in \Delta_{K-1}} -\langle p, Dq \rangle - H(p) \tag{4–24}$$

where $p = (p_1, \ldots, p_K)$, $q = (q_1, \ldots, q_K)$ and

$$H(p) = \lambda \sum_{i,j=1}^{m,n} log\langle f(i,j), p(i,j) \rangle = \lambda \sum_{i,j=1}^{m,n} log \sum_{k=1}^{K} f_k(i,j) p_k(i,j).$$

The descent direction for $min_{p\in R^K} -\langle p, Dq\rangle - H(p)$ is $Dq + \nabla_p H(p)$. So, the evolution of membership p (primal step) is

$$p^{(n+1)} = P_{\Delta_{K-1}}(p^{(n)} + \tau_n(Dq^{(n)} + \nabla H(p^{(n)}))) \tag{4–25}$$

where $P_{\Delta_{K-1}}(v)$ is the projection to the simplex defined by

$$P_{\Delta_{K-1}}(x) \triangleq min_{z\in\Delta_{K-1}}\|z - x\| \tag{4–26}$$

for $\forall x \in R^K$, where $\|\cdot\|$ denotes the Euclidean distance. We will see a novelty method for the projection to simplex in the next section.

Since the first variation of (4–24) with respect to q_i is Dp_i, the dual step is

$$q_i^{(n+1)} = P_{X^K}(q_i^{(n)} + \theta^{(n)} Dp_i^{(n)}) \tag{4–27}$$

where P_{X^K} is the projection to space T defined by

$$[P_{X^K}(x)]_l = \frac{x_l}{max\{\|x\|_2, 1\}}. \tag{4–28}$$

where l denotes the number of component of a vector.

Therefore, the bi-direction projected PDHG algorithm for minimizing energy functional $(4 - -7)$ is given by

$$\begin{cases} p^{(n+1)} = P_{\Delta_{K-1}}(p^{(n)} + \tau_n(Dq^{(n)} + \nabla H(p^{(n)}))) \\ q^{(n+1)} = P_{X^K}(q^{(n)} + \theta^{(n)} Dp^{(n)}) \end{cases} \tag{4–29}$$

where the projection to simplex Δ_{K-1} is given explicitly in next section.

4.2.3 Projection to simplex Δ_{K-1}

Now, we want to deduce the explicit solution for $P_{\Delta_{K-1}}(x)$. Let x^* denote the solution. Define

$$\Pi_{\Delta_{K-1}}(x) = \begin{cases} 0 \quad if \quad x \in \Delta_{K-1} \\ \infty \quad otherwise \end{cases} \tag{4–30}$$

Then by Moreau's Identity, we have

$$\begin{aligned} x^* = P_{\Delta_{K-1}}(x) \triangleq min_{z\in\Delta_{K-1}}\|z-x\| = arg\ min_{x\in R^K}\left\{\frac{1}{2}\|x-y\|^2 + \Pi_{\Delta_{K-1}}(x)\right\} \\ = y - arg\ min_{z\in R^K}\left\{\Pi^*_{\Delta_{K-1}}(z) + \frac{1}{2}\|z-y\|^2\right\} \end{aligned} \tag{4–31}$$

where $\Pi^*_{\Delta_{K-1}}(z)$ is the Legendre-Fenchel transform of $\Pi_{\Delta_{K-1}}$. By definition, we have

$$\begin{aligned} \Pi^*_{\Delta_{K-1}}(z) =& sup_{x\in R^K}\left\{\langle x, z\rangle - \Pi_{\Delta_{K-1}}(x)\right\} \\ =& sup_{x\in\Delta_{K-1}}\langle x, z\rangle = max_{1\le i\le K} z_i \triangleq t \end{aligned} \tag{4–32}$$

So, we finally get the solution

$$x_i^* = y - z_i^* = max\ (y_i - t, 0), \quad i = 1,\ 2,\ ,\ ...,\ K \tag{4–33}$$

which is simple and very fast.

4.3 Results

Since the proposed model is an extension of [64], we present all experiment results compared with the model developed in paper [64] (We temporarily call this model MAP-AFCM model).

The first experiment aims at testing robustness to noise. In Fig. 4-1, the original image contains obviously three phases. We added a mixed Gaussian noise with zero mean and an overall variance 0.03. First, we applied MAP-AFCM model. We choose $\lambda_1 = 5$, and stop iterations using criterion $max_{1\le i\le 3}\{|c(i)_{new} - c(i)_{old}|\} < 0.001$, where $c(i)_{old}$ denotes the old mean before each iteration, and $c(i)_{new}$ denotes the new mean after each iteration (the same for the rest experiments). Then we applied our model (4–7) to the image. Obviously, the result of the new model is much better, where the first line shows the segmentation using MAP-AFCM, the second line shows the segmentation using the proposed model. For each line, from left to right are original image, three membership functions and hard segmentation after thresholding.

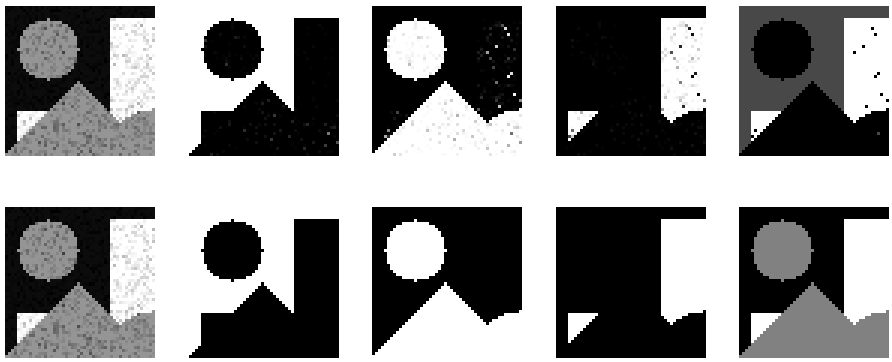


Figure 4-1. Experiment 1: Robustness to noise.

Discussion: This big difference comes from the difference of the fitting terms in two models. Note that in MAP-AFCM model, to make the fitting term small enough, the image intensity at each point must be very close to the mean of its phase. Thus it is sensitive to noise. Comparatively, in Model (4–7), the effect of isolated noise to the energy functional can be counteracted by the variances appeared in the denominators of the fitting term. So the new model is more robust to noise.

The second experiment aims at comparing robustness to bias. In Fig. 4-2, the first line is the original biased image and its ground truth of all three membership functions. The second line and the third line show the soft segmentations obtained using MAP-AFCM model and the proposed model, respectively. Obviously, the proposed model gives more precise result compared with the ground truth since there is no bias in the segmentation.

Our third experiment aims to give a comparison between variances fixed and variances updated in the new model. For all the five lines, from left to right are the original image, three membership functions and hard segmentation, respectively. From the first line to the fourth line are the results with variances fixed. For example, we set $\sigma_i^2 = 0.005$ for all $(1 \leq i \leq 3)$ in the first line, and we set $\sigma_i^2 = 0.010$ for all $(1 \leq i \leq 3)$ in the second line, and so on. However, the last line is the result where variances are updated, and we obtained the final variances for the three phases, which are $\sigma_1 = 0.0069, \sigma_2 = 0.0193$, and $\sigma_3 = 0.0135$, respectively. Obviously, the last row

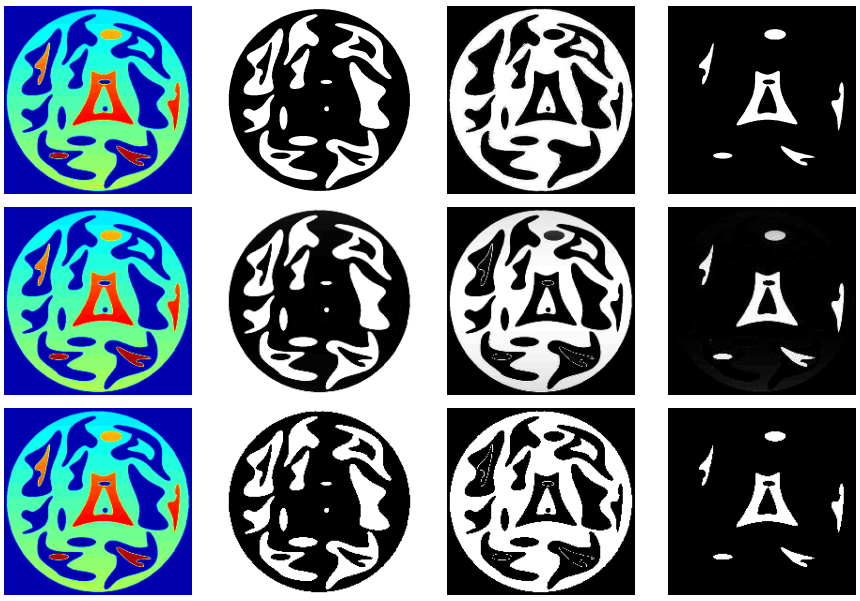


Figure 4-2. Experiment 2: Robustness to bias.

gives the best result. This experiment shows that updating variances is better than fixing variances and assuming all of them are equal. Since FCM (fuzzy c-mean) model is a special case when all variances are fixed and the same, this experiment shows that the proposed model outperforms FCM model.

Finally, we test our model using real images. In Fig. 4-4, the liver is not very clear due to the existence of bias. Using MAP-AFCM leads to a wrong result where a big part of the liver was incorrectly classified to background as shown in the first line. This can be easily seen from the hard segmentation. However, using the proposed model can get much better result as shown in the second line. This is because the fitting term in the model contains bias, as well as variance. By calculating the variances of the three phases, they are $0.013, 0.011$ and 0.002, respectively. This fact also proves that it is reasonable to assume that different phases may have different variances as in our model. In the figure, the first line shows the segmentation using MAP-AFCM model, the second line shows the segmentation using the proposed model. From left to right: original image, three membership functions and hard segmentations, respectively.

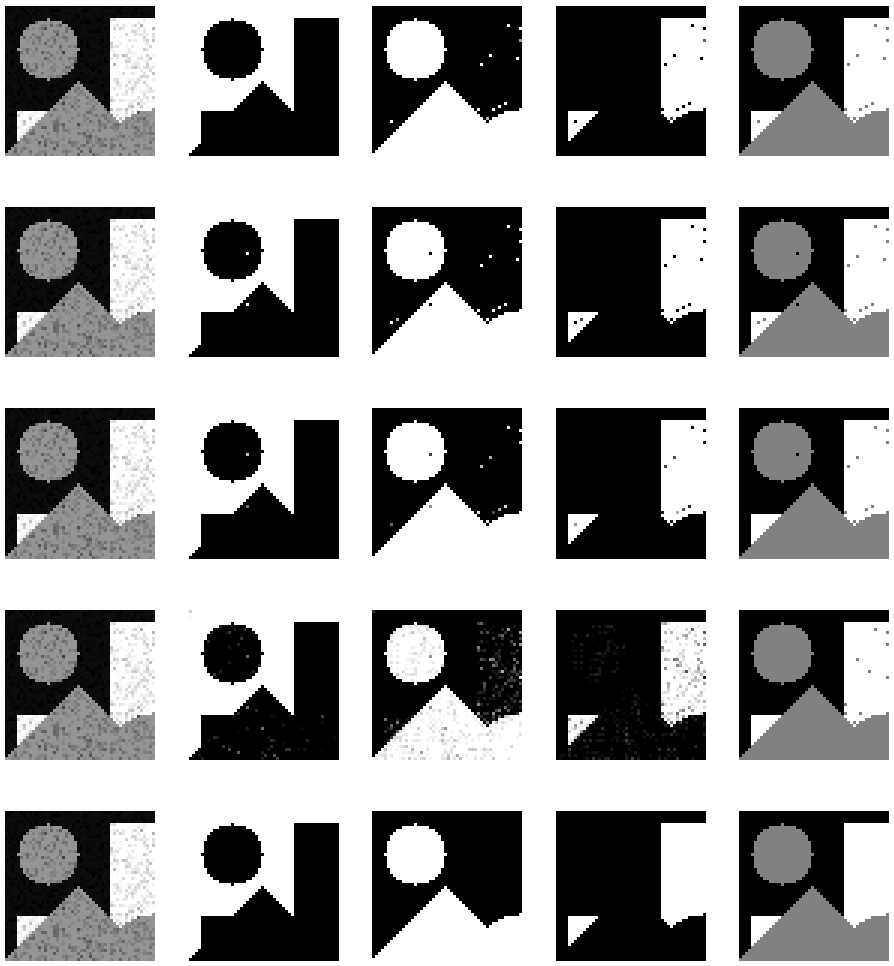


Figure 4-3. Comparison between variances fixed and updated.

Figure 4-4. MRI liver segmentation.

As we mentioned at the beginning of the paper, one of the most important application of soft segmentation is partial volume segmentation of MRI brain images. Fig. 4-5 gives a comparison in MRI brain image soft segmentation, where the first line shows the segmentation using MAP-AFCM model, the second line shows the segmentation using the proposed model. From left to right: original image, three membership functions (white matter, gray matter and CSF (cerebrospinal fluid)) and hard segmentations, respectively.. There is a big difference between the soft segmentations (the membership functions). By using MAP-AFCM model, most pixels are classified to be partial volume, i.e., its intensity is neither close to 1, nor close to 0 (In the figure, brightness of intensity means close to 1, darkness means close to 0, and intensity between brightness and darkness means partial volume). However, this is not true because it is well known that partial volume of MRI brain image should appear mostly often at the boundary of different tissues. Comparatively, using the proposed model can get more reasonable results, where the partial volume only appears at the boundary of different tissues.

We also present some natural images for comparison. In Fig. 4-6, the left image is the original image, and the middle one and the right one are hard segmentations after thresholding using MAP-AFCM model and the proposed model, respectively, where from left to right: original image and three phases of hard segmentations. Line 1 is of MAP-AFCM model and Line 2 is of the proposed model. In Fig. 4-7, we present all three phases of hard segmentations after thresholding using different models. In Fig. 4-8, the first column is the original image. We present all membership functions and hard segmentations for readers to compare. Line 1 is the membership functions using MAP-AFCM model, Line 2 is the membership functions using the proposed model, Line 3 is the hsrd segmentations using MAP-AFCM model, and Line 4 is the hard segmentations using the proposed model. For all three examples, the results using our model are all better than using AFCM model.

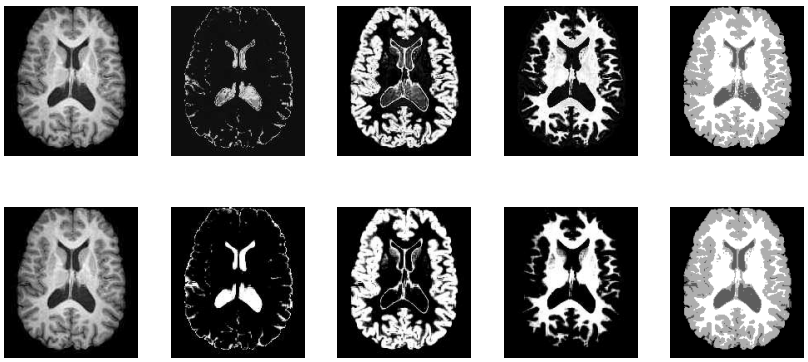


Figure 4-5. MRI brain image segmentations.

Figure 4-6. Natural image segmentation after thresholding.

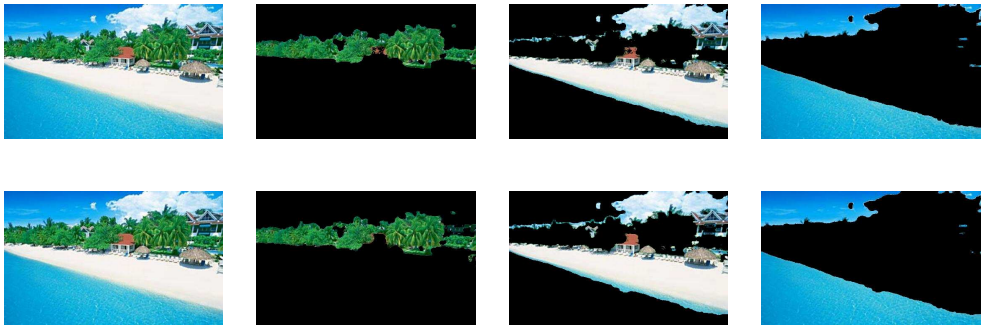


Figure 4-7. Natural image segmentation after thresholding.

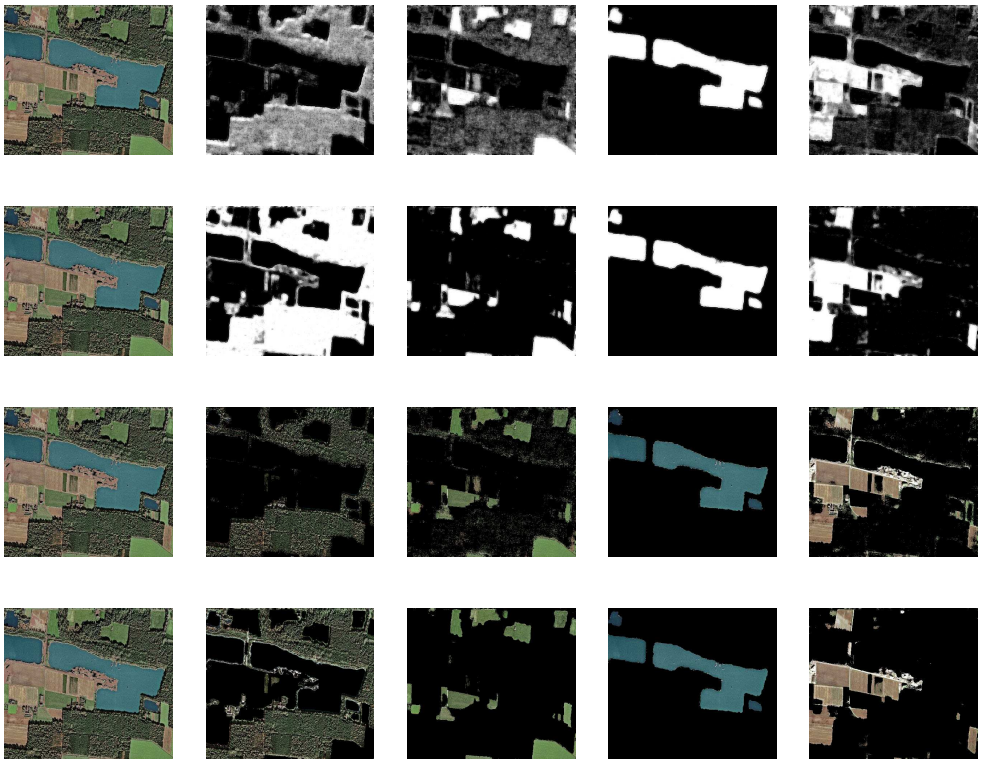


Figure 4-8. Natural image segmentation after thresholding.

CHAPTER 5
SEMI-SUPERVISED MULTIPHASE IMAGE SEGMENTATION

In a supervised segmentation model, the features obtained by learning procedure is a significant input, while in a semi-segmentation model or an unsupervised segmentation model, the input data of the model can only be the given image. Supervised segmentation, although takes a lot of time for training a large selected set of data, can perform a segmentation very fast as long as the learned features are good enough. A supervised segmentation model usually works efficiently for a specific kind of images, while a semi-segmentation model can work out a good result for any kind of images but takes more time for each implementation.

Interactive image segmentation is a technique using human interfering, usually combined with semi-supervised image segmentation. Since at each time of semi-supervised segmentation, some regions must be assigned to specific classes before each implementation is carried out. The performance of the segmentation is hereafter depends on the assignment significantly. Under the assumption that the assignment to each class at the beginning is correct, it is generally true that the more the regions are assigned , the better the result should be. Therefore, a good semi-segmentation model is such one that needs less assignment of the region but still gives a good enough result. Meanwhile, since we can't guarantee that if the result will be good enough or not during an assignment, an interactive interfering is usually needed in such a procedure to improve the result. Such a technique is called interactive image segmentation.

Image matting is a 2-D interactive semi-supervise image segmentation technique. It belongs to soft segmentations. In digital matting, a foreground element is extracted from an image by estimating a color and opacity for the foreground element at each pixel. The opacity value at each pixel is typically called its *alpha*, and the *opacity* image, taken as a whole, is referred to as the *alpha matte* (between 0 and 1) or *key*. Matting is used in order to composite the foreground element into a new scene. Matting and compositing

Figure 5-1. Image matting.

were originally developed for film and video production. The most popular image matting methods contain Poisson Matting [90], Bayesian Matting [35] and Spectral Matting [61].

Formally, image matting methods take as input an image I, which is assumed to be a composite of a foreground image F and a background image B. The color of the i-th pixel is assumed to be a linear combination of the corresponding foreground and background colors,

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \tag{5–1}$$

where α_i is the pixels foreground opacity. In natural image matting, all quantities on the right hand side of the equation are unknown. Obviously, this is a severely under-constrained problem, and user interaction is required to extract a good matte. Most recent methods expect the user to provide a trimap as a start. Such an example is shown in Figure 5-1. The trimap is a rough (typically hand-drawn) segmentation of the image into three regions: foreground (shown in white), background (shown in black) and unknown (shown in gray). Given the trimap, these methods typically solve for F, B, and α simultaneously. This is typically done by iterative nonlinear optimization, alternating the estimation of F and B with that of α. In practice, this means that for good results the unknown regions in the trimap must be as small as possible.

Just like image matting, other supervised semi-supervised image segmentations are mostly developed for two-phase images. In this chapter, we developed a new semi-supervised multi-phase image segmentation frame work based on the model studied in Chapter 4.

5.1 From Unsupervised Segmentation to Semi-Supervised Segmentation

In this section, we introduce a semi-supervised image segmentation framework based on piecewise smoothed soft Mumford-Shah model.

Given a source image $I(x)$, we assume the image contains N classes. Let $u_i(x)$ denote the i-th pattern and $p_i(x)$ be the i-th membership function. A piecewise smoothed soft Mumford-Shah model is defined as follow:

$$\begin{aligned} E(u;p) = &\sum_{i=1}^{N} \frac{1}{2} \int_{\Omega} (I(x) - u_i(x))^2 p_i(x) dx \\ &+ \sum_{i=1}^{N} \lambda \int_{\Omega} |\nabla u_i(x)| dx + \sum_{i=1}^{N} \mu \int_{\Omega} |\nabla p_i(x)| dx \end{aligned} \tag{5–2}$$

The iterations based on fast gradient-descent method are

$$\begin{cases} \dfrac{\partial E}{\partial u_i} = -\lambda div(\dfrac{\nabla u_i}{|\nabla u_i|}) + (u_i - I)p_i \\ \dfrac{\partial E}{\partial p_i} = -\mu div(\dfrac{\nabla p_i}{|\nabla p_i|}) + (u_i - I)^2 \end{cases} \tag{5–3}$$

The primal-dual form of (5–2) with respect to u is

$$\min_{u} \max_{|v| \le 1} E(u, v; p) = \sum_{i=1}^{N} \frac{1}{2} \int_{\Omega} (I(x) - u_i(x))^2 p_i(x) dx + \sum_{i=1}^{N} \lambda \int_{\Omega} u_i(x) div\ v_i dx \tag{5–4}$$

The primal-dual form of (5–2) with respect to p is

$$\min_{p \in \Delta_{N-1}} \max_{|q| \le 1} E(u; p, q) = \sum_{i=1}^{N} \frac{1}{2} \int_{\Omega} (I(x) - u_i(x))^2 p_i(x) dx + \sum_{i=1}^{N} \lambda \int_{\Omega} p_i(x) div\ q_i dx \tag{5–5}$$

The iteration on u and v is

$$\begin{cases} \dfrac{\partial u_i}{\partial t} = -[(u_i - I)p_i + \lambda div\ v_i] \\ \dfrac{\partial v_i}{\partial t} = -\lambda \nabla u_i \end{cases} \tag{5–6}$$

The iteration on p and q is

$$\begin{cases} \dfrac{\partial p_i}{\partial t} = -\left[(u_i - I)^2 + \mu div\ q_i\right] \\ \dfrac{\partial q_i}{\partial t} = -\mu \nabla p_i \end{cases} \tag{5–7}$$

In an interactive semi-supervised image segmentation, an image is assumed to include two parts: the known part $\Omega_i,\ i = 1, 2, \ldots, N$ and the unknown part Ω_U. Only the unknown part needs to be applied for segmentation. The above model is then to find segmentation only for domain Ω_U, i.e., to minimize the following energy functional:

$$\begin{aligned} E(p) = &\sum_{i=1}^{N} \frac{1}{2} \int_{\Omega_U} (I(x) - u_i(x))^2 p_i(x)\, dx \\ &+ \sum_{i=1}^{N} \lambda \int_{\Omega_U} |\nabla u_i(x)|\, dx + \sum_{i=1}^{N} \mu \int_{\Omega_U} |\nabla p_i(x)|\, dx \end{aligned} \tag{5–8}$$

If we solve this problem still using previous procedures (5–6) and (5–7), then it is not a supervised segmentation since we didn't use known information to instruct segmentation for unknown area. The key point of our work is to update each pattern u_i based on the nearest point principle, i.e.,

$$u_i(x) = average\{u_i(y) |\ y = arg\min_y \{|x - y|, y \in \Omega_i\}\} \tag{5–9}$$

wihch is referred from the third step of Poisson Matting [90].

With initially given patterns $u_i(x)$ and under smoothness constraint of $p_i(x)$, the objective energy functional becomes

$$\begin{aligned} E(p) = &\sum_{i=1}^{N} \frac{1}{2} \int_{\Omega_U} (I(x) - u_i(x))^2 p_i(x)\, dx \\ &+ \sum_{i=1}^{N} \mu \int_{\Omega_U} |\nabla p_i(x)|\, dx \end{aligned} \tag{5–10}$$

where each $u_i(x)$ is determined by (5–9).

Considering that the close relation of points when they are near to each other and the loose relation when they are far away, we add a distance factor to the fitting term. The energy functional is therefore becomes

$$\begin{aligned} E(p) = &\sum_{i=1}^{N} \frac{1}{2} \int_{\Omega_U} (I(x) - u_i(x))^2 p_i(x) e^{\alpha d_i(x)} dx \\ &+ \sum_{i=1}^{N} \mu \int_{\Omega_U} |\nabla p_i(x)| dx \end{aligned} \tag{5–11}$$

where α is a parameter and $d_i(x)$ is the nearest distance obtained from 5–9. The factor $e^{\alpha d_i(x)}$ will force the influence to be ignored when the distance of two points is far away.

Correspondingly, we rewrite the energy with indication functions and replace the total variation by weighted total variation, we get the final energy functional

$$\begin{aligned} E(p) = &\sum_{i=1}^{N} \int_{\Omega} (I(x) - u_i(x))^2 p_i(x) e^{\alpha d_i(x)} \chi_{\Omega_U}(x) dx \\ &+ \sum_{i=1}^{N} \mu \int_{\Omega} |\nabla p_i(x)| g(\nabla \tilde{I}(x)) \chi_{\Omega_U}(x) dx \end{aligned} \tag{5–12}$$

where $\tilde{I}$ is a smoothness of I and $g(x_1, x_2)$ is an edge function usually defined as $\frac{1}{1 + x_1^2 + x_2^2}$.

5.2 Solve Semi-Supervised Multiphase Image Segmentation

The primal-dual form of (5–12) is

$$\begin{aligned} E(p) = &\sum_{i=1}^{N} \int_{\Omega} (I(x) - u_i(x))^2 p_i(x) e^{\alpha d_i(x)} \chi_{\Omega_U}(x) dx \\ &+ \sum_{i=1}^{N} \mu \sup_{|q_i| \le 1, q_i \in C_c^1(\Omega, R^2)} \int_{\Omega} p_i(x) div\ (g(\nabla \tilde{I}(x)) \chi_{\Omega_U}(x) q_i(x)) dx \end{aligned} \tag{5–13}$$

So, the iteration on p and q are, respectively,

$$\begin{cases} \dfrac{\partial p_i}{\partial t} = -\left[(u_i - I)^2 e^{\alpha d_i(x)} \chi_{\Omega_U}(x) + \mu\ div\ (g(\nabla \tilde{I}(x)) \chi_{\Omega_U}(x) q_i(x))\right] \\ \dfrac{\partial q_i}{\partial t} = -\mu (\nabla p_i) g(\nabla \tilde{I}(x)) \chi_{\Omega_U}(x) \end{cases} \tag{5–14}$$

In our framework, the new memberships $p_i(x)$ obtained from above iterations are actually a temporary one. We still need to update the memberships based on the following rule: if $p_i(x) > 0.95$ and $I(x) \approx u_i(x)$ for some $1 \leq i \leq N$ and $x \in \Omega_U$, then put x to Ω_i^+. So, the known parts are updated by

$$\Omega_i = \Omega_i \cup \Omega_i^+. \tag{5–15}$$

Corresponding, the unknown part is updated according to the following equation

$$\Omega_U = \Omega - \bigcup_{i=1}^{N} \Omega_i \tag{5–16}$$

We now describe the complete algorithm. Given an image I defined in a domain Ω, if the image contains N classes, then the complete algorithm for our semi-supervised multiphase image segmentation is given as below.

1. **Initialization.**
   (a) Initialize known parts Ω_i^0 using brush;
   (b) Initialize unknown part by $\Omega_U^0 = \Omega - \bigcup_{i=1}^{N} \Omega_i^0$;
   (c) Initialize memberships: For each $1 \leq i \leq N$ and $x \in \Omega_i^0$, set $p_i^0(x) = 1$ and $p_j^0(x) = 0$ for $j \neq i$; for $x \in \Omega_U^0$, set $p_i(x)$ randomly.
   (d) Initialize patterns: For each $1 \leq i \leq N$ and $x \in \Omega_i^0$, set $u_i^0(x) = I(x)$; For any $x \in \Omega_U^0$, set $u_i^0(x)$ in terms of the nearest point principle as (5–9);
2. **Iterations.**
   (a) Update memberships $p_i^k(x)$ by (5–14);
   (b) Update known areas Ω_i^k by (5–15);
   (c) Update unknown area Ω_U^k by (5–16);
   (d) Update patterns $u_i^k(x)$ by (5–9);
3. **Termination** The iterations will be terminated if $\Omega_U = \emptyset$.

(A) (B) (C)

Figure 5-2. Comparison with synthetic image.

5.3 Experiments

In Figure 5-2, we show how semi-supervised algorithm works differently from general unsupervised segmentation. The original image contains two objects with same intensities as shown in Figure 5-2(B). Suppose that only the left one is the object that we need to separate. Under unsupervised segmentation, the segmentation of the foreground will be exactly the same as shown in Figure 5-2(B). However, if we circle a green mask as the seed for the background and circle a red mask as the seed for the foreground as shown in Figure 5-2(A), then the segmentation of the foreground is shown as in Figure 5-2(C), which is the desired one. Figure 5-3 and Figure 5-4 show the segmentation results of a flower and a MRI medical image respectively with a comparison between using unsupervised segmentation and using supervised segmentation. In Figure 5-3, (A1) is the original image, (A2) is the original image with assigned class masks, (B1) and (C1) are the unsupervised segmentation, and (B2) and (C2) are supervised segmentation. The differences are obvious. In Figure 5-4, (A1) is the original image, (A2) is the original image with assigned class masks, (B1)-(D1) are unsupervised segmentation, and (B2)-(D2) are supervised segmentation.

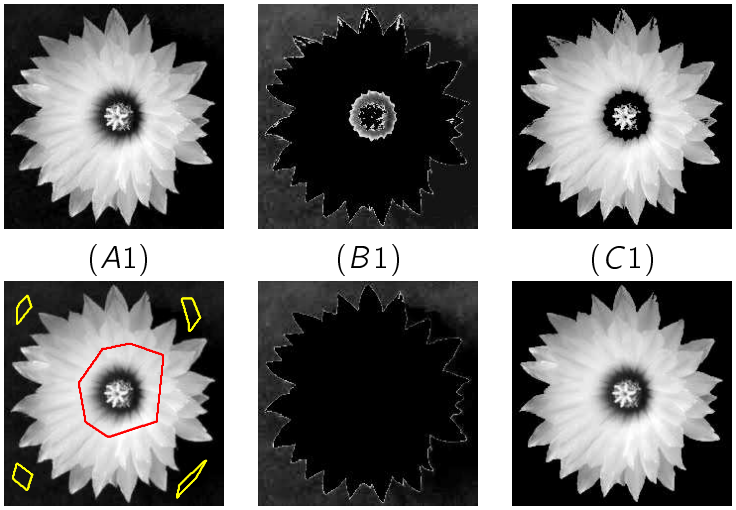


Figure 5-3. Comparison with flower.

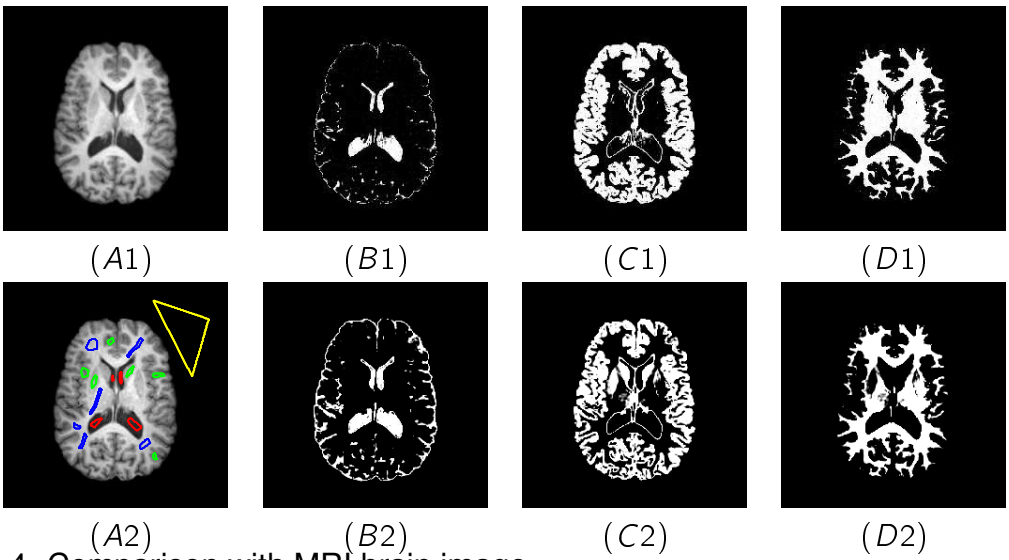


Figure 5-4. Comparison with MRI brain image.

CHAPTER 6
FUTURE WORK

In our first model in Chapter 2, we used the same numerical implementation method as in Sine-Sinc model. The method can guarantee to converge to a local minimum or a saddle point. That means, we can not find a global minimum for that model. The future work is to find a better numerical method so that better result can be expected, or find a way to convert the model to a convex functional, so that a global minimizer can be guaranteed.

In Chapter 3, the constructed function $h(x)$ can still be improved. Actually, the particular function $h(x)$ in our implementation satisfy $\sum_{k=0}^{K} h(z(x) - k) = 1$ for all x in the domain only when the domain is infinite. When the domain is finite, cutting-off must be used to make the equation hold at the ends of the domain. Therefore, how to overcome this problem is still an open problem.

In Chapter 4, we noticed the energy functional is convex with respect to all variables except for the variances. This hinds to find a global minimizer when alternating iterations are carried out. One way is to fix the variances based on prior knowledge of the given image. However, how to choose the variances is still a problem. If variances are not fixed, how to guarantee the numerical method converge to an ideal solution is another work to do.

In Chapter 5, the frame work of semi-supervised image segmentation is still based on intensity. When the image contains some texture features, the frame work does not work very efficiently. In this case, feature-based model must be used in the frame work. Let $F : \Omega \subset R^n \rightarrow Z \subset R^m$ be a function which maps an n-dimensional image domain to a multi-dimensional (m-dimensional) space of contextual features Z. For each point $x \in \Omega$, $F(x)$ is a vector containing image statistics or features. Such features can encode contextual knowledge about the region of interest and its neighboring structures (e.g., size, shape, orientation, relationships to neighboring structures, etc.).

Feature-based image segmentation is extensively used in texture segmentation and some medical image segmentation. Therefore, an immediate future work is to develop a feature-based semi-supervised multiphase image segmentation frame work.

In addition, an important application of soft segmentation is partial volume segmentation. How to combine non-local information into 3-D soft segmentation to improve partial volume segmentation is a big issue. Further work also includes how to convert some existing multiphase segmentation models for hard segmentation to soft segmentation so that a global minimizer can be found.

REFERENCES

[1] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, pp. 193–199, 2002.

[2] L. Ambrosio and V. M. Tortorelli, "Approximation of functionals depending on jumps by elliptic functionals via $\Gamma - convergence$," *Communications on Pure and Applied Mathematics*, vol. 43, pp. 999–1036, 1990.

[3] L. Ambrosio and V. M. Tortorelli, "On the approximation of free discontinuity problems," *Bollettino della Unione Matematica Italiana*, vol. 6, pp. 105–123, 1992.

[4] J.-F. Aujol, G. Gilboa, T. Chan and S. Osher, "Structure-texture image decomposition modeling algorithms and parameter selection," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 111–136,2006.

[5] E. Bae, J. Yuan and X-C Tai, "Global Minimization for Continuous Multiphase Partitioning Problems using a Dual Approach," *UCLA CAM Report*, 09-75, 2009.

[6] E. Bae, J. Yuan, X.-C. Tai and Y. Boykova, "Fast continuous max-flow approach to non-convex multilabeling problems," *UCLA CAM Report*, 10-62, 2010.

[7] M. Belkin, P. Niyogi, and V. Sindhwani, "On manifold regularization," *Proc. Intl. Workshop on Artificial Intelligence and Statistics*, Barbados, Jan. 2005.

[8] A. Bertozzi06, , S. Esedoglu, and A. Gilette, "Inpainting of binary images using the chan hiliard equation," *UCLA CAM Report*, 06-07, 2006.

[9] J.C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithm," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 2, pp. 1–8, 1980.

[10] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/ snake model," *Jour. of Math. Imaging and Vision*, 2007.

[11] B.Bourdin, A.Chambolle, "Design-dependent loads in topology optimization," *ESAIM Control Optim. Calc. Var.*, vol. 9, pp. 19–48(2003).

[12] Y. Boykov, and M.-P.Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," *International Conference on Computer Vision*, vol. I, pp. 105C-112, July, 2001.

[13] Y. Boykov, and V. Kolmogorov, "An experimental comparison of min-cut/max-flowalgorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp.1124C-1137, 2004.

[14] Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of Computer Vision (IJCV)*, vol. 70, no. 2, pp. 109–131, 2006.

[15] X. Bresson and T.F. Chan, " Non-local unsupervised variational image segmentation models," *UCLA CAM Report*, 08-23, 2008.

[16] X. Bresson, et al, "Fast Global Minimization of the Active Contour/Snake Model," *J. Math. Imaging Vis.*, vol. 28, pp. 151–167, 2007.

[17] E.-S. Brown, T.-F. Chan, and Xavier Bresson, "A Convex Approach for Multi-phase Piecewise Constant Mumford-Shah Image Segmentation," *UCLA Cam Report*, 09-66, July, 2009.

[18] E.-S. Brown, T.-F. Chan, and X. Bresson, "Globally Convex Chan-Vese Image Segmentation," *UCLA CAM Report*, 10-44, 2010.

[19] M. Burger, R. Stainko, "Phase-field relaxation of topology optimization with local stress constraints," *SIAM Journal on Control and Optimization*, vol. 45, pp. 1447–1466, 2006.

[20] J. Cahn, J. Hilliard, "Free anergy of a non-uniform system-interfacial free energy," *Journal of Chem. Phys.*, vol. 28, no. 1, pp. 258–267, 1958.

[21] J. Carter, "Dual methods for total variation-Based image restoration," Ph.D. thesis, UCLA, 2001.

[22] V. Caselles, F. Catte, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numer. Math.*, vol. 66, pp. 1–31, 1993.

[23] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 1, pp. 61–79, 1997.

[24] M. Celenk, "A color clustering technique for image segmentation," *Computer Vision, Graphics and Image Processing*, Vol. 52, pp. 145–170, 1990.

[25] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.

[26] J.-F. Aujol and A, "Chambolle, Dual norms and Image Decomposition Models," *International Journal of Computer Vision*, vol. 63, no. 1, pp. 85–104, 2005.

[27] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien, (Eds.), *Semi-Supervised Learning*, MIT Press, 2006.

[28] T. Chan, G. Golub, and P. Mulet, "A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration," *SIAM Journal on Scientific Computing*, vol. 20, no. 6, pp. 1964–1977, 1999.

[29] Tony F. Chan and Luminita A. Vese, "Active Contours Without Edges," *IEEE Tran. on Image Processing*, Vol. 10, NO. 2, , pp. 266–277, Feb., 2001.

[30] T. F. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM J. Appl. Math.*, vol. 66, pp. 1632–1648, 2006.

[31] S. Chen and D. Zhang, "Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure," *IEEE Transactions on Systems Man and Cybernetics*, vol. 34, no. 4, pp. 1907–1916, 2004.

[32] S.Thiruvenkadam, S.Arcot and Y.Chen, "A PDE Based Method For Fuzzy Classification Of Medical Images," *Proceedings of International Conference on Image Processing*, vol. 481, pp. 1805–1808, 2006.

[33] J.h. An and Y.Chen, "Region based image segmentation using modified Mumford-Shah algorithm," *First International Conference on Scale Space and Variation Variational Methods in Computer Vision*, vol. 4485, pp. 733–742, 2007.

[34] F. Chen, Y.Chen and H. D. Tagare, "An improvement of sine-sinc model based on logrithm of likelihood," *Proc. of IPCV 08*, vol. 1, no. 1, pp. 222–227, 2008.

[35] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski, "A Bayesian approach to digital matting," *CVPR*, 2001.

[36] G. Chung, L. Vese, "Energy minimization based segmentation and denoising using a multilayer level set approach," *Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 3757, no. 1, pp. 439–455, 2005.

[37] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, "Combinatorial Optimization," *John Wiley Sons*, 1998.

[38] A. Dempster, N. Laird, D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal Royal Statis. Society B*, vol. 39, pp. 1–8, 1977.

[39] J.C. Dunn, " A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Jounal. of Cybernetics*, vol. 3, pp. 32–57, 1973.

[40] S. Esedoglu and Y.-H. Tsai, "Threshold dynamics for the piecewise constant Mumford-Shah functional," *J. Comput. Phys.*, vol. 211, pp. 367-384, 2006.

[41] E. Esser, X. Zhang and T. Chan, " A General Framework for a Class of First Order Primal-Dual Algorithms for TV Minimization," *UCLA Cam report 09-67*, August, 2009.

[42] D. Eremina, et.al, "Investigation on an EM framework for partial volume segmentation," *Proceedings of SPIE-Medical Imaging* Vol. 6144, pp. 1–9, 2006.

[43] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[44] L. Ford and D. Fulkerson, "Flows in Networks," *Princeton University Press*, 1962.

[45] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *UCLA cam report 06-47*, 2006

[46] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *Journal of the Association for Computing Machinery*, vol. 35, no. 4, pp. 921-C940, 1988.

[47] T. Goldstein, X. Bresson, S. Osher, "Global minimization of Markov Random Fields with Applications to Optical Flow," *UCLA CAM Report 09-77*, 2009.

[48] D. M. Greig, B. T. Porteous, and A. H. Seheult, " Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society, Series B*, vol. 51, no. 2, pp. 271–279, 1989.

[49] X. Han, C. Xu and J. Prince, "A topology preserving level set method for geometric deformable models," "IEEE Trans. Patt. Anal. Mach. Intell.," vol. 25, pp. 755–768, 2003.

[50] N. Houhou, X. Bresson, A. Szlam, T. F. Chan, and J.-P. Thiram, "Semi-supervised segmentation based on nonlocal continuous min-cut," *Proc. of the Intl. Conf. on Scale Space and Variational Methods in Computer Vision*, LNCS 5567, Voss, Norway, June, 2009.

[51] M. Kass, A.Witkin and D. Terzopoulos, "Snakes: active contour models," *International J. Comp. Vis.*, vol. 1, pp. 321–331, 1987.

[52] Y. Gu, L. Wang and X.-C. Tai, "A Direct Approach towards Global Minimization for Multiphase Labeling and Segmentation Problems," *IEEE Tran. on Image Processing*, Vol. 99, pp. 1–20, 2012.

[53] C. Li, R. Huang, Z. Ding, C. Gatenby and D. Metaxas, "A variational level set approach to segmentation and bias correction of images with intensity inhomogeneity," *MICCAI 2008, Part II*, LNCS 5242, pp. 1083–1091, 2008.

[54] Y. Huang, M. Ng and Y. Wen, " A Fast Total Variation Minimization Method for Image Restoration," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 7, pp. 774–795, 2008.

[55] H. Ishikawa and D. Geiger, "Segmentation by grouping junctions," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 125C-131, 1998.

[56] H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1333–1336, 2003.

[57] Y. Jung, S. Kang and J. Shen, " Multiphase image segmentation via Modica-Mortola phase transition," *SIAM Applied Math.*, vol. 67, no. 1, pp. 1213–1232, 2007.

[58] B. Kimia, A. Tannenbaum and S. Zucker, "Shapes, shocks, and deformations I: the components of twodimensional shape and the reaction-diffusion space," *Intl J. Comp. Vis.*, vol. 15, pp. 189–224, 1995.

[59] S. Lee, "Efficient multistage approach for unsupervised image classification," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Anchorage, AK, USA, pp. 1581–1584, 2004.

[60] K. V. Leeput, F. Maes, D. Vandermeulen and P. Suetens, "A unifying framework for partial volume segmentation of brain MR images," "*IEEE Trans. on Med. Imaging*, Vol. 22, pp. 105–119, 2003.

[61] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral Matting," *IEEE Trans. on PAMI*, Vol. 30, no. 10, pp. 1699–1712, 2008.

[62] C. Li, C. Kao, J. C. Gore and Z. Ding, "Minimization of Region-Scalable Fitting Energy," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 17, no. 10, pp. 1940–1949, Oct., 2008.

[63] Fang Li, Michael K. NG, and Chunming Li, Variational Fuzzy Mumford-Shah Model for Image Segmentation.

[64] X.Li, L.Li, H.Lu and Z.Liang, "Partial Volume Segmentation of Brain Magnetic Resonance Images Based on Maximum a Posteriori Probability," *Med.Physis*, vol. 32, no. 7, pp. 2337–2345, 2005.

[65] M.L.J. Lie and X.C. Tai. "A variant of the level set method and applications to image segmenattion," *AMS Mathematics of Computations*, vol. 75, no. 1, pp. 1155–1174, 2006.

[66] R. Malladi, J. A. Sethian and B. C. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 17, pp. 158–175, 1995.

[67] U. Maulik and S. Bandyopadhyay, "Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1075C-1081, 2003.

[68] U. Maulik and I. Saha, "Modified differential evolution based fuzzy clustering for pixel classification in remote sensing imagery," *Pattern Recogn.*, vol. 42, no. 9, pp. 2135C-2149, 2009.

[69] L. Modica and S. Mortola, "Un esempio di Γ-convergenza," *Boll. Un. Mat. Ital.*, vol. 14, pp. 285–299, 2007.

[70] L. Modica, "The gradient theory of phase transitions and the minimal interface criterion. Arch," *Rational Mech. Anal.*, vol. 98, no. 1, pp. 123–142, 1987.

[71] B. Mory and R. Ardon, "Fuzzy region competition: A convex two-phase segmentation framework," *International Conference on Scale-Space and Variational Methods in Computer Vision, Proceedings*, pp. 214–226, 2007.

[72] B.Mory, R.Ardon and J.P. Thiran, "Variational Segmentation using Fuzzy Region Competition and Local Non-Parametric Probability Density Functions," *ICCV*, pp. 1–8, 2007.

[73] L. Modica and S. Mortola, "Un esempio di Gamma-convergenza," *Bollettino della Unione Matematica Italiana B*, vol. 5, no. 1, pp. 285–299, 1977.

[74] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Appl.Math.*, vol. 42, pp. 577–685, 1989.

[75] W. J. Niessen, K. L. Vincken, J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever, "Multiscale segmentation of three-dimensional MR brain images," *Int. J. Comput. Vis.*, vol. 31, pp. 185–202, 1999.

[76] S. Osher and J.A. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton Jacobi formulations," *J. Comput. Phys.*, vol. 79, pp. 12-49, 1988.

[77] S.J. Osher, "A level set formulation for the solution of the Dirichlet problem for Hamilton-Jacobi equations," *SIAM J. Anal.*, vol. 24, pp. 1145–1152, 1993.

[78] S. Osher and R. Fedkiw, "Level Set Methods and Dynamic Implicit Surfaces," *Springer-Verlag*, New York, 2002.

[79] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *Int. Journal of Computer Vision*, vol. 3, no. 46, 2002.

[80] D. Peng, B. Merriman, S. Osher, H. Zhao, and M. Kang, "A PDE-based fast local level set method," *J. Comp. Phys.*, vol. 155, pp. 410–438, 1999.

[81] D. L. Pham and J. L. Prince, "An adaptive fuzzy C-means algorithm for the image segmentation in the presence of intensity inhomogeneities," *Pattern Recognit. Lett.*, vol. 20, pp. 57–68, 1998.

[82] T. Pock, A. Chambolle, H. Bischof, and D. Cremers, "An algorithm for minimizing the mumford-shah functional," *IEEE Conference on Computer Vision (ICCV)*, 2009.

[83] L. Rudin, S.J. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259–268, 1992.

[84] M. Rumpf and B. Wirth, "A nonlinear elastic shape averaging approach," *SIAM Journal on Imaging Sciences*, 2008.

[85] Linda G. Shapiro and George C. Stockman, "Computer Vision," New Jersey, Prentice-Hall, ISBN 0-13-030796-3, pp. 279–325, 2001.

[86] J. A. Sethian, "Level set methods and fast marching methods," Cambridge: Cambridge University Press, 1999.

[87] Shen.J. " A Stochastic Variational Model for Soft Mumford-Shah Segmentation," *International Journal of Biomedical Imaging*, vol 2006, pp. 1–14, 2006.

[88] Y. Jung, , S. Kang, and J. Shen, "A stochastic-variational model for soft mumford-shah segmentation," *SIAM Applied Math.*, vol. 67, no. 1, pp. 1213–1232, 2007.

[89] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[90] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum, "Poisson Matting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 315–321, 2004.

[91] M. Sussman and E. Fatemi, "An efficient, interfacepreserving level set redistancing algorithm and its application to interfacial incompressible fluid flow," *SIAM J. Sci. Comp.*, vol. 20, pp. 1165–1191, 1999.

[92] X.-C. Tai and C.-H. Yao, "Image segmentation by piecewise constant mumford-shah model without estimating the constants," *Journal of Computational Math.*, vol. 24, pp. 435–443, 2006.

[93] H. D. Targare, Y. Chen and R. K. Fulbright, "A reparameterized level set algorithm and its comparison with EM-based partial volume segmentation of MR brain images," *SPIE*, 2008.

[94] D. Tuia, M. Kanevski, J. Munoz Mari, and G. Camps-Valls, "Cluster-based active learning for compact image classification," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Honolulu, Hawaii, USA, pp. 2824–2827, 2010.

[95] J.-B. Hiriart-Urruty and C. Lemarechal, "Convex analysis and minimization algorithms," *Grundlehren der mathematischen Wissenschaften*, pp. 305–306, Springer-Verlag, New York, 1993.

[96] O. Veksler, "Image segmentation by nested cuts," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 339–344, 2000.

[97] B. Vemuri and Y. Chen, "Joint image registration and segmentation," *Geometric Level Set Methods in Imaging, Vision, and Graphics*, Springer, pp. 251–269, 2003.

[98] L. Vese, T. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *Int J.of Comp. Vision* vol. 50, no. 3, pp. 271–293, 2002.

[99] C. Vogel, " Computational methods for inverse problems," *SIAM*, Philadelphia, 2002.

[100] Fei Wang, Jingdong Wang, Changshui Zhang, and Helen C. Shen, "Semi-supervised classification using linear neighborhood propagation," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, New York, NY, USA, June, 2006.

[101] M.Y.Wang, S.Zhou, "A Variational Method for Structural Topology Optimization," *CMES*, Vol. 6, no. 6, pp. 547–566.

[102] W. Wells, E. Grimson, R. Kikinis, and F. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.* vol. 15, pp. 429–442, 1996.

[103] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1101C1113, 1993.

[104] C. Xu and J. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Imag. Proc.*, vol. 7, pp. 359–369, 1998.

[105] A.L. Yuille and A. Rangarajan, "The concave-convex procedure (CCCP)," *Neural Computations*, vol. 15, pp. 915–936, 2003.

[106] B.M.H.-K. Zhao, T. Chan, S. Osher, "A variational level set approach to multiphase motion," *J. Comput. Phys.*, vol. 127, no. 1, pp. 179–195, 1996.

[107] Dengyong Zhou, Olivier Bousquet, Thomas N. Lal, Jason Weston, and Bernhard Scholkopf, "Learning with local and global consistency," *Advances Neural Information Processing Systems*, Vancouver, BC, Canada, Dec. 2003.

[108] M. Zhu and T. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm For Total Variation Image Restoration. CAM report, cam08-34, 2008.

[109] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, September, 1996.

[110] Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty, "Semi-supervised learning with Gaussian fields and harmonic functions," *Proc. Intl. Conf. on Machine Learning*, Washington, D.C, USA, Aug. 2003.

[111] Xiaojin Zhu, "Semi-supervised learning literature survey," *Computer Sciences TR 1530*, University of WinconsinCMadison, 2008.

BIOGRAPHICAL SKETCH

Fuhua Chen was born in Nanjing, China in 1966. He went to the Department of Mathematics in 1984, and got his bachelor's degree in 1988 and master's degree in 1991 both from Nangjing University. He then worked at Hohai University till 2001 as an assistant professor and a lecturer. From September 1999 to July 2003, he was a part-time graduate student at the Nanjing University of Science & Technology reading for his doctorate degree in computer science with research interest in pattern-recognition and artificial-intelligence. In July 2001, he immigrated to Canada and became a Canadian citizen in 2006. From September 2004, he was a graduate student at the University of Windsor and got his master's degree in mathematics and statistics in 2006. From January 2006 to May 2006, he was a postdoctoral fellow at Yangming University where his research was focused on positron-emission-tomography. From August 2006, he went to the University of Florida reading for his doctorate degree in mathematics with research interest in image-processing.