Utilization of spatial information for segmentation of cell nuclei in fluorescence microscopy image

Amin Gharipour
School of Information and Communication Technology
Griffith Sciences
Griffith University

Submitted in fulfillment of the requirements of the degree of
Doctor of Philosophy

September 2015
Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments.

Amin Gharipour
September 2015
To be published manuscripts from the thesis


Published manuscripts from the thesis


Other publications during candidature


Abstract

Numerous aspects of analyzing and quantifying fluorescence microscopy images rely on quantitative cell nucleus image analysis. Specifically, the basis for all automated cell image analysis in high-throughput applications is image segmentation. Semi-automatic and manual segmentation methods are tedious, require intensive labor, and suffer from inter- and intra-person variability. Therefore, automatic methods with the ability to deal with different cell types and image artifacts are required.

The goal of this thesis is motivated by the fact that in cell nuclei image segmentation, the spatial relationship of the pixels can be utilized as an important characteristic that improves the performance of segmentation methods. Therefore, this thesis aims to use spatial information in cell nuclei image segmentation and proposes several implicit models under different assumptions. First, we assume that the local image data are Gaussian and propose an implicit model based on the Bayesian classification risk and anisotropic weighting scheme for fluorescence microscopy image segmentation. The proposed algorithm obtains the lowest MAD measure for all four experiments. The MAD values are approximately 18%, 7%, and 12% better than the state-of-the-art methods we compared with for U20S cells, NIH3T3 cells, and Synthetic cells, respectively. Under the non-Gaussianity assumption, we utilize the correntropy criterion and propose the local implicit model. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 7%, and 5% better than the state-of-the-art methods we compared with for U20S cells and BBBC005 cells, respectively. Further, we assume that the local image data are nonlinearly separable and propose the local model based on multiple kernels mapping where a linear combination of multiple kernels is utilized to implicitly
map the original local image data into data of a higher dimension. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 5%, and 6% better than the state-of-the-art methods we compared with for U20S data set and NIH3T3 data set, respectively. Finally, a novel implicit segmentation method based on the multiple kernels is proposed. In this method, a new flexible framework to fuse different pixel information is introduced where different pixel information represented by different kernels is combined in the kernel space to produce a new kernel. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 2%, 5%, 2% and 4% better than the state-of-the-art methods we compared with for U20S cells, NIH3T3 cells, BBBC005 data set, and Synthetic cells, respectively.
Acknowledgments

First and foremost, I would like to express my deepest appreciation to my principal supervisor Prof. Alan Wee-Chung Liew for his continuous support during my Ph.D. His patience, motivation, immense knowledge, and guidance helped me in all the time of research. My parents taught me the first lessons in my life and motivated me to explore and find new horizons. I like to appreciate their unconditional love and all of their efforts. I would also like to thank my brother, Prof. Mohammad Gharipour, and my sister, Dr. Mojgan Gharipour. Thank you for supporting me for everything, and especially I can't thank you enough for encouraging me throughout this experience.
Acronyms

FISH: Fluorescence In-Situ Hybridization
IHC: Immunohistochemistry
GFP: Green Fluorescent Protein
CLSM: Confocal Laser Scanning Microscope
NA: Numerical Aperture
CCD: Charged Coupled Device
Dice FP: Dice False Positive
Dice FN: Dice False Negative
MAD: Mean Absolute Distance
ACWE: Active Contour without Edge
BLS: Bayesian-based Level Set Approach
PS: Piecewise Smooth
LBF: Local Binary Fitting
LGDFE: Local Gaussian Distribution Fitting Energy
LCV: Local Chan–Vese
SPF: Region-based Signed Pressure Force
AACWE: Adaptive Active Contours Without Edges
LRAC: Localizing region-based active contours
SBGFRLS: Selective Binary and Gaussian Filtering Regularized Level Set
LLC: Locally Linear Classification
LGDF: Local Gaussian Distribution Fitting
LLBWIP: Local Level set Model based on the Bayesian Risk and Weighted Image Patch
MA: Merging Algorithm
OT: Otsu Thresholding
WA: Watershed Algorithm
RSFE: Region-Scalable Fitting Energy
DRLSE: Distance Regularized Level Set Energy
LSBR: Level Set based on the Bayesian Risk
MSE: Mean Square Error
MACE: Correntropy-based Minimum Average Correlation Energy
LLCK: Local Level Set model based on the Correntropy-based K-means Clustering
MKLLS: Multiple Kernel Local Level Set
KLS: Kernel-based Level Set approach
LSMKD: Level Set Segmentation with Multiple Kernels Induced Data Term
Contents

Abstract ............................................................................................................................... iv
Acknowledgments ............................................................................................................... vi
Acronyms .......................................................................................................................... vii
Contents .............................................................................................................................. ix
List of Figures ................................................................................................................... xii
List of Tables .................................................................................................................... xiv

Chapter 1 .......................................................................................................................... 1
  1.1 Why segmentation? ......................................................................................................... 1
  1.2 Aim of the research ......................................................................................................... 2
  1.3 Thesis outline, and original contributions .................................................................... 2

Chapter 2 .......................................................................................................................... 5
  2.1 Introduction ....................................................................................................................... 6
  2.2 Specimen preparation ....................................................................................................... 7
  2.3 Image acquisition ............................................................................................................. 9
  2.4 Fluorescence microscopy image segmentation ............................................................. 12
    2.4.1 Image segmentation .................................................................................................. 12
    2.4.2 Datasets .................................................................................................................. 14
    2.4.3 Segmentation performance evaluation .................................................................... 15
  2.5 Summary ....................................................................................................................... 16

Chapter 3 .......................................................................................................................... 17
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3.1 Introduction</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>3.2 Parametric deformable models</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3.3 Implicit deformable models</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>3.4 Discussion</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>3.5 Summary</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Chapter 4</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>4.1 Introduction</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>4.2 LLBWIP</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>4.2.1 Modelling</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>4.2.2 Level set formulation</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>4.2.3 Minimization</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>4.3 Experimental results</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>4.4 Summary</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Chapter 5</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>5.1 Introduction</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>5.2 LLCK</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>5.2 Experimental results</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>5.3 Summary</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Chapter 6</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>6.1 Introduction</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>6.2 MKLLS</td>
<td>67</td>
</tr>
</tbody>
</table>
6.3 Experimental results ................................................................. 72
6.4 Summary .................................................................................. 75

Chapter 7 ..................................................................................... 76
7.1 Introduction ............................................................................. 77
7.2 LSMKD ..................................................................................... 78
7.3 Experimental results ............................................................... 83
7.4 Summary ................................................................................ 87

Chapter 8 ..................................................................................... 88
8.1 Research accomplishments ..................................................... 88
8.2 Recommendations for future work ........................................... 90

Bibliography .................................................................................. 92
List of Figures

Figure 2.1: Absorption and emission spectrum of 4,6-diamidino-2-phenylindole (DAPI) fluorophore. ................................................................. 6

Figure 2.2: Fluorescence in-situ hybridization (FISH) principle. Image source: National Human Genome Research Institute, http://www.genome.gov. .......................... 8

Figure 2.3: A schematic of (a) widefield and (b) confocal fluorescence microscope systems................................................................. 10

Figure 2.4: Original images of the four different datasets. (a) U20S cells, (b) NIH3T3 cells, (c) BBBC005, (d) Synthetic microscopy images from SIMCEP. ......................... 14

Figure 3.1: Implicit shape representation using the distance transform [48]. ............... 22

Figure 4.1. Two image patches selected from a synthetic image, (middle column) the calculated mean-square deviation and (right column) the estimated weight vectors. .... 43

Figure 4.2. Segmentation results of the proposed approach. (a) original images, (b) segmentation results................................................................. 46

Figure 4.3(a) Comparisons of Heaviside function with different $\varepsilon$ values (b) Comparisons of Dirac delta function with different $\varepsilon$ values. ................................. 50

Figure 5.1: Graphical representation of $\Omega_1 \cap N \mathbf{x}$. The red circle denotes the circular neighbourhood of $\mathbf{x}$, $N \mathbf{x}$. $\Omega_1$ and $\Omega_2$ partition $N \mathbf{x}$ into two subregions $\Omega_1 \cap N \mathbf{x}$, and $\Omega_2 \cap N \mathbf{x}$. (The reader is referred to the online version of this chapter for the colour image.) ......................................................... 60

Figure 6.1. Illustration of nonlinear 2-D data separation with mapping: The data is non linearly separable in the data space. Mapping the data to a feature (kernel) space and, then, separating it in the induced space with linear methods is possible. For the purpose
of display, the feature space in this example is of the same dimension as the original data space. In general, however, the feature space is of higher dimension. ........................ 68
List of Tables

Table 3.1. Summarization of most relevant proposed implicit techniques. .................. 34

Table 4.1. Quantitative results for U20S data set for the different segmentation approaches. .................................................................................................................... 50

Table 4.2. Quantitative results for NIH3T3 cells data set for the different segmentation approaches. .................................................................................................................... 51

Table 4.3. Quantitative results for BBBC005 data set for the different segmentation approaches. .................................................................................................................... 52

Table 4.4. Quantitative results for synthetic data set for the different segmentation approaches. .................................................................................................................... 53

Table 4.5. Ranking of segmentation methods according to the Jaccard, the Hausdorff distance, and the MAD metrics. .................................................................................................................... 54

Table 4.6. Standard deviations of LLBWIP results. .................................................................................................................... 55

Table 5.1. Quantitative results for U20S data set for the different segmentation approaches. .................................................................................................................... 63

Table 5.2. Quantitative results for BBBC005 data set for the different segmentation approaches. .................................................................................................................... 64

Table 5.3. Standard deviations of LLCK results. .................................................................................................................... 64

Table 6.1. Quantitative results for U20S data set for the different segmentation approaches. .................................................................................................................... 73

Table 6.2. Quantitative results for NIH3T3 data set for the different segmentation approaches. .................................................................................................................... 74

Table 6.3. Standard deviations of MKLLS results. .................................................................................................................... 74
Table 7.1. Quantitative results for U2OS data set for the different segmentation approaches. ................................................................. 84
Table 7.2. Quantitative results for NIH3T3 cells data set for the different segmentation approaches. ................................................................. 85
Table 7.3. Quantitative results for BBBC005 data set for the different segmentation approaches. ................................................................. 86
Table 7.4. Quantitative results for synthetic data set for the different segmentation approaches. ................................................................. 86
Table 7.5. Standard deviations of LSMKD results. ................................................................. 87
Chapter 1

Introduction

1.1 Why segmentation?

Numerous aspects of analyzing and quantifying fluorescence microscopy images rely on quantitative cell nucleus image analysis [1-3]. Specifically, the basis for all automated cell image analysis in high-throughput applications is cell image segmentation. Semi-automatic and manual segmentation methods are tedious, require intensive labor, and suffer from inter-and intra-person variability. Therefore, automatic methods with the ability to deal with different cell types and image artifacts are required.
1.2 Aim of the research

A popular trend in image segmentation is to use spatial information, which addresses the issues of intensity inhomogeneity and sensor noise, within the segmentation frameworks. The main reason behind the idea of using spatial information for image segmentation is that the pixels in an image are highly correlated, i.e. the pixels in the immediate neighborhood have nearly the same feature data. Therefore, the spatial relationship between neighboring pixels is an important characteristic that can be of great aid in image segmentation. However, the performance of a segmentation algorithm is fundamentally bounded by the assumptions it made. As the first step in this research, we studied and identified limitations already reported in the literature.

Since most of the proposed approaches assume that the global image data or local image data can be described by Gaussian distribution or are linearly separable [4-11], the second and most important purpose of the study is to develop new frameworks for fluorescence microscopy image segmentation using spatial information and non-Gaussian techniques such as the correntropy criterion and kernel methods.

1.3 Thesis outline, and original contributions

The work performed in this research is divided into seven main chapters. The following briefly describes the contents of each chapter.

Chapter 2 provides the fundamental concepts of fluorescence microscopy imaging and introduces fluorescence microscopy image segmentation as one of the main image processing tasks common in this field. Also, the datasets which are used in this thesis are considered and the region-based and contour-based measures, which can be used for segmentation performance evaluation, are introduced.
Chapter 3 provides a detailed review of the literature about deformable models, which can be considered as a major category of cell segmentation techniques. This chapter provides a required theoretical background of deformable approaches for image segmentation and an overview of existing methods.

Chapter 4 discusses the first contribution of this thesis towards fluorescence microscopy image segmentation. An implicit model in a variational level set formulation, which is based on the image patch information is used to segment the image. Compared to previous approaches, we define a novel local energy functional based on the Bayesian classification risk for an image patch. Also, a weighting scheme is used to enable the pixels in each image patch to have anisotropic weights. A manuscript that highlights the outcomes of the chapter is under review in *Pattern recognition*.

Chapter 5 provides a novel implicit approach for cell nucleus segmentation in fluorescence microscopy images. In this approach, a level set model based on local spatial information and correntropy-based k-means clustering (LLCK) is proposed to segment the fluorescence microscopy images. Compared to previous approaches, the main advantages of our segmentation method can be highlighted as follows. First, due to the correntropy criterion, the LLCK segmentation algorithm is robust to outliers. Second, since the spatial relationship of the pixels in a local neighborhood can be utilized as an important characteristic that improves the performance of level set segmentation methods, the LLCK segmentation algorithm can efficiently segment the images with intensity inhomogeneity by employing the local image information. This method is accepted in the DICTA 2015 conference.

The third contribution of the thesis is illustrated in Chapter 6. This chapter introduces a novel local model based on the linear composite of multiple kernels for cell nucleus
segmentation. In this approach, a data term around a neighborhood of a point is defined based on the local image intensities data which are mapped implicitly into data of a higher dimension using the linear combination of multiple kernels. Finally, the data term is integrated over the entire image domain to form a double integral energy. This method is accepted in the Dicta 2015 conference.

Chapter 7 discusses the fourth contribution of this thesis. In this chapter, a novel segmentation method with multiple kernels induced data term is proposed as a framework for image segmentation problems. The proposed approach provides a new flexible vehicle to fuse different pixel information in image segmentation problems. That is, different pixel information represented by different kernels is combined in the kernel space to produce a new kernel. A manuscript from the outcome of this chapter is under preparation for submission to a high impact journal.

Chapter 8 summarizes the main conclusions obtained in this research. Also, suggestions for further study are provided.
In this chapter, the fundamental concepts of fluorescence microscopy imaging are briefly reviewed, and fluorescence microscopy image segmentation as one of the critical processing tasks common in this field is introduced. The specimen preparation techniques and staining protocols are described, followed by a description of the acquisition principles of fluorescence microscopes and basic components of these devices. Then, the characteristic properties of the acquired images and the segmentation challenges such as noise and intensity inhomogeneity are presented. Finally, the datasets which are used in this thesis are considered and the region-based and contour-based measures, which can be used for segmentation performance evaluation, are introduced.
2.1 Introduction

Fluorescence microscopy imaging technique plays a significant role in various fields including biological science, as a result of it being able to reveal features that are not observable with traditional optical microscopy. In this thesis, the utilization of fluorescence microscopy imaging in cell biology is briefly considered. Since the cell structures and components of interest are not observable directly, fluorescent markers are introduced into the cell to visualize them. Using the fluorescent markers, cells, and sub-microscopic cellular components can be recognized with a high degree of specificity amid non-fluorescing material.

![Absorption and emission spectrum of 4,6-diamidino-2-phenylindole (DAPI) fluorophore.](image)

Figure 2.1: Absorption and emission spectrum of 4,6-diamidino-2-phenylindole (DAPI) fluorophore.

The fundamental principle utilized by fluorescence microscopes is a physical phenomenon known as fluorescence. In fluorescence, the emission of light of a particular wavelength is activated by the molecular absorption of a photon in the fluorescent material. It was first proposed in 1852 by a British scientist Sir George G. Stokes in his chapter on "Refrangibility." He observed that red light was emitted by the mineral fluorite when it was illuminated by light in the ultraviolet spectrum [12]. It has
also been noted that the emitted light always has a longer wavelength due to the fact that a small amount of the excitation light energy is converted to heat. An example of the excitation and emission spectrum of the so-called blue fluorophore (DAPI) is depicted in Figure 2.1.

In the rest of this chapter, Section 2 describes the basic approaches for introducing fluorescent markers into a biological specimen. In section 3, the fundamentals of data acquisition of the fluorescently stained material and the properties of the acquired data are given. Section 4 considers fluorescence microscopy image segmentation, as one of the most common image processing tasks in this field. Section 4 also describes the datasets and the segmentation performance measures used in this research.

2.2 Specimen preparation

For cell nucleus visualization, cell staining techniques are utilized. In this process, fluorescent dyes attached to molecular probes that are capable of specific binding to the desired objects have to be introduced into the specimen. After specimen staining, a fluorescence microscope is utilized to observe the target objects. Fluorescence in-situ hybridization (FISH) and immunohistochemistry (IHC) [13] are two commonly used staining protocols.

The FISH staining technique is commonly applied for visualization of cell nuclei due to its capability of visualizing selected regions in the nucleic acids – DNA and RNA. An illustration of the principle of FISH is shown in Figure 2.2. First, a sequence complementary to the target DNA or RNA sequence is synthesized (so-called probe), and fluorescent dye is utilized to directly or indirectly (using the biotin) label them. The next step is hybridization where both the target and the complementary sequence are denatured. Then, the labeled probe is let to anneal to the target sequences by dropping
them on the slide with the specimen where they can be combined. The probe that did not anneal to any target sequence is then washed away. In IHC protocol, the localization of particular proteins is enabled by immunohistochemistry. This process is based on the antibodies, which are Y-shaped proteins that recognize and bind to a specific molecule on the target protein called antigen, produced directly by the immune system. To observe the target structures, a fluorescent dye is utilized for direct or indirect labeling processes. It is notable that both FISH and IHC can be used at the same time. Also, it is possible to combine several fluorescent dyes provides that their emission and excitation spectra do not overlap too much.

Figure 2.2: Fluorescence in-situ hybridization (FISH) principle. Image source: National Human Genome Research Institute, http://www.genome.gov.

The main drawback of the above staining techniques is that they cannot be applied to experiments with living cells because they require fixation of cells on a microscopic slide and chemical or thermal treatment. To overcome this problem, a different approach is taken in which the fluorescent material is produced by the target structures of the modified living cells. Since the green fluorescent protein (GFP), which was first
isolated from the jellyfish Aequorea Victoria [14], produce a highly observable and non-toxic substance, they are utilized for this approach.

In order to obtain a good image quality, the following issues should be considered carefully. First, the staining technique has to be selected carefully. To avoid intensity inhomogeneity, places within the cell nucleus have to be reachable for the fluorescent dyes. Incomplete washing is the cause of some fluorescent dyes remaining in wrong places and producing wrong signals. Also, the observed image can be corrupted by foreign particles like dust or other impurities.

2.3 Image acquisition

A fluorescence microscope is utilized to visualize the fluorescently marked specimen. The schematic layout of a fluorescence microscope is showed in Figure 2.3. Fluorescence microscopes use the concept of conventional optical microscopes while their acquisition capabilities are extended using additional components.

In the fluorescence acquisition mode, the light is radiated from a high-luminance light source (typically a mercury or xenon arc lamp or laser) over the excitation filter. In the excitation filter, the wavelengths, which correspond to the absorption spectrum of the fluorophore, are permitted to go through. The radiated light reflects from the dichroic mirror and reaches the specimen through the objective. Electrons can be in different energy states. The ground state is a very stable state for an electron and has the lowest energy level [15, 16]. The electrons of the fluorochrome are elevated to a higher energy level, an excited state. When an electron returns to the ground state, this energy is released in the form of emitted light. The emitted light passes through the dichroic mirror and the emission filter. It guarantees that only light emitted by the fluorophore reaches the eye or a detector usually consisting of a CCD (charged-coupled device)
camera or a photomultiplier tube [17]. The difference between the absorption and emission spectra of the observed fluorophore needs to be large enough to distinguish between the excitation and emission light, such as in Figure 2.1. This aspect is important for the functionality of a fluorescence microscope. When the specimen is stained with several fluorescent dyes, the selection of what is to be observed is based on the combination of excitation and emission filters. Also, when some biological material naturally fluoresces without intervention, the autofluorescence effect can be suppressed by the combination of excitation and emission filters.

Figure 2.3: A schematic of (a) widefield and (b) confocal fluorescence microscope systems.

The image quality can be affected by the different choice of hardware instruments and their setup. Two-dimensional image data can be produced by the widefield (conventional) fluorescence microscopy. The axial direction (z-axis) is used to aggregate the intensity of a pixel at a given lateral coordinate. The final image can be corrupted by the reflected and out-of-focus light. In the confocal mode, most of the light from the specimen that is not from the microscope’s focal plane are excluded, and only
a thin cross-section of the specimen is represented. The obtained image from the confocal mode has less haze and better contrast than that of a conventional microscope. One of the two most common confocal systems is the confocal laser scanning microscope (CLSM). In this system, the scanning process is performed point by point, line-by-line and one plane after another using a focused laser beam and the image is progressively formed in the computer memory. While the main advantage of this approach is that out-of-focus light is effectively reduced, the high-intensity laser source is utilized in this system and therefore, is much more expensive. The confocal spinning disk microscope [18] can provide a cheaper alternative to CLSM. In this system, a rotating disk with many pinholes is utilized to achieve the confocal effect while the light is projected through hundreds of pinholes. It increases the acquisition speed with the whole observable area attained throughout a single revelation of the disc. Using specialized acquisition techniques such as the two-photon [19], 4Pi[19] or STED [20], can improve the results.

Microscope objectives are the most important imaging component, and also the most complex for the reason that they are responsible for primary image formation and play a key role in determining the quality of images that the microscope can produce. The magnification factor, the numerical aperture (NA) and the level of correction of monochromatic and chromatic optical aberrations are main features of the microscope objectives. The numerical aperture of a microscope objective can be defined as a measure of the light efficiency of the objective and determines the fine specimen detail at a fixed object distance. The optical resolution of the microscope system can be affected by the choice of the objective. When the higher resolution is required, oil-immersion or water-immersion objectives can be utilized since they have a higher numerical aperture.
2.4 Fluorescence microscopy image segmentation

The first part of this section provides a brief introduction on fluorescence microscopy image segmentation. The second part presents the datasets which were used in this thesis. The third part introduces region-based and contour-based measures which can be used for segmentation performance evaluation.

2.4.1 Image segmentation

Fluorescence microscopy has become an essential tool in biology during the last two decades. In fluorescent microscopy, motorized microscope stage is utilized to control the movement in both lateral and axial directions. Also, most other components including filter selection can be remotely controlled by a computer. By making fluorescent microscope a fully computerized instrument capable of unsupervised acquisition over long periods of time generating large 2D dataset, it can be used in a wide variety of biological experiments.

The acquired image has a size that differs by tens to several hundreds of megabytes. Humans cannot perform manual analysis of such enormous amounts of data in a reasonable time. Therefore, robust and automatic or, at least, semi-automatic image processing methods play an important role in the quantitative analysis of fluorescent microscopy images. Image segmentation, which is commonly defined as a partitioning of the input image into multiple disjoint regions, can be considered as an intermediate step in image analysis. It aims to find which part of the data array makes up an object in the real world. Segmentation is necessary for some important tasks such as measurement, visualization, registration, reconstruction, and content-based search, each having specific requirements. In fluorescence microscopy, accurate localization of the boundaries of the observed fluorescently labeled structural and functional units (cell
nuclei, genes, etc.) are particularly important. Quantitative cell nucleus image analysis plays an important role in numerous applications of analyzing and quantifying fluorescence microscopy images, and fluorescence microscopy image segmentation is one of the primary steps for all automatic image analysis in many high-throughput applications. It is a critical step because errors in the segmentation process almost certainly lead to inaccuracies in the subsequent quantitative analysis.

Fluorescence microscopy image segmentation is a challenging task because of the following problems:

- **Intensity Inhomogeneity:** this is a common problem in many real-world images from different modalities [21, 22]. In particular, it is frequently observed in a fluorescence microscopy image. As an example, the intensity inhomogeneity in cell images can appear as a variation of intensity throughout the image. Therefore, the intensities of the same region change with locations in the image.

- **Noise:** The signal measured for each pixel in fluorescence microscopy image can be affected by local noise which can seriously mislead automatic segmentation methods. In practice, in fluorescence microscopy imaging, the image can be corrupted by two types of noise: intrinsic and extrinsic. Intrinsic noise arises when a photon hits the detector screen and, therefore, creates a random number of light photons. Extrinsic noise is introduced by other sources during the image acquisition process. As an example, a charged coupled device (CCD) camera, which results in the production of negligibly small extrinsic noise, is utilized in most of the fluorescence microscope systems.
2.4.2 Datasets

In this thesis, the proposed approaches are applied to 2D fluorescence microscopy images of cell nuclei from four experiments which include different cell types. Two datasets from [23] which have ground truth are used. The first data set consists of 48 images, each with a size of 1349×1030 pixels and has 1831 U2OS Hoechst stained cell nuclei (see Figure 2.4 (a)). The second data set contains 49 images, each with a size of 1344×1024 pixels and has 2178 NIH3T3 Hoechst stained cell nuclei (see Figure 2.4 (b)). It is noted that since the images of the second data set suffer from intensity inhomogeneity, in comparison with the images in the first set, automatic analysis of the second set is more challenging.

![Figure 2.4: Original images of the four different datasets. (a) U2OS cells, (b) NIH3T3 cells, (c) BBBC005, (d) Synthetic microscopy images from SIMCEP.](image-url)
The third dataset consists of 17 images from BBBC005 [24], each with a size of 520 × 696 pixels and has approximately 159 cell nuclei (see Figure 2.4 (c)). Finally, fourth dataset consists of synthetic images of cell populations with realistic properties generated with the SIMCEP simulation tool [25]. The SIMCEP simulation tool is available from http://www.cs.tut.fi/sgn/csb/simcep/. The generated set includes 20 simulated images of cell populations with ground truth, each with a size of 400×400 pixels and includes approximately 400 cell nuclei (see Figure 2.4(d)).

2.4.3 Segmentation performance evaluation

In this thesis, the performance of the proposed segmentation algorithms is evaluated using region-based and contour-based measures. For region-based measure, the Jaccard coefficient [26], which is widely used to measure spatial overlap, as well as Dice false positive (Dice FP) and Dice false negative (Dice FN) are used. The Dice FP is used to measure the over-segmentation and Dice FN gives a measure of under-segmentation. For contour-based measures, the Hausdorff distance and mean absolute contour distance (MAD) are used. The Jaccard coefficient can be calculated as:

\[
\text{Jaccard}(R,S) = \frac{|R \cap S|}{|R \cup S|} \times 100
\]  

(2.1)

where R is the binary reference image and S is the binary segmented image. The Dice FP and Dice FN are obtained using the following equation:

\[
\text{Dice FP}(R,S) = \frac{2|R \cap S|}{|R| + |S|} \times 100
\]  

(2.2)
Dice \(FN(R, S) = \frac{2|R \cap S|}{|R| + |S|} \times 100\) \hspace{1cm} (2.3)

where \(\overline{R}\) and \(\overline{S}\) are the complements of the \(R\) and \(S\) respectively.

The Hausdorff distance can be calculated as:

\[\text{Hausdorff } (R, S) = \max_{i \in S_c} \{D(i)\}\] \hspace{1cm} (2.4)

where \(D(i)\) denotes the minimal Euclidean distance of pixel \(i\) to the contour of the reference object and \(S_c\) denotes the contour of the segmented object. For all \(n_{S_c}\) pixels on \(S_c\), the MAD is defined as:

\[\text{MAD} = \frac{1}{n_{S_c}} \sum_{i \in S_c} |D(i)|\] \hspace{1cm} (2.5)

2.5 Summary

In this chapter, first, a brief introduction of fluorescence microscopy imaging modality has been given. Then, fluorescence microscopy image segmentation and related challenges have been considered. Also, we introduced the datasets which are used in this thesis. The last part introduced region-based and contour-based measures which can be used for segmentation performance evaluation.
Chapter 3

Deformable Model Segmentation Approaches

In recent years, many methods have been applied to the segmentation of fluorescence microscopy images. Deformable models, which can capture a wide spectrum of different shapes, can be considered as a major category of cell segmentation techniques. This chapter provides the required theoretical background of deformable model approaches for image segmentation and an overview of existing methods.
3.1 Introduction

Recently, many methods have been proposed for the segmentation of cell nuclei in fluorescence microscopy images [1-3, 27-31]. Deformable models, which can capture a wide spectrum of different shapes, can be considered as a major category of cell segmentation techniques [32]. Since these approaches are numerically robust, deformable models have been widely used in fluorescence microscopy image segmentation. The aims of this chapter are to provide the reader an intuitive but also mathematical description of these model-based methods as well as describe their implementation aspects. Deformable models can be considered as curves or surfaces which deform based on the internal and external forces to segment the image. The internal forces are used to preserve the shape smoothness of the model while the external forces are utilized to drive the model to the desired region boundaries.

There are two main types of deformable models: parametric models [33], which use an explicit representation of objects [28-30], and implicit models [31, 34-45]. Parametric models are curves whose deformations can be determined by the displacement of a discrete number of control points along the curve. For two-dimensional image segmentation, parametric models can be defined as surfaces with the control points defining two-dimensional deformable grids. Usually, the convergence rate, which depends on the predetermined number of control points, is considered as the main advantage of parametric models. However, the main drawback of parametric models is that they are topology dependent. Therefore, only a single region can be captured.

Implicit models define the shape from the n-dimensional to an n + 1- dimensional domain using distance transformation. As will be explained below, such transformation
has three main advantages. First, since the shape is defined in a domain with dimensionality similar to the dataset space, a more mathematically straightforward integration of shape and appearance (image features) in the model definition can be provided by implicit models. Second, the shape is implicitly defined by implicit models using the control (deformation) points at the image pixels’ positions. Finally, methods that use such representations can be made topology independent. Therefore, multiple regions can be captured by a single model. This characteristic makes the model robust to initializations.

The remainder of this chapter is organized as follows. Section 2 reviews parametric models. In Section 3, the required background for the rest of the thesis is provided, and the mathematical definitions of the most representative implicit models are explained. Section 4 summarizes the chapter.

3.2 Parametric deformable models

Parametric deformable models, which was introduced in 1988 [33], utilize parametric curves to represent the model shape. In these models, the deformations are determined by geometry and kinematics as well as dynamics. From the mathematical point of view, parametric deformable models can be considered as splines, whose state (position and dynamics) is determined by an energy function, and their evolution is an energy minimization problem.

Let \( \Omega = \{ \Omega_i \}_{i=1}^2 \) denotes the image domain, \( C \subset \Omega \) denotes the closed, smooth segmenting curve defined by a set of ordered points \( \mathbf{x} \subset \Omega \). If the parametric domain be represented by \( s \), then the curve can be defined as \( s \rightarrow C(s) = \{ \mathbf{x}(s) \} \). The model that undergoes deformation based on the energies of \( \mathbf{x}(s) \) can be defined by this curve.
The internal energy of an active contour can be considered as the summation of forces applied to the curve to preserve its smoothness. The curve $C$ can be extended and shrunk depending on the internal energy. Hence, the smoothness of the model can be determined by the internal energy. Mathematically, the internal energy can be formulated as:

$$E_{\text{int}}(C) = -\int_0^1 e_{\text{int}}(C(s)) \, ds$$

(3.1)

where the individual energies are defined as follows:

$$e_{\text{int}}(C(s)) = \alpha(s) \left| \frac{\partial C}{\partial s} \right| + \beta(s) \left| \frac{\partial^2 C}{\partial^2 s} \right|$$

(3.2)

where $\alpha$ and $\beta$ are used to regulate the relative importance of the two smoothness terms. The length of the curve is determined by the first derivative of $C$ corresponds to the first-order smoothness of the curve. The second derivative can determine the smoothness in the direction normal to the curve. Therefore, minimum length and maximum smoothness can be obtained by the minimization of the energy in (3.1) and (3.2).

Commonly, image gradient is used by the parametric models to guide external image forces that drive a shape-based model. The external image forces can be formulated as follows:

$$E_{\text{ext}}(C) = -\int_0^1 \left| \nabla I(C(s)) \right| \, ds$$

(3.3)
where $\hat{I} = G_\sigma * I$ denotes the image $I$ after smoothing with a Gaussian kernel of standard deviation $\sigma$ and the image gradient along the curve $C$ is denoted by $\nabla I(C(s))$. The maximum of the accumulative image gradient along the curve is obtained by minimizing (3.3).

One of the main drawbacks of the parametric deformable models is that since they depend on image gradient information, they are sensitive to noise and spurious edges. Therefore, it is required to initialize the curve close to the boundary to avoid getting stuck in local minima.

Some parametric deformable methods utilize region-based external energies or combinations of both edge and region-based terms to overcome the mentioned problem. The main assumption behind these methods is that the parametric model $C$ segments the image into foreground and background region, each of which has different statistics. This difference locally derives the model evolution. In [46], a generalized energy function based on active contours and region growing is proposed, and the minimization of the objective function is guaranteed to converge to a local minimum.

However, because of the significant difference in representation for shape and appearance, the problem of unifying shape and appearance is not addressed. A region based module is utilized to develop a rough binary mask of the region of interest in other hybrid segmentation frameworks [47]. This rough boundary estimation is used to initialize the curve for a deformable model. Then, the gradient information is used to deform the curve.

### 3.3 Implicit deformable models

The implicit models form the second class of deformable models. The level-set based shape representation is used in these models, where curves are transformed into higher
dimensional scalar functions based on the scalar distance function, as shown in Figure 3.1. The surfaces are the distance functions while the gray planes indicate the zero level (zero distance). The blue circles represent the desired boundaries while the red disks correspond to the positive distance values from the current contour in each iteration [48].

The scalar distance can be formulated as follows:

\[
\begin{cases}
\phi(x) > 0, & \text{x is inside } \mathcal{C} \\
\phi(x) = 0, & \text{x is on } \mathcal{C} \\
\phi(x) < 0, & \text{x is outside } \mathcal{C}
\end{cases}
\]  

(3.4)

where \(\phi(x): \Omega \rightarrow \mathbb{R}\) a scalar Lipchitz continuous level set function.

![Figure 3.1: Implicit shape representation using the distance transform [48].](image)

The model’s shape is embedded in a higher-dimensional space of distance transforms, such that the zero-level of a scalar (Euclidean distance) function
corresponds to the evolving curve. The interface can delineate two regions in the image domain $\Omega$. The level set represents the model shape implicitly, where $\mathbf{x}$ denotes the image pixel location in Cartesian coordinates. The model's shape is transformed into a distance image by this representation $\varnothing(\mathbf{x})$. There are two main advantages for using this definition. First, since the image and the shape have the same dimensionality, the integration of the shape with the appearance can be straightforward. Second, the $C^1$ continuity constraint (first order smoothness) is satisfied by the shape distance function.

A typical objective function for edge-based implicit models that drives the front propagation of a level set (distance) function, is [49]:

$$E(C) = \int_0^1 g(|\nabla I(C(s))|)|C'(s)|ds$$

(3.5)

where $g(|\nabla I|) = \frac{1}{1+|\nabla I|^2}$ and the front curve of the evolving level set function is denoted by $C$. For the objective function minimization, the front curve deforms along its normal direction $C''(s)$, and the speed function $g(|\nabla I|)$ controls the speed. It is noted that the speed function definition is based on image gradient $\nabla I$, and it takes the positive value in homogeneous areas and zero at ideal edges. Using the above shape representation, Mumford and Shah's idea for image segmentation can be explained as follows: an observed image $I$ is given, the segmentation algorithm has to find a decomposition $\Omega_i$ of $\Omega$ and an optimal piecewise smooth estimation $u$ of $I$, under the following condition: $I$ changes smoothly within each $\Omega_i$, and rapidly or discontinuously through the boundaries of $\Omega_i$.

The following minimization problem has been proposed [50] to solve this problem:
inf \left\{ F^{MS}(I,C) = D_{MS} + \mu \int_{\Omega \setminus C} |\nabla I(x)|^2 \, dx + \rho \cdot |C| \right\} \tag{3.6}

where \(|C|\) indicates the length of \(C\), \(\mu\) and \(\rho\) are positive constants, and the first term, the data term, \(D_{MS}\) is defined as:

\[ D_{MS} = \int_\Omega |I(x) - u(x)|^2 \, dx \tag{3.7} \]

The existence of a minimizer for (3.6) has been considered [50, 51]. Assume that \(v_i\) is defined as follows:

\[ \{ v_i | \forall \Omega_i, v_i = \text{mean}(I(x)) \text{ in } \Omega_i, \text{for } i = 1,2 \} \tag{3.8} \]

Now the above model can be reduced to a minimal partition problem if and only if the segmented image can be restricted to a piecewise constant function and the following minimization can be implemented to solve it.

\[ \text{Min } E^{MS}(I,C) = D_{MS}(I,v_i) + \rho \cdot |C| \tag{3.9} \]

where \(D_{MS}(I,v_i) = \sum_i \int_{\Omega_i} |I(x) - v_i|^2 \, dx\). The main drawback of this method is that since the energy functional is not convex, in practice the optimal solution of (3.2) cannot be obtained easily. This method is also highly sensitive to inhomogeneity caused by noise and artifacts in real images.

Based on the Mumford and Shah model, Chan and Vese proposed a new version of “Active Contour without Edge” (ACWE) [52]:

\[ E^{ACWE}(I,v_i,C) = D_{ACWE}(I,v_i) + \rho \cdot |C| \tag{3.10} \]
where $\rho$ is positive constants and $\mathcal{D}_{\text{ACWE}}$ is defined as:

$$
\mathcal{D}_{\text{ACWE}}(I, \nu_i) = \int_{\text{inside}(c)} |I(x) - \nu_1|^2 \, dx + \int_{\text{outside}(c)} |I(x) - \nu_2|^2 \, dx
$$

(3.11)

where $\nu_1, \nu_2$ represent the average of intensities of $I$ inside and outside the segmenting curve, respectively. Let us define $M_1(\emptyset(x))$ and $M_2(\emptyset(x))$ as follows:

\[
\begin{align*}
M_1(\emptyset(x)) &= H(\emptyset(x)) \\
M_2(\emptyset(x)) &= 1 - H(\emptyset(x))
\end{align*}
\]

(3.12)

Then $\mathcal{D}_{\text{ACWE}}(I, \nu_i, \emptyset)$ is defined as follows:

$$
\mathcal{D}_{\text{ACWE}}(I, \nu_i, \emptyset(x)) = \sum_{i=1}^2 \int_{\Omega} |I(x) - \nu_i|^2 M_i(\emptyset(x)) \, dx
$$

(3.13)

The final energy functional of ACWE model can be defined as follows:

$$
E_{\text{ACWE}}(\nu_1, \nu_2, \emptyset(x)) = \mathcal{D}_{\text{ACWE}}(I, \nu_i, \emptyset) + \int_{\Omega} \delta(\emptyset(x)) |\nabla \emptyset(x)| \, dx
$$

(3.14)

where the Heaviside function $H(z)$ and Dirac delta function $\delta(z)$ are defined by $H(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{if } z \leq 0 \end{cases}$ and $\delta(z) = \frac{\partial H(z)}{\partial z}$, respectively. Using the standard technique of Euler-Lagrange, the level set function can be updated as follows:

$$
\frac{\partial \emptyset}{\partial t} = \delta(\emptyset) \left[ \mu \, \text{div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) - e_{\text{ACWE}_1} + e_{\text{ACWE}_2} \right]
$$

(3.15)

where $e_{\text{ACWE}_1}$ and $e_{\text{ACWE}_2}$ are defined as:
\[
\begin{align*}
  e_{\text{ACWE}_1} &= (I(x) - \nu_1)^2 \\
  e_{\text{ACWE}_2} &= (I(x) - \nu_2)^2
\end{align*}
\]  
(3.16)

At each iteration \( \nu_1 \) and \( \nu_2 \) are updated using the following equation:

\[
\nu_i = \frac{\int_{\Omega} I(x) M_i(\emptyset(x)) \, dx}{\int_{\Omega} M_i(\emptyset(x)) \, dx}
\quad \text{for } i = 1, 2.
\]  
(3.17)

One of the main assumptions behind the ACWE method is that each region in the image has a distinct mean intensity, and these intensities are assumed to be homogeneous in each region. The ACWE’s data term is based on the piecewise constant segmentation model [52-54] and can be considered as a specific case of the Gaussian distribution. However, it is impossible to guarantee this characteristic for the general image, which limits the application of the ACWE method.

The Bayesian based level set approach (BLS), as a generalization of the ACWE method, is proposed [55] where a more general variational formulation is obtained from the maximization of the a posteriori segmentation probability given an observed data. The data term of the BLS model is defined as follows:

\[
\mathcal{D}_{\text{BLS}}(I, \Sigma_i, \nu_i, \emptyset(x)) = \sum_{i=1}^{2} \int_{\Omega} -\log P(I(x)|x \in \Omega_i) M_i(\emptyset(x)) \, dx
\]  
(3.18)

where \( P(I(x)|x \in \Omega_i) \) is defined as:

\[
P(I(x)|x \in \Omega_i) = \frac{1}{\Sigma_i \sqrt{2\pi}} \exp\left( -\frac{(\nu_i - I(x))^2}{2\Sigma_i^2} \right)
\]  
(3.19)
where $\nu_i$ and $\Sigma_i$ denote the mean and the covariance matrix of the Gaussian distribution, respectively. The final energy functional of BLS model can be defined as follows:

$$E_{BLS} \left( I, \Sigma_i, \nu_i, \emptyset((\mathbf{x})) \right) = D_{BLS} \left( I, \Sigma_i, \nu_i, \emptyset \right) + \int_\Omega \delta \left( \emptyset(\mathbf{x}) \right) |\nabla \emptyset(\mathbf{x})| d\mathbf{x}$$

(3.20)

For the fixed $\emptyset$ and $i = 1, 2$, $E_{BLS}$ can be minimized with respect to the $\nu_i$ and $\Sigma_i$ by the following update equation:

$$\nu_i = \frac{\int_\Omega I(\mathbf{x})M_i(\emptyset(\mathbf{x}))d\mathbf{x}}{\int_\Omega M_i(\emptyset(\mathbf{x}))d\mathbf{x}}$$

(3.21)

$$\Sigma_i = \frac{\int_\Omega (\nu_i - I(\mathbf{x}))^2 M_i(\emptyset(\mathbf{x}))d\mathbf{x}}{\int_\Omega M_i(\emptyset(\mathbf{x}))d\mathbf{x}}$$

(3.22)

Minimization of the energy functional $E_{BLS}$ with respect to $\emptyset$ can be achieved by solving the following gradient descent flow equation:

$$\frac{\partial \emptyset}{\partial t} = \delta \left( \emptyset \right) \left[ \mu \text{ div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) - e_{BLS_1} + e_{BLS_2} \right]$$

(3.23)

where $e_{BLS_1}$ and $e_{BLS_2}$ are defined as follows:

$$\begin{cases} e_{BLS_1} = \log(\Sigma_1) + \frac{(\nu_1 - I(\mathbf{x}))^2}{2\Sigma_1^2} \\ e_{BLS_2} = \log(\Sigma_2) + \frac{(\nu_2 - I(\mathbf{x}))^2}{2\Sigma_2^2} \end{cases}$$

(3.24)

To improve the ACWE and BLS methods, some methods are proposed [4-7, 56-58]. The algorithm proposed in [56], in which an edge-based indicator was added into the region based model, usually yields better results in comparison with ACWE in terms of
accuracy of segmentation and locating low contrast edges. By minimizing the Mumford–Shah functional, two similar region-based active contour models which are based on a piecewise smooth description of the images, and widely known as piecewise smooth (PS) models, are proposed [57, 58]. Recently, local intensity information has been incorporated into the implicit models [4-7] for more accurate segmentation, especially in the presence of intensity inhomogeneity. The local binary fitting (LBF) model is proposed to overcome the difficulty caused by intensity inhomogeneities [4, 5]. The LBF model utilizes the local intensity information and defines two spatially varying fitting functions \( \nu_1(\mathbf{x}) \) and \( \nu_2(\mathbf{x}) \) for the local mean intensities approximation.

For a given point, \( \mathbf{x} \in \Omega \), the data term of the local intensity fitting energy is defined as:

\[
\mathcal{D}_{\text{LBF}_x} = \int_{\text{inside}(c)} \omega(\mathbf{x} - \mathbf{y})|I(\mathbf{x}) - \nu_1(\mathbf{x})|^2 \, d\mathbf{y} + \int_{\text{outside}(c)} \omega(\mathbf{x} - \mathbf{y})|I(\mathbf{x}) - \nu_2(\mathbf{x})|^2 \, d\mathbf{y}
\]

(3.25)

where \( \omega(.) \) denotes a Gaussian function (more details about \( \omega \) is presented in Chapter 5). The above local data term, \( \mathcal{D}_{\text{LBF}_x} \), is defined for a centre point \( \mathbf{x} \). Finally, the aim is to define the data term of the LBF model for all the centre points \( \mathbf{x} \) in the image domain \( \Omega \), therefore \( \mathcal{D}_{\text{LBF}}(I, \nu_1(\mathbf{x})) = \int_\Omega \mathcal{D}_{\text{LBF}_x} \, d\mathbf{x} \) denotes the final data term. Using the level set function to represent the contour, \( \mathcal{D}_{\text{LBF}} \) can be rewritten as:

\[
\mathcal{D}_{\text{LBF}}(I, \nu_1(\mathbf{x}), \emptyset) = \sum_{i=1}^2 \int_\Omega \left( \int_\Omega \omega(\mathbf{x} - \mathbf{y})|I(\mathbf{x}) - \nu_i(\mathbf{x})|^2 M_i(\emptyset(\mathbf{x})) \, d\mathbf{y} \right) \, d\mathbf{x}
\]

(3.26)

For more accurate computation involving the level set function and its evolution, the level set function is regularized by penalizing its deviation from a signed distance
function. Also, the smoothness of the boundary is controlled by regularizing the length of the contour. The two terms above can be described by the following energy function:

\[ \mathcal{R}(\phi) = \mathcal{L}(\phi) + \mathcal{P}(\phi) \]  

(3.27)

where \( \mathcal{L}(\phi) \) and \( \mathcal{P}(\phi) \) are defined as follows:

\[
\begin{align*}
\mathcal{L}(\phi(x)) &= \int_{\Omega} \delta(\phi) |\nabla(\phi)| \, dx \\
\mathcal{P}(\phi(x)) &= \int_{\Omega} \frac{1}{2} (|\nabla \phi| - 1)^2 \, dx
\end{align*}
\]

(3.28)

Therefore, the final energy functional of LBF model can be defined as follows:

\[ E_{\text{LBF}} = D_{\text{LBF}}(I, \nu_i(x), \phi) + \mathcal{R}(\phi) \]  

(3.29)

For the fixed \( \phi \) and \( i = 1, 2 \), \( E_{\text{LBF}} \) can be minimized with respect to the \( \nu_i(x) \) by the following update equation:

\[ \nu_i(x) = \frac{\omega(x) I(x) M_i(\phi(x))}{M_i(\phi(x))} \]  

(3.30)

Minimization of the energy functional \( E_{\text{LBF}} \) with respect to \( \phi \) is achieved by solving the following gradient descent flow equation:

\[ \frac{\partial \phi}{\partial t} = \delta(\phi) \left( \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - e_{\text{LBF}_1} + e_{\text{LBF}_2} \right) + \left( \nabla^2 \phi - \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right) \]  

(3.31)

where \( e_{\text{LBF}_1} \) and \( e_{\text{LBF}_2} \) are defined as:
\[
\begin{aligned}
e_{\text{LBF}_1} &= \int_{\Omega} \omega(\mathbf{x} - \mathbf{y}) |I(\mathbf{x}) - v_1(\mathbf{x})|^2 \, d\mathbf{y} \\
e_{\text{LBF}_2} &= \int_{\Omega} \omega(\mathbf{x} - \mathbf{y}) |I(\mathbf{x}) - v_2(\mathbf{x})|^2 \, d\mathbf{y}
\end{aligned}
\]  
(3.32)

Since the local intensity mean is utilized in the LBF model, it provides desirable segmentation results in the presence of intensity inhomogeneity. Also, the regularization term is used to avoid the time-consuming re-initialization step. Also, the LBF model is sensitive to the noise and initialization [8], which limits its practical applications.

Similar to the LBF model, the local Gaussian distribution fitting energy method (LGDFE) [7] is proposed based on the BLS to overcome the difficulty caused by intensity inhomogeneities. The LGDFE model describes local image intensities by Gaussian distributions with different means and variances. For a given point, \( \mathbf{x} \in \Omega \), the data term of the LGDFE is defined as:

\[
\mathcal{D}_{\text{LGDFE}_x} = \sum_{i=1}^{2} \int_{\Omega_i} -\omega(\mathbf{x} - \mathbf{y}) \log p_x(I(\mathbf{y})|\mathbf{y} \in \Omega_i) \, d\mathbf{y}
\]  
(3.33)

where \( p_x(I(\mathbf{y})|\mathbf{y} \in \Omega_i) \) is defined as:

\[
p_x(I(\mathbf{y})|\mathbf{y} \in \Omega_i) = \frac{1}{\Sigma(\mathbf{x})_i \sqrt{2\pi}} \exp\left(-\frac{(v_i(\mathbf{x}) - \mathbf{x})^2}{2\Sigma(\mathbf{x})_i^2}\right)
\]  
(3.34)

where \( v_i(\mathbf{x}) \) and \( \Sigma(\mathbf{x})_i \) denote the local intensity mean and the covariance matrix of the Gaussian distribution. Finally, the aim is to define the data term of the LGDFE model for all the centre points \( \mathbf{x} \) in the image domain \( \Omega \), therefore \( \mathcal{D}_{\text{LGDFE}}(I, v_i(\mathbf{x}), \Sigma(\mathbf{x})_i) = \int_{\Omega} \mathcal{D}_{\text{LGDFE}_x} \, d\mathbf{x} \) denotes the final data term. Using the level set function to represent the contour \( C \), \( \mathcal{D}_{\text{LGDFE}} \) can be rewritten as:
\[ \mathcal{D}_{\text{LGDFE}}(l, v_i(x), \Sigma(x)_i, \emptyset) = \]
\[ \sum_{i=1}^{2} \int_{\Omega} \left( \int_{\Omega} -\frac{\omega(x - y)}{M_i(\emptyset(y))} \log P_x(l(y) | y \in \Omega_i) \, dy \right) \, dx \]  
(3.35)

Therefore, the final energy functional of LGDFE model can be defined as follows:

\[ E_{\text{LGDFE}} = \mathcal{D}_{\text{LGDFE}}(l, v_i(x), \Sigma(x)_i, \emptyset) + \mathcal{R}(\emptyset) \]  
(3.36)

For the fixed \( \emptyset \) and \( i \in [1,2] \), \( E_{\text{LGDFE}} \) can be minimized with respect to the \( v_i(x) \) by the following update equation:

\[ v_i(x) = \frac{\int_{\Omega} \omega(y - x) l(y) M_i(\emptyset(y)) \, dy}{\int_{\Omega} g(y - x) M_i(\emptyset(y)) \, dy} \]  
(3.37)

\[ \Sigma(x)_i^2 = \frac{\int_{\Omega} \omega(y - x) (v_i(x) - l(x))^2 M_i(\emptyset(y)) \, dy}{\int_{\Omega} \omega(y - x) M_i(\emptyset(y)) \, dy} \]  
(3.38)

Minimization of the energy functional \( E_{\text{LGDFE}} \) with respect to \( \emptyset \) is achieved by solving the following gradient descent flow equation:

\[ \frac{\partial \emptyset}{\partial t} = \delta(\emptyset) \left( \text{div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) - e_{\text{LGDFE}1} + e_{\text{LGDFE}2} \right) + \left( \nabla^2 \emptyset - \text{div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) \right) \]  
(3.39)

where \( e_{\text{LGDFE}1} \) and \( e_{\text{LGDFE}2} \) are defined as:

\[ e_{\text{LGDFE}1} = \int_{\Omega} \omega(y - x) \left( \log \sigma_1(x) + \frac{(v_1(x) - l(x))^2}{2\Sigma(x)_1^2} \right) \, dy \]

\[ e_{\text{LGDFE}2} = \int_{\Omega} \omega(y - x) \left( \log \sigma_2(x) + \frac{(v_2(x) - l(x))^2}{2\Sigma(x)_2^2} \right) \, dy \]  
(3.40)
Another approach which has been proposed to improve the ACWE method is the local Chan–Vese model (LCV) which uses both global image information and locally filtered image information for image segmentation [6]. For the LCV model, the data term, $\mathcal{D}_{LCV}$, consists of two terms: global term which uses the $\mathcal{D}_{ACWE}$, and local term $\mathcal{D}_{LACWE}$. The local term uses the difference image $\tilde{I}(x) = g_q I(x) - I(x)$, where $g_q$ is an averaging filter with $q \times q$ size. $\mathcal{D}_{LCV}$ can be formulated as follows:

$$\mathcal{D}_{LCV}(I, \tilde{I}, v_i, d_i, \phi) = \mathcal{D}_{ACWE}(I, v_i, \phi) + \mathcal{D}_{LACWE}(\tilde{I}, d_i, \phi)$$

(3.41)

where $\mathcal{D}_{LACWE}(\tilde{I}, d_i, \phi)$ is defined as:

$$\mathcal{D}_{LACWE}(\tilde{I}, d_i, \phi) = \sum_{i=1}^{2} \int_{\Omega} |\tilde{I}(x) - d_i|^2 M_i(\phi(x)) dx$$

(3.42)

where $d_i$ denotes the intensity averages of the difference image. The final energy functional of LCV model can be defined as follows:

$$E_{LCV} = \mathcal{D}_{LCV}(I, \tilde{I}, v_i, d_i, \phi) + \mathcal{R}(\phi)$$

(3.43)

In numerical implementation, for fixed $\phi$ and $i = 1, 2$, the following variational formulations can be obtained to minimize $E_{LCV}$:

$$v_i = \frac{\int_{\Omega} I(x) M_i(\phi(x)) dx}{\int_{\Omega} M_i(\phi(x)) dx}$$

(3.44)

$$d_i = \frac{\int_{\Omega} \tilde{I}(x) M_i(\phi(x)) dx}{\int_{\Omega} M_i(\phi(x)) dx}$$

(3.45)

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[ \mu \text{div} \left( \frac{v\phi}{|v\phi|} \right) - e_{ACWE_1} + e_{ACWE_2} - e_{LACWE_1} + e_{LACWE_2} \right]$$
\[ + \left( \nabla^2 \phi - \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right) \]  

(3.46)

where \( e_{\text{LACWE}_1} \) and \( e_{\text{LACWE}_2} \) are defined as follows:

\[
\begin{align*}
\left( e_{\text{LACWE}_1} = \right) & \left( I(x) - d_1 \right)^2 \\
\left( e_{\text{LACWE}_2} = \right) & \left( I(x) - d_2 \right)^2
\end{align*}
\]

(3.47)

It is noted that \( \mathcal{D}_{\text{LACWE}} \) is similar to the \( \mathcal{D}_{\text{ACWE}} \) but the original image \( I(x) \) is replaced by \( I(x) \) and \( I(x) \) is transformed to \( I(x) \) using the high-pass filter.

An improved implicit model is proposed in [8] which introduces energy functional with a local intensity fitting term and auxiliary global intensity fitting term. The Local intensity fitting term is dominant near object boundaries and attracts the contour toward object boundaries. The auxiliary global intensity fitting term incorporates global image information to improve the robustness of the proposed method. An implicit model with selective local or global segmentation is proposed in [9] which utilizes the advantages of the ACWE and GAC[49, 59] models. The statistical information inside and outside the contour is used to construct a region-based signed pressure force (SPF) function [60], which can control the direction of evolution. In [10], a local intensity clustering property is derived, and, therefore, a local clustering criterion function for the intensities in a neighborhood of each point is defined. To define the energy functional, the local clustering criterion is integrated over the neighborhood center to define an energy functional, which is converted to a level set formulation. An implicit model for bias correction and segmentation for images with intensity inhomogeneities is proposed in [11]. The proposed method defines a function for clustering the image pixels in a small neighborhood. The cluster centers in this objective function have a multiplicative factor that estimates the bias within the neighborhood. An implicit model for image
segmentation based on the local intensity is proposed in [61]. In particular, an improved region fitting term is utilized to partition the regions of interests in images depending on the local statistics regarding the intensity and the magnitude of the gradient in the neighborhood of a contour. In [62], an adaptive active contours without edges (AACWE) model is proposed based on the ACWE. In AACWE, the Sobolev gradient [63] is utilized to develop the segmentation framework based on the PDE formulation for two-phase image segmentation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACWE</td>
<td>ACWE is proposed based on the techniques of curve evolution and Mumford-Shah functional.</td>
<td>[52]</td>
</tr>
<tr>
<td>BLS</td>
<td>BLS proposed a variational formulation obtained from a Bayesian model.</td>
<td>[55]</td>
</tr>
<tr>
<td>LBF</td>
<td>LBF is a region-based active contour model that is able to utilize image information in local regions.</td>
<td>[4]</td>
</tr>
<tr>
<td>RSFE</td>
<td>In RSFE, a data fitting energy is defined regarding a contour and two fitting functions that locally approximate the image intensities on the two sides of the contour.</td>
<td>[5]</td>
</tr>
<tr>
<td>LRAC</td>
<td>LARC considers local rather than global image statistics and evolves a contour based on local information.</td>
<td>[64]</td>
</tr>
<tr>
<td>LGDFE</td>
<td>In LGDFE, the local image intensities are described by Gaussian distributions with different means and variances.</td>
<td>[7]</td>
</tr>
<tr>
<td>LGIF</td>
<td>LGIF induces a local force to attract the contour and stops it at object boundaries, and an auxiliary global intensity fitting term, which drives the motion of the contour far away from object boundaries.</td>
<td>[8]</td>
</tr>
</tbody>
</table>
SBGFRLS first selectively penalizes the level set function to be binary, and then uses a Gaussian smoothing kernel to regularize it. [9]

LCV utilizes both global image information and local image information for image segmentation. [6]

LGDF: In LGDF the local image intensities are characterized by Gaussian distributions with different means and variances. [76]

LSBR proposes an alternative criterion derived from the Bayesian risk classification error for image segmentation. [82]

KLS: KLS investigates level set multiphase image segmentation by kernel mapping and piecewise constant modeling of the image data. [102]

AACWE: AACWE is based on the ACWE and the Sobolev gradient. [62]

LLC: In each local region, LLC defines a locally weighted least squares energy to fit a linear classifier. [68]

### 3.4 Discussion

However, it should be emphasized that in our view, the above modifications do not mean that limitations on implicit models are removed. Defining a proper data term is a key issue in image segmentation using implicit models. The data term can be defined based on the different assumptions and utilizes different sources of information (intensity information, spatial information). Many of the above implicit models still suffer from intensity inhomogeneity, outliers, and intensity overlapping. For example, the ACWE and the BLS models can suffer from noise and intensity inhomogeneity. The LBF model and the LGDFE are sensitive to the noise and initialization [8], which limits their practical applications. Also, most of the above approaches assume that the global image data or local image data can be described by Gaussian distribution or are linearly separable [4-11].

In addition to the incorporation of local spatial information, two main categories of local implicit models are proposed. In the first category [4, 5, 7, 11, 64-69], a local data
term is defined based on the local image data around a neighborhood of a point and a weighting function. The local energy is then integrated over the entire image domain to form a final data term. Implicit approaches in the second category [6, 8, 70-73], use information from different sources and the data term consists of two terms, i.e., global term and local term.

To improve over existing approaches in the first category, in Chapter 4, we propose an implicit model based on the Bayesian classification risk and local spatial information. To improve the robustness of the model to noise and outliers, we utilize image patch information and define an anisotropic weighting scheme. In Chapter 5, we show that the data terms of the ACWE model and the LBF model can be rewritten based on the k-means clustering. Since the LBF model uses the mean square error (MSE) and is based on the Gaussianity and linearity assumptions, it is sensitive to noise and outliers. We applied the correntropy criterion to tackle this difficulty. Specifically, we propose an implicit model based on the correntropy criterion and show that the kernel function used in the correntropy criterion satisfies Mercer’s Theorem [74]. To define a more general data term based on the non-Gaussianity assumption, in Chapter 6, the local image intensities data are mapped implicitly into a higher dimension using a linear ensemble of basis kernels.

To improve proposed approaches in the second category, in Chapter 7, we assume that the global image data are nonlinearly separable and utilize the multiple kernels to define the data term. With multiple kernels, the data term gain more flexibility on kernel selections and also reflect the fact that, in practical learning problems, data from multiple heterogeneous or homogeneous sources can be used. It is noted that, in this chapter, we use the term “multiple kernel” in a wider sense than the one used in chapter
six. We utilize composite kernels constructed by multiple kernels for flexible information fusion.

3.5 Summary

This chapter provides the theoretical background of deformable approaches for image segmentation and an overview of existing methods. We provide a theoretical and mathematical background of some well-known existing implicit models. The data terms of the methods have been considered, and we highlighted differences between contributions in terms of their approach, assumptions, and key steps. To keep the review concise, we did not enter into details of the derivation of each work.
In this chapter, a new implicit model in a variational level set formulation is developed where our new energy functional minimizes the Bayesian classification risk. For the proposed data term, it has been assumed that the image data in the local image patch follows the Gaussian distribution and is linearly separable. Since the model based on this assumption can be sensitive to noise and outliers, a weighting scheme is used to enable the pixels in each image patch to have anisotropic weights. The performance of the proposed framework is evaluated using a large number of fluorescence microscopy images from four datasets with different cell types. A quantitative comparison is also performed with several existing segmentation approaches. The proposed algorithm obtains the lowest MAD measure for all four experiments. The MAD values are approximately 18%, 7%, and 12% better than the state-of-the-art methods we compared with for U20S cells, NIH3T3 cells, and Synthetic cells, respectively.
4.1 Introduction

The goal of this chapter is motivated by the fact that in cell nuclei image segmentation, pixels in an image patch possess nearly the same intensity. Therefore, the spatial relationship of the pixels in an image patch can be utilized as an important characteristic that improves the performance of segmentation methods [6, 7, 75, 76]. Consequently, the goal is to develop a segmentation algorithm based on image patch information [19, 77-79].

The contribution of this chapter is as follows. A new approach is introduced for cell nucleus segmentation in fluorescence microscopy images. An implicit model in a variational level set formulation, which is based on the image patch information is proposed to segment the image. Compared to previous approaches, we define a novel local energy functional based on the Bayesian classification risk [80-82] for an image patch. Also, a weighting scheme is used to enable the pixels in each image patch to have anisotropic weights.

The remainder of this chapter is organized as follows. Section 2 introduces the proposed approaches. In Section 3, experimental results are presented and analyzed using different cell types. The results of the proposed methods are also compared with previous approaches. This chapter is summarized in Section 4.

4.2 LLBWIP

4.2.1 Modelling

Assuming that an image is formed by two regions, cell and background pixels, the following two hypotheses can be used to characterize the image segmentation.

- A null hypothesis $H_1$, in which the cell is absent,
• An alternative hypothesis $H_2$, in which the cell is present.

The segmentation method is utilized to decide which hypothesis is correct. Therefore, one of two decisions can be made:

• $D_1$: the classifier declares that the cell is absent.

• $D_2$: the cell is present, and thus should be chosen by segmentation procedure.

The following four conditional probabilities are defined for the combinations of decisions in the hypothesis test:

(i) $P(D_1|H_1)$ is the probability of declaration that the cell is absent when it is actually absent.

(ii) $P(D_2|H_1)$ is the probability of declaration that the cell is present when it is absent.

(iii) $P(D_1|H_2)$ is the probability of declaration that the cell is absent when it is present.

(iv) $P(D_2|H_2)$ is the probability of declaration that the cell is present when it is actually present.

Using the statistical terminology, the first probability, which is the probability of rejecting the null hypothesis $H_1$ when it is actually true, is called type I risk. On the other hand, $P(D_1|H_2)$, which can be considered the probability of accepting $H_1$ when $H_1$ is actually false, is called type II risk. The consequence of each combination of hypothesis and decision is quantified with an associated loss. The losses of $P(D_1|H_1)$, $P(D_2|H_1)$, $P(D_1|H_2)$, and $P(D_2|H_2)$ can be denoted as $L(1, 1)$, $L(2, 1)$, $L(1, 2)$, and $L(2, 2)$, respectively. $L(1, 1)$ and $L(2, 2)$ can be viewed as the losses arising from the correct decision while $L(2, 1)$ and $L(1, 2)$ denote the losses that arise from the incorrect decision. In this study, $L(1, 1)$ and $L(2, 2)$ are set to zero. $L(2, 1)$ and $L(1, 2)$ are
expected to be one. Now the Bayesian risk for segmenting an image into cell and background can be written as follows:

\[ r = P(H_1)P(D_2|H_1) + P(H_2)P(D_1|H_2) \]  

(4.1)

Let \( \Omega = \{\Omega_i\}_{i=1}^{2} \) denotes the image domain and \( I: \Omega \rightarrow \mathbb{R}^+ \) denotes a given image, for each point \( x \) in the image domain \( \Omega \), the image patch centred on \( x \) can be represented as:

\[ P_x = (I(y), y \in N_x) \]  

(4.2)

where \( N_x \) can be considered as a \( q \times q \) neighbourhood of point \( x \). Now the image patch \( P_x \) with domain \( N_x \) can be partitioned by \( \{\Omega_i\}_{i=1}^{2} \) into the following disjoint regions:

\[ R_1 = \{\Omega_1 \cap N_x\} \text{ and } R_2 = \{\Omega_2 \cap N_x\}, \text{ where } N_x = \bigcup_{i=1}^{2} R_i, R_i \cap R_j = \emptyset \text{ } \forall i \neq j. \]

\( D_1 \) and \( D_2 \) can now be redefined as follows; all pixels of the image patch \( P_x \) that lead the segmentation procedure to choose decision \( D_1 \) fall in the region \( R_1 \), whereas all pixels of the image patch \( P_x \) that result in decision \( D_2 \) fall in the region \( R_2 \). The two hypotheses, \( H_1 \) and \( H_2 \), are considered in association with pdfs \( P(I(x)|H_1) \) and \( P(I(x)|H_2) \), respectively, where \( I(x) \) denotes a pixel value of the \( N_x \). The integral of \( P(I(x)|H_1) \) over the region \( R_2 \) represents the risk \( P(D_1|H_1) \) and the integral of \( P(I(x)|H_2) \) over the region \( R_1 \) represents the risk \( P(D_1|H_2) \). Now the data term, based on the Bayesian risk for segmenting an image patch \( P_x \) into cell and background, can be written as follows:

\[ \mathcal{D}_{LLBWIP_x} = \mathcal{D}_{LLBWIP_x1} + \mathcal{D}_{LLBWIP_x2} \]  

(4.3)

where \( \mathcal{D}_{LLBWIP_x1} \) and \( \mathcal{D}_{LLBWIP_x2} \) are defined as follows:
\[
\begin{align*}
\mathcal{D}_{LLBWIP_{x_1}} &= \int_{R_1} \ln(P(I(y) \mid y \in R_1)P(y \in R_1)) dy \\
\mathcal{D}_{LLBWIP_{x_2}} &= \int_{R_2} \ln(P(I(y) \mid y \in R_2)P(y \in R_2)) dy
\end{align*}
\] (4.4)

Figure 4.1. Two image patches selected from a synthetic image, (middle column) the calculated mean-square deviation and (right column) the estimated weight vectors.

Since dissimilar pixels should have different weights to reflect their decrease of importance [83-86], a weighting function needs to be defined and incorporated into (4.4) to constrain the influence of dissimilar pixels. To define the weighting function for each pixel \(y\), first, its mean square deviation \(\sigma_{yx}\) is calculated as follows:

\[
\sigma_{yx} = \left( \frac{\int_{y \in N \setminus \{y\}} (I(y') - I(y))^2}{n_x - 1} \right)^{\frac{1}{2}}
\] (4.5)

Then, the following exponential kernel function is utilized to produce the weight for pixel \(y\):

\[
\zeta_{yx} = \exp \left( - \left( \sigma_{yx} - \frac{\int_{y \in N_x} \sigma_{yx}}{n_x} \right) \right)
\] (4.6)
Finally, the weights are normalized by:

$$\kappa_{yx} = \frac{\eta_{yx}}{\int_{y \in N_x} \eta_{yx}}$$  \hspace{1cm} (4.7)

The weighting function can now be defined as follows:

$$\omega_y = \begin{cases} 
\kappa_{yx}, & \text{for } I(y) \in P_x \\
0, & \text{for } I(y) \notin P_x 
\end{cases}$$ \hspace{1cm} (4.8)

It is noted that $\omega_y$ is well defined on $N_x$ and $\int_{N_x} \omega_y \, dy = 1$. With this weighting function, for $y \in N_x$ if the mean square deviation of $I(y)$ is far away from the average mean-square deviation in the $P_x$, $I(y)$ is expected to have a very small weight. The calculated weight vectors of two different image patches are shown in Figure 4.1. The pixel values in both patches are given in the first column. The mean square deviations $\sigma_{yx}$ are given in the middle column and the right column shows the obtained weights. The first image patch (the upper row) is selected from the boundary area while the second image patch (the bottom row) is selected from the homogeneous area. In the first patch, the weight of the pixel with high intensity (over 200) tends to zero. Also in this patch the pixels with intensity values 49 and 81 are considered as noisy pixels and have negligible weights. In the second image patch, pixels with intensity values 81, 80, 161 and 164 are all considered as noise and their weights also tend to zero. As can be seen from the calculated weight vectors in both examples, the anisotropic weighting method can effectively reduce the impact of edges and noisy pixels on images. Using this weighting function, (4.4) is reformulated as:
\[
\begin{align*}
\mathcal{D}_{\text{LLBWIP}} x_1 &= \int_{\Omega_1} \omega \ln(P(I(\mathbf{y}) | \mathbf{y} \in R_2) P(\mathbf{y} \in R_2)) \, d\mathbf{y} \\
\mathcal{D}_{\text{LLBWIP}} x_2 &= \int_{\Omega_2} \omega \ln(P(I(\mathbf{y}) | \mathbf{y} \in R_1) P(\mathbf{y} \in R_1)) \, d\mathbf{y}
\end{align*}
\]

We assume that \(P(I(\mathbf{y}) | \mathbf{y} \in R_2)\) and \(P(I(\mathbf{y}) | \mathbf{y} \in R_1)\) follow the Gaussian distribution. The Gaussian distribution is one of the most commonly utilized density estimators, due to the fact that analytical methods can be utilized to estimate its parameters. Since the Gaussian function has the highest entropy compared with all distributions with the same variance, it is the best model when only the mean and variance are available. The utilization of the Gaussian distribution is also justified by the central limit theorem. Based on the central limit theorem, in fairly general terms, when the number of random variables increases, the mean of a set of random variables approaches a Gaussian distribution [87]. Now (4.9) can be written as follows:

\[
\begin{align*}
\mathcal{D}_{\text{LLBWIP}} x_1 &= \int_{\Omega_1} \omega \ln \left( \mathcal{N}(I(\mathbf{y}) | \mu_2(\mathbf{y}), \Sigma_2(\mathbf{y})) P(\mathbf{y} \in R_2) \right) \, d\mathbf{y} \\
\mathcal{D}_{\text{LLBWIP}} x_2 &= \int_{\Omega_2} \omega \ln \left( \mathcal{N}(I(\mathbf{y}) | \mu_1(\mathbf{y}), \Sigma_2(\mathbf{y})) P(\mathbf{y} \in R_1) \right) \, d\mathbf{y}
\end{align*}
\]

(4.10)

where \(\mu_i(\mathbf{y})\) and \(\Sigma_i(\mathbf{y})\) denote the mean and the covariance matrix, respectively, of the Gaussian distribution.

To incorporate the bias field that corrects for the intensity inhomogeneity into the data term \(\mathcal{D}_{\text{LLBWIP}} x\), the true intensity \(v_i\) in each region \(R_i\) is assumed to be constant. Then \(\mu_i(\mathbf{y})\) can be approximated as follows [84, 85]:

\[
\mu_i(\mathbf{y}) = b(\mathbf{y}) v_i \quad \text{for } i=1, 2.
\]

(4.11)

where \(b(\mathbf{y})\) denotes the bias field at each pixel. Therefore, (4.10) can be reformulated as:
\[
\begin{align*}
\mathcal{D}_{\text{LLBWIP}_x_1} &= \int_{\Omega_1} \omega_y \ln \left( \mathcal{N}(l(y) | b(y), \Sigma_2(y)) P(y \in R_2) \right) \, dy \\
\mathcal{D}_{\text{LLBWIP}_x_2} &= \int_{\Omega_2} \omega_y \ln \left( \mathcal{N}(l(y) | b(y), \Sigma_1(y)) P(y \in R_1) \right) \, dy
\end{align*}
\]

(4.12)

Figure 4.2. Segmentation results of the proposed approach. (a) original images, (b) segmentation results.

Two prior probabilities \(P(y \in R_2)\) and \(P(y \in R_1)\) are still unknown. Instead of assuming that the prior probabilities are equal for both regions [70], an iterative algorithm derived from the concavity of the Kullback–Leibler information number is
adopted for prior probabilities estimation [82, 88]. Finally, the data term is obtained by integrating over all \( x \) in the image domain \( \Omega \),

\[
D_{\text{LLBWIP}} = \int_{\Omega} D_{\text{LLBWIP}}(x) \, dx.
\]

### 4.2.2 Level set formulation

We assume that the two regions \( R_1 \) and \( R_2 \) can be represented by the regions separated by the Lipchitz function \( \Phi \), \( \Omega_1 = \{ \Phi(x) > 0 \} \) and \( \Omega_2 = \{ \Phi(x) \leq 0 \} \). By incorporating the Heaviside function \( H \), \( D_{\text{LLBWIP}_1} \) and \( D_{\text{LLBWIP}_2} \) are rewritten as follows:

\[
\begin{align*}
D_{\text{LLBWIP}_1} &= \int_{\Omega} \omega_y \ln \left( \mathcal{N} \left( I(y) | b(y)v_2, \Sigma_2(y) \right) P(y \in R_2) M_1(\Phi(y)) \right) dy \\
D_{\text{LLBWIP}_2} &= \int_{\Omega} \omega_y \ln \left( \mathcal{N} \left( I(y) | b(y)v_1, \Sigma_1(y) \right) P(y \in R_1) M_2(\Phi(y)) \right) dy
\end{align*}
\]

(4.13)

Therefore, the energy functional \( E_{\text{LLBWIP}} \) can be rewritten as follows:

\[
E_{\text{LLBWIP}} = \int_{\Omega} \left( D_{\text{LLBWIP}_1} + D_{\text{LLBWIP}_2} \right) dx + \gamma \mathcal{L}(\Phi) + \mathcal{P}(\Phi)
\]

(4.14)

where \( \delta(z) = \frac{\partial H(z)}{\partial z} \) and \( \mathcal{L}(\Phi) \) is used to control the smoothness of the zero level set and avoids the formation of small, isolated regions in the final segmentation. \( \mathcal{P}(\Phi) \) is used to eliminate the need for re-initialization in our method [6]. The parameter \( \gamma \) can be understood as the parameter which controls the penalization effect of the length term. For small \( \gamma \), smaller objects will be detected while for large values of \( \gamma \), larger objects will be detected. In this work, \( \gamma \) is set to 255\( \times \)255\( \times \)10\(^{-3} \) [11]. The Heaviside function \( H(z) \) can be approximated by a smooth function \( H_{\varepsilon}(z) \), which is defined as:
\[ H_\varepsilon(z) = \frac{1}{2} \left[ 1 + \frac{2}{\pi} \arctan \left( \frac{z}{\varepsilon} \right) \right] \]  

(4.15)

where \( \varepsilon \) is a positive constant (see Figure 4.3 (a)). Therefore, the energy functional is approximated as follows:

\[ E_{\text{LLBWIP}_\varepsilon} = \int_{\Omega} \left( D_{\text{LLBWIP}_{x_1}\varepsilon} + D_{\text{LLBWIP}_{x_2}\varepsilon} \right) dx + \gamma L_\varepsilon(\emptyset) + \mathcal{P}(\emptyset) \]

(4.16)

where \( D_{\text{LLBWIP}_{x_1} \varepsilon} \), \( D_{\text{LLBWIP}_{x_2} \varepsilon} \), and \( L_\varepsilon(\emptyset) \) approximate \( D_{\text{LLBWIP}_{x_1}} \), \( D_{\text{LLBWIP}_{x_2}} \), and \( L(\emptyset) \), respectively. Figure 4.2 shows visual examples of the segmentation results for the proposed functional.

4.2.3 Minimization

In numerical implementation, for fixed \( \emptyset \) and \( i = 1,2 \), the variables \( \Sigma_i(y) \), \( b(y) \), and \( v_i \) are updated as follows:

\[ \Sigma_i(y) = \frac{\int_{\Omega} \omega_y (l(y) - b(y)v_i)^2 P(y \in R_i) M_{i,\varepsilon}(\emptyset(y))) dy}{\int_{\Omega} \omega_y P(y \in R_i) M_{i,\varepsilon}(\emptyset(y))) dy} \]  

(4.17)

\[ b(y) = \frac{\sum_{i=1}^{2} \int_{\Omega} \omega_y (l(y)\Sigma_i(y)^{-1}v_i) P(y \in R_i) M_{i,\varepsilon}(\emptyset(y))) dy}{\sum_{i=1}^{2} \int_{\Omega} \omega_y (v_i(y)\Sigma_i(y)^{-1}v_i) P(y \in R_i) M_{i,\varepsilon}(\emptyset(y))) dy} \]  

(4.18)

\[ v_i = \frac{\int_{\Omega} \omega_y b(y)\Sigma_i(y)^{-1} P(y \in R_i) l(y) M_{i,\varepsilon}(\emptyset(y))) dy}{\int_{\Omega} \omega_y M_{i,\varepsilon}(\emptyset(y)) dy} \]  

(4.19)
Minimization of the energy functional $E_{\text{LLBWIP}}$ with respect to $\emptyset$ is achieved by solving the following gradient descent flow equation:

$$\frac{\partial \emptyset}{\partial t} = \delta (\emptyset) F + \left( \nabla^2 \emptyset - \text{div}\left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) \right)$$

(4.20)

where $F$ is defined as:

$$F = \left( e_{\text{LLBWIP}_1} - e_{\text{LLBWIP}_2} + \gamma \text{div}\left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) \right)$$

(4.21)

where $e_{\text{LLBWIP}_1}$ and $e_{\text{LLBWIP}_2}$ are defined as follows:

$$\begin{cases}
  e_{\text{LLBWIP}_1} = \int_{\Omega} \omega_y P(y \in R_2) \left( \ln \left( \Sigma_2(y) \right) + \frac{(I(y) - b(y)v_2)^2}{\Sigma_2(y)} \right) dy \\
  e_{\text{LLBWIP}_2} = \int_{\Omega} \omega_y P(y \in R_1) \left( \ln \left( \Sigma_1(y) \right) + \frac{(I(y) - b(y)v_1)^2}{\Sigma_1(y)} \right) dy
\end{cases}$$

(4.22)
The parameter $\varepsilon$ can affect the speed function. As can be seen Figure 4.3(b), for the large value of $\varepsilon$, for instance, $\varepsilon = 4$, the weight $\delta_\varepsilon(\emptyset)$ becomes very small and consequently $\delta_\varepsilon(\emptyset)\mathcal{F}$ in (4.20) is trivial while for small values of $\varepsilon$, for example, $\varepsilon = 0.1$, only a very small range of level sets is weighted. The convergence rate of the energy minimization is slow in both cases and therefore, in practice, $\varepsilon$ is set to 1 or 1.5 [52, 89].

### 4.3 Experimental results

<table>
<thead>
<tr>
<th>Approach</th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
<th>Dice FP</th>
<th>Dice FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA[2]</td>
<td>83.8</td>
<td>4.6</td>
<td>13.6</td>
<td>8.3</td>
<td>9.7</td>
</tr>
<tr>
<td>WA[90]</td>
<td>52.4</td>
<td>11.2</td>
<td>34.1</td>
<td>30.5</td>
<td>31.5</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.1</td>
<td>11.7</td>
<td>33.9</td>
<td>12.7</td>
<td>13.3</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>77.6</td>
<td>8.3</td>
<td>24.1</td>
<td>11.8</td>
<td>12.2</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>71.7</td>
<td>6.4</td>
<td>19.4</td>
<td>15.6</td>
<td>16.4</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>72.9</td>
<td>7.8</td>
<td>21.3</td>
<td>15.3</td>
<td>14.7</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td>88.7</td>
<td>4.2</td>
<td>12.8</td>
<td>5.4</td>
<td>7.1</td>
</tr>
<tr>
<td>Three-step[32]</td>
<td>88.4</td>
<td>4.7</td>
<td>13.4</td>
<td>5.3</td>
<td>5.2</td>
</tr>
<tr>
<td>LSBR[82]</td>
<td>83.2</td>
<td>5.8</td>
<td>19.8</td>
<td>11.8</td>
<td>9.1</td>
</tr>
<tr>
<td>LLBWIP</td>
<td><strong>91.6</strong></td>
<td><strong>3.5</strong></td>
<td><strong>12.7</strong></td>
<td><strong>4.7</strong></td>
<td><strong>3.9</strong></td>
</tr>
</tbody>
</table>

*The best results are indicated by bold values.*

The proposed approach is applied to 2D fluorescence microscopy images of cell nuclei from four experiments which have been introduced in Chapter 2, Section 2.4.2. The
performance of the LLBWIP algorithm is evaluated using region-based and contour-based measures. For region-based measure, the Jaccard coefficient [26], Dice FP and Dice FN are used. For contour-based measures, the Hausdorff distance and MAD are used. The mathematical description of the utilized measures has been given in Chapter 2, Section 2.4.3.

Table 4.2. Quantitative results for NIH3T3 cells data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
<th>Dice FP</th>
<th>Dice FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA[2]</td>
<td>53.4</td>
<td>6.2</td>
<td>18.7</td>
<td>28.2</td>
<td>31.7</td>
</tr>
<tr>
<td>WA[90]</td>
<td>45.2</td>
<td>5.9</td>
<td>19.1</td>
<td>35.4</td>
<td>40.5</td>
</tr>
<tr>
<td>OT[91]</td>
<td>47.3</td>
<td>12.6</td>
<td>37.4</td>
<td>38.4</td>
<td>33.5</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>61.2</td>
<td>7.5</td>
<td>22.2</td>
<td>26.1</td>
<td>22.9</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>62.8</td>
<td>6.8</td>
<td>21.5</td>
<td>20.4</td>
<td>23.5</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>61.6</td>
<td>7.8</td>
<td>23.3</td>
<td>25.3</td>
<td>23.7</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td>73.9</td>
<td>4.4</td>
<td>14.2</td>
<td>16.4</td>
<td>13.5</td>
</tr>
<tr>
<td>Three-step[32]</td>
<td>70.8</td>
<td>5.7</td>
<td>16.4</td>
<td>15.5</td>
<td>19.7</td>
</tr>
<tr>
<td>LSB[82]</td>
<td>64.2</td>
<td>7.2</td>
<td>19.8</td>
<td>21.2</td>
<td>20.4</td>
</tr>
<tr>
<td>LLBWIP</td>
<td><strong>75.9</strong></td>
<td><strong>4.1</strong></td>
<td>14.3</td>
<td><strong>12.7</strong></td>
<td><strong>12.2</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

Table 4.1 shows the quantitative results for the different performance measures which are averaged over all images in the U20S cells. The quantitative results for NIH3T3 cells are reported in Table 4.2 while the quantitative results for BBBC005 cells are reported in Table 4.3. Table 4.4 reports the quantitative results for images of
synthetic cells. As a comparison, we also reported the results of the Merging algorithm (MA) [2], the Otsu thresholding (OT) [91], the Watershed algorithm (WA) [90], and several level set-based methods, namely the BLS [55], the region-scalable fitting energy functional (RSFE) [5], the distance regularized level set method (DRLSE) [89], two step and three step approaches [32] and the level set method based on the Bayesian risk (LSBR) [82].

Table 4.3. Quantitative results for BBBC005 data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>BBBC005 cells (17 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaccard</td>
</tr>
<tr>
<td>MA[2]</td>
<td>80.4</td>
</tr>
<tr>
<td>WA[90]</td>
<td>75.3</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.0</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>68.0</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>70.6</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>70.4</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td><strong>85.6</strong></td>
</tr>
<tr>
<td>Three-step[32]</td>
<td>83.2</td>
</tr>
<tr>
<td>LSBR[82]</td>
<td>75.2</td>
</tr>
<tr>
<td>LLBWIP</td>
<td>83.4</td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

As can be seen from Table 4.1, our approach obtains the best results for the Jaccard coefficient, Hausdorff distance, MAD, Dice FP, and Dice FN for the U20S data set. Specifically, our approach yields better results than all the level set based approaches we tested. Note that for the Dice FP value and the Dice FN value, our approach yields significantly smaller values. LLBWIP also obtains the best results for the Jaccard
coefficient, MAD, Dice FP, and Dice FN for the NIH3T3 data set as shown Table 4.2. For the Hausdorff distance, the best result is obtained by Two-step approach, although the result of the proposed approach is very close.

As can be seen from Table 4.3, for the BBBC005 cells, LLBWIP yields significantly better results for the Hausdorff distance and Dice FN. In particular, we obtain a significantly smaller Dice FN value. For the Jaccard coefficient and Dice FP, the best results are obtained by Two-step approach and MA, respectively, while the results of LLBWIP are close behind. For the synthetic cell images results shown in Table 4.4, the best result for the Dice FN value is obtained by LSBR, but our approach is only slightly poorer. However, LLBWIP obtains significantly better results for the Jaccard coefficient, MAD, Hausdorff distance and Dice FP.

**Table 4.4. Quantitative results for synthetic data set for the different segmentation approaches.**

<table>
<thead>
<tr>
<th>Approach</th>
<th>Synthetic cells (20 images)</th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
<th>Dice FP</th>
<th>Dice FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA[2]</td>
<td></td>
<td>65.2</td>
<td>1.6</td>
<td>5.1</td>
<td>18.6</td>
<td>24.5</td>
</tr>
<tr>
<td>WA[90]</td>
<td></td>
<td>64.5</td>
<td>1.2</td>
<td>5.6</td>
<td>22.3</td>
<td>21.8</td>
</tr>
<tr>
<td>OT[91]</td>
<td></td>
<td>63.2</td>
<td>1.7</td>
<td>7.4</td>
<td>22.7</td>
<td>24.2</td>
</tr>
<tr>
<td>BLS[55]</td>
<td></td>
<td>72.4</td>
<td>0.9</td>
<td>4.8</td>
<td>15.4</td>
<td>18.5</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td></td>
<td>69.5</td>
<td>1.5</td>
<td>7.1</td>
<td>20.4</td>
<td>15.5</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td></td>
<td>62.0</td>
<td>1.3</td>
<td>6.7</td>
<td>22.9</td>
<td>23.0</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td></td>
<td>76.7</td>
<td>1.2</td>
<td>4.8</td>
<td>12.7</td>
<td>13.4</td>
</tr>
<tr>
<td>Three-step[32]</td>
<td></td>
<td>73.1</td>
<td>1.4</td>
<td>5.3</td>
<td>13.2</td>
<td>16.7</td>
</tr>
<tr>
<td>LSBR[82]</td>
<td></td>
<td>74.6</td>
<td>1.4</td>
<td>4.1</td>
<td>18.1</td>
<td>10.9</td>
</tr>
<tr>
<td>LLBWIP</td>
<td></td>
<td><strong>83.3</strong></td>
<td><strong>0.8</strong></td>
<td><strong>3.7</strong></td>
<td><strong>10.1</strong></td>
<td>11.8</td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.
For each data set, the ranking of the algorithms according to the values obtained by Jaccard, Hausdorff distance, and MAD, is depicted in Table 4.5. For each metric, the differences between the methods are only considered statistically significant if the 95% confidence intervals of the estimate of the true means do not overlap. To acquire the ranking for each data set 135 t-tests are performed (45 possible pairings of the methods for each of the three metrics). Bonferroni correction of the confidence intervals is used to adjust for multiple comparisons. Also, standard deviations of LLBWIP results for different data sets are summarized in Table 4.6.

**Table 4.5. Ranking of segmentation methods according to the Jaccard, the Hausdorff distance, and the MAD metrics.**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Metric</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>U20S</td>
<td>Jaccard</td>
<td>LLBWIP &gt; Two-step &gt; Three-step &gt; MA = LSBR &gt; BLS = OT = DRLSE = RSFE &gt; WA</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>LLBWIP &gt; Two-step &gt; MA &gt; Three-step &gt; LSBR &gt; RSFE &gt; DRLSE = BLS &gt; WA = OT</td>
</tr>
<tr>
<td></td>
<td>Hausdorff</td>
<td>LLBWIP = Two-step &gt; Three-step = MA &gt; RSFE = LSBR = DRLSE &gt; BLS &gt; WA = OT</td>
</tr>
<tr>
<td>NIH3T3</td>
<td>Jaccard</td>
<td>LLBWIP &gt; Two-step &gt; Three-step &gt; LSBR &gt; RSFE = DRLSE = BLS &gt; MA &gt; WA = OT</td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>LLBWIP = Two-step &gt; Three-step &gt; WA = MA = RSFE &gt; LSBR = BLS = DRLSE &gt; OT</td>
</tr>
<tr>
<td></td>
<td>Hausdorff</td>
<td>Two-step = LLBWIP &gt; Three-step = MA = WA &gt; LSBR = RSFE = BLS = DRLSE &gt; OT</td>
</tr>
<tr>
<td></td>
<td>Jaccard</td>
<td>Two-step &gt; LLBWIP = Three-step &gt; MA &gt; OT = WA = LSBR &gt; RSFE = DRLSE = BLS</td>
</tr>
<tr>
<td>BBBC005</td>
<td>MAD</td>
<td>LLBWIP = WA = Two-step = Three-step = MA &gt; RSFE &gt; LSBR &gt; BLS &gt; OT &gt; DRLSE</td>
</tr>
<tr>
<td></td>
<td>Hausdorff</td>
<td>LLBWIP = Two-step &gt; Three-step &gt; BLS = LSBR &gt; DRLSE &gt; WA = OT = MA &gt; RSFE</td>
</tr>
<tr>
<td></td>
<td>Jaccard</td>
<td>LLBWIP &gt; Two-step &gt; Three-step &gt; LSBR &gt; BLS &gt; RSFE &gt; MA = WA = OT = DRLSE</td>
</tr>
<tr>
<td>Synthetic cells</td>
<td>MAD</td>
<td>LLBWIP = BLS &gt; Two-step = WA &gt; LSBR = DRLSE = Three-step = RSFE &gt; RSFE = OT</td>
</tr>
</tbody>
</table>
The performance of our algorithm increases with the increasing size of the $q \times q$ neighborhood. For the larger value of $q$ ($q=5$), our approach yields better results, while the time cost also increased significantly. Taking both the segmentation performance and the computation time into consideration, we suggest $3 \leq q \leq 4$. It is important to note that all the level set based approaches use the standard level set scheme with gradient descent optimization and the convergence criterion is defined as the MAD between the contours of segmented objects in two consecutive iterations. Based on the performed analysis in [92], for a given $n \times n$ image, the computational complexity of a level set model is $O(n^2)$ for each evolving level set. LLBWIP uses information of the image patches which would increase the computational cost. However, since $q \leq n$ and $q$ is fixed in LLBWIP, the computational complexity of LLBWIP can be approximated by $O(n^2)$.

### Table 4.6. Standard deviations of LLBWIP results.

<table>
<thead>
<tr>
<th></th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
</tr>
</thead>
<tbody>
<tr>
<td>U2OS cells</td>
<td>6.9</td>
<td>0.6</td>
<td>3.3</td>
</tr>
<tr>
<td>NIH3T3 cells</td>
<td>7.4</td>
<td>1.1</td>
<td>4.0</td>
</tr>
<tr>
<td>BBBC005 cells</td>
<td>5.4</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Synthetic cells</td>
<td>5.7</td>
<td>0.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>

### 4.4 Summary

A novel algorithm for cell nuclei image segmentation based on level set formulation, called LLBWIP, has been introduced. In LLBWIP, a novel local energy functional
based on the Bayesian classification risk for an image patch has been proposed. A weighting scheme is used to enable the pixels in each image patch to have anisotropic weights in the local energy functional. The final energy functional is then obtained by integrating the local energy over the entire image domain. Experiments demonstrated that the proposed approaches are robust and capable of producing significantly more accurate segmentation results than many state-of-the-art approaches we compared with. However, similar to the LBF model and LGDFE, the proposed approach assumes that the local image data follows a Gaussian distribution. In the next chapter, we propose an approach based on the non-Gaussianity assumption.
In this chapter, we show that the data terms of AWCE and LBF models can be rewritten based on k-means clustering. ACWE and LBF models assume that image data are globally and locally Gaussian, respectively. In practice, the image data may not be Gaussian. To perform cell image segmentation under non-Gaussianity, we propose a new segmentation method via correntropy-based k-means clustering in a variational level set formulation. The performance of the proposed method is evaluated using a large number of fluorescence microscopy images. A quantitative comparison is also performed with several state-of-the-art segmentation algorithms. The proposed approach is applied to 2D fluorescence microscopy images of cell nuclei from U20S data set and BBBC005 data set which have been introduced in Chapter 2, Section 2.4.2. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 7%, and 5% better than the state-of-the-art methods we compared with for U20S cells and BBBC005 cells, respectively.
5.1 Introduction

The mean square error (MSE) is the most widely used dissimilarity measure for quantifying how different the intensities of two pixels are. Segmentation algorithms using this dissimilarity measure rely heavily on the Gaussianity and linearity assumptions. Correntropy was introduced by Santamaria et al. [93] who applied it for blind deconvolution. Subsequent work has used correntropy for blind source separation [94] and as a measure for determining nonlinear dynamics [95]. A correntropy-based minimum average correlation energy (MACE) filter is given in [96]. Further discussions on correntropy can be found in [97] and [98].

In this chapter, a level set model based on the local spatial information and correntropy-based k-means clustering (LLCK) is proposed to segment the fluorescence microscopy images. Compared to previous approaches, the main advantages of our segmentation method can be highlighted as follows. First, due to the correntropy criterion, the LLCK segmentation algorithm is robust to outliers. Second, since the spatial relationship of the pixels in a local neighborhood can be utilized as an important characteristic that improves the performance of level set segmentation methods, the LLCK segmentation algorithm can efficiently segment the images with intensity inhomogeneity by employing the local image information.

The remainder of this chapter is organized as follows. Section 2 introduces the proposed approach. In Section 3, experimental results are presented and analyzed using images of different cell types. The results of the proposed methods are also compared with several existing approaches. Finally, Section 4 summarizes the chapter.
5.2 LLCK

K-means clustering can be considered as one of most utilized clustering methods and is simple and computationally fast. For image segmentation, k-means clustering segments an image into regions where each region belongs to one of the k clusters [99]. The algorithm consists of two steps. In the first step, the centroids are calculated and in the second step, each point is taken to belong to the cluster with the nearest centroid. K-means clustering algorithm, which utilizes the Euclidean distance to measure the distance of the nearest centroid, is an iterative algorithm which minimizes the sum of distances from each pixel to its cluster centroid, over all clusters. The continuous form of the k-means clustering for segmenting an image into cell and background can be written as follows:

$$E_{k\text{-means}}(I, v, \theta(x, i)) = \sum_{i=1}^{2} \int_{\Omega_i} |I(x) - v_i|^2 \theta(x, i) \, dx$$  \hspace{1cm} (5.1)

where $v_i$ denotes the center of each cluster and $\theta(x, i)$ denotes the membership function which satisfies the following two conditions: $\theta(x, i) \in \{0, 1\}$ and $\sum_{i=1}^{2} \theta(x, i) = 1$. From Chapter 3, we know that the data term of ACWE can be written as follows:

$$\mathcal{D}_{ACWE}(I, v, \theta(x)) = \sum_{i=1}^{2} \int_{\Omega} |I(x) - v_i|^2 \theta(x, i) \, dx$$  \hspace{1cm} (5.2)

The energy functional (5.1) provides the data term similar to $\mathcal{D}_{ACWE}(I, v, \theta(x))$ [52].

For $N_x$, a circular neighbourhood of point $x$, the data term of the LBF model $\mathcal{D}_{LBF_x}(3.25)$, using the continuous form of the k-means clustering for segmenting the $N_x$ into cell and background, can be rewritten as follows:
\[
\mathcal{D}_{\text{LBF}_x} = \sum_{i=1}^{2} \int_{R_i} \theta(y, i) |l(y) - v_i(x)|^2 dy
\]

(5.3)

where \(v_i(x)\) denotes the centre of each cluster and \(\theta(y, i)\) denotes the membership function which satisfies the following two conditions: \(\theta(y, i) \in \{0,1\}\) and \(\sum_{i=1}^{2} \theta(y, i) = 1\).

Since (5.3) utilized the MSE criterion to measure the distance between pixels and cluster centers, it is sensitive to noise and outliers. The correntropy criterion [100] can be applied to tackle this difficulty. Therefore, the new data term can be written as:

\[
\mathcal{D}_{\text{LLCK}_x} = -\sum_{i=1}^{2} \int_{R_i} \theta(y, i) \sigma^2 g(|l(y) - v_i(x)|) dy
\]

(5.4)

where \(g(q) = \exp\left(-\frac{q^2}{2\sigma^2}\right)\) is a Gaussian function with kernel width \(\sigma\). Using the correntropy-based k-means clustering, \(\mathcal{D}_{\text{LLCK}_x}\) can adaptively emphasize the samples.
that are close to their corresponding cluster centres. Therefore, the effect of outliers can be reduced.

The intensities $I(y)$ involved in the $D_{LLCK_x}$ are in a local region centred at the point $x$.

The following truncated Gaussian function with a localization property is utilized to control the size of the $N_x$.

$$
\omega(d) = \begin{cases} 
\frac{1}{a} \exp \left( - \frac{|d|^2}{2\mathcal{K}^2} \right) & \text{if } |d| \leq \rho \\
0 & \text{if } |d| > \rho 
\end{cases}
$$

(5.5)

where $\alpha$ denotes a constant such that $\int \omega(d) = 1$ and $\mathcal{K}$ is a scale parameter. Therefore (5.4) can be rewritten as follows:

$$
D_{LLCK} = - \sum_{i=1}^{2} \int_{\Omega_i} \omega(x-y) \theta(y_i) \sigma^2 g(|I(y) - v_i(x)|) dy
$$

(5.6)

Using the Heaviside function, (5.6) can be rewritten as:

$$
D_{LLCK} = - \sum_{i=1}^{2} \int_{\Omega} \omega(x-y) M_i(\emptyset(y)) \sigma^2 g(|I(y) - v_i(x)|) dy
$$

(5.7)

Considering all points $x$ in the image domain $\Omega$, the final energy functional can be written as follows:

$$
E_{LLCK} = - \sum_{i=1}^{2} \int_{\Omega} \int_{\Omega} \omega(x-y) M_i(\emptyset(y)) \sigma^2 g(|I(y) - v_i(x)|) dy dx + \gamma L(\emptyset) + P(\emptyset)
$$

(5.8)

Using the $H_\epsilon(z)$, the energy functional is approximated as follows:

$$
E_{LLCK_\epsilon} = - \sum_{i=1}^{2} \int_{\Omega} \int_{\Omega} \omega(x-y) M_i(\emptyset(y)) \sigma^2 g(|I(y) - v_i(x)|) dy dx + \gamma L_\epsilon(\emptyset)
$$
In this chapter, \( \gamma \) is set to \( 255 \times 255 \times 10^{-3} \) and \( \varepsilon \) is set to 1 [52, 89]. For more details about \( \gamma \) and \( \varepsilon \) please see Chapter 4, Section 4.3.

For a fixed level set function \( \varnothing \), the update equations for the \( \nu_i(x) \) \( i \in \{1,2\} \) can be written as follows:

\[
\nu_i(x) = \frac{\omega(x - y) M_{1e}(\varnothing(x)) \sigma^2 g(||y - \nu_i(x)||)}{\omega(x - y) M_{1e}(\varnothing(x)) \sigma^2 g(||y - \nu_i(x)||)}
\]  

(5.10)

Minimization of the energy functional \( E_{\text{LLCK}_{\varepsilon}} \) with respect to \( \varnothing \) is achieved by solving the following gradient descent flow equation:

\[
\frac{\partial \varnothing}{\partial t} = \delta_{\varepsilon}(\varnothing) \mathcal{F} + \left( \nabla^2 \varnothing - \text{div} \left( \frac{\nabla \varnothing}{|\nabla \varnothing|} \right) \right)
\]  

(5.11)

where \( \mathcal{F} \) is defined as:

\[
\mathcal{F} = \left( \mathcal{F}_1 - \mathcal{F}_2 + \gamma \text{div} \left( \frac{\nabla \varnothing}{|\nabla \varnothing|} \right) \right)
\]  

(5.12)

where \( \mathcal{F}_1 \) and \( \mathcal{F}_2 \) are defined as:

\[
\begin{align*}
\mathcal{F}_1 &= \int_{\Omega} \omega(x - y) M_{1e}(\varnothing(y)) \sigma^2 g(||y - \nu_1(x)||) \, dy \\
\mathcal{F}_2 &= \int_{\Omega} \omega(x - y) M_{2e}(\varnothing(y)) \sigma^2 g(||y - \nu_2(x)||) \, dy
\end{align*}
\]  

(5.13)
5.2 Experimental results

The effectiveness of the proposed approach is illustrated using 2D fluorescence microscopy images of cell nuclei from [23, 24] which have been introduced in Chapter 2, Section 2.4.2. We compare the proposed method with OT [91], WA [90], and three level set-based methods, namely BLS [55], RSFE [5], and DRLSE [89]. The quantitative results for U20S cells are reported in Table 5.1. Table 5.2 reports the quantitative results for BBBC005 cells. All performance measures are averaged over all images in the U20S and the BBBC005 datasets. The mathematical description of the utilized measures has been proposed in Chapter 2, Section 2.4.3.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
<th>Dice FP</th>
<th>Dice FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA[90]</td>
<td>52.4</td>
<td>11.2</td>
<td>34.1</td>
<td>30.5</td>
<td>31.5</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.1</td>
<td>11.7</td>
<td>33.9</td>
<td>12.7</td>
<td>13.3</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>77.6</td>
<td>8.3</td>
<td>24.1</td>
<td>11.8</td>
<td><strong>12.2</strong></td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>71.7</td>
<td>6.4</td>
<td>19.4</td>
<td>15.6</td>
<td>16.4</td>
</tr>
<tr>
<td>ACWE[52]</td>
<td>69.3</td>
<td>8.6</td>
<td>27.1</td>
<td>21.2</td>
<td>15.1</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>72.9</td>
<td>7.8</td>
<td>21.3</td>
<td>15.3</td>
<td>14.7</td>
</tr>
<tr>
<td>LLCK</td>
<td><strong>83.1</strong></td>
<td><strong>6.2</strong></td>
<td><strong>18.8</strong></td>
<td><strong>8.3</strong></td>
<td>12.6</td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

As can be seen from Table 5.1, our approach yields the best results for the Jaccard coefficient, MAD, Hausdorff distance, and Dice FP for the U20S data set. Furthermore, our approach yields significantly better results than the level set based approaches. Note that for the Dice FP value, LLCK obtains significantly smaller values. For the Dice FN, the best result is obtained by the BLS approach while the result of the proposed
approach is comparable to the best results of the other approaches. It is apparent from Table 5.2 that the LLCK obtains the best results for the Jaccard coefficient, Dice FP, and Dice FN. The BLS obtains the best result for the Hausdorff distance. For the MAD, the best result is obtained by WA approach, although the result of the proposed approach is very close. Also, standard deviations of LLCK results for different data sets are summarized in Table 5.3.

Table 5.2. Quantitative results for BBBC005 data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>BBBC005 data set (17 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaccard</td>
</tr>
<tr>
<td>WA[90]</td>
<td>75.3</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.0</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>68.0</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>70.6</td>
</tr>
<tr>
<td>ACWE[52]</td>
<td>65.7</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>70.4</td>
</tr>
<tr>
<td>LLCK</td>
<td><strong>79.8</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

Table 5.3. Standard deviations of LLCK results.

<table>
<thead>
<tr>
<th></th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
</tr>
</thead>
<tbody>
<tr>
<td>U20S cells</td>
<td>6.4</td>
<td>0.8</td>
<td>5.9</td>
</tr>
<tr>
<td>BBBC005 cells</td>
<td>4.9</td>
<td>0.3</td>
<td>0.6</td>
</tr>
</tbody>
</table>

5.3 Summary

In this chapter, a novel level-set based algorithm for cell nuclei image segmentation has been developed. We utilized correntropy-based k-means algorithm, which can reduce
the effects of noise and outliers, to define the energy functional around the neighborhood of a point. Due to the correntropy criterion, the clustering algorithm adaptively increases the weights of samples that are close to their clusters and decreases the weights of samples that are away from their clusters. As a result, compared with LBF model, our proposed segmentation algorithm is robust to outliers. Our results for both U20S data set and BBBC005 data set show that the LLCK can produce accurate segmentation results. As it has been explained, LLCK utilizes the Gaussian kernel which satisfies Mercer’s Theorem [101]. Recent advances in kernel methods and their applications have highlighted the need to consider multiple kernels or composite kernels instead of a single fixed kernel. Therefore, next chapter proposes a local implicit model based on the multiple kernels mapping.
Chapter 6

A Multi-Kernel Local Implicit Model for Segmentation

In this chapter, a multiple kernel local segmentation algorithm is introduced as a framework for fluorescence microscopy cell image segmentation. In this framework, a new data term in a variational level set formulation based on the piecewise constant model and multiple kernels mapping is proposed where a linear combination of multiple kernels is utilized to implicitly map the original local image data into data of a higher dimension. We evaluate the performance of the proposed method using a large number of fluorescence microscopy images. A quantitative comparison is also performed with some state-of-the-art segmentation approaches. The proposed approach is applied to 2D fluorescence microscopy images of cell nuclei from U20S data set and NIH3T3 data set which have been introduced in Chapter 2, Section 2.4.2. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 6%, and 7% better than the state-of-the-art methods we compared with for U20S data set and NIH3T3 data set, respectively.
6.1 Introduction

The goal of this chapter is motivated by the fact that in images segmentation, pixels in a local neighborhood possess nearly the same intensity. Therefore, the spatial relationship of the pixels in a local neighborhood can be utilized as an important characteristic that improves the performance of intensity-based segmentation methods. This chapter introduces a novel implicit model called “multiple kernel local level set segmentation method (MKLLS)” based on the linear composite of multiple kernels for cell nucleus segmentation in fluorescence microscopy images. In this approach, first, we define a data term around a neighborhood of a point. In LBF model, the local image data belonging to the two regions, are assumed to be linearly separable, but this is not always the case in practice. To address the case that the local image data is not linearly separable, the image data can be implicitly transformed via a nonlinear mapping into data of a higher dimension so that local image data of the two regions becomes linearly separable (see Figure 6.1) [102]. Therefore, to form the data term, the local image intensities data are mapped implicitly into data of a higher dimension using the linear combination of multiple kernels. Finally, the data term is integrated over the entire image domain to obtain the overall energy functional.

The remainder of the chapter is organized as follows. The proposed approach is discussed in section 2. In section 3, experimental results are presented and analyzed using different images. The summary is reported in section 4.

6.2 MKLLS

Using the ACWE data term, the data term for segmenting the $N_x$ can be written as follows:
\[ \mathcal{D}_{MKLLS_x}(I, \nu_i) = \sum_{i=1}^{2} \int_{R_i} ||I(y) - \nu_i||^2 dy \]

\[ \mathcal{D}_{MKLLS_x}(I, \nu_i) = \sum_{i=1}^{2} \int_{R_i} ||I(y) - \nu_i||^2 dy \]

(6.1)

Figure 6.1. Illustration of nonlinear 2-D data separation with mapping: The data is non-linearly separable in the data space. Mapping the data to a feature (kernel) space and, then, separating it in the induced space with linear methods is possible. For the purpose of display, the feature space in this example is of the same dimension as the original data space. In general, however, the feature space is of higher dimension.

Since the image data in \( N_x \) can be non-linearly separable between the object and background regions, the Mercer kernel function can be used to transform the image data implicitly. Therefore, \( \mathcal{D}_{MKLLS_x} \) can be reformulated as follows:

\[ \mathcal{D}_{MKLLS_x}(I, \nu_i) = \sum_{i=1}^{2} \int_{R_i} ||\Phi(I(y)) - \nu_i||^2 dy \]

where \( \Phi(.) \) is a nonlinear mapping for image data in \( N_x \) to a higher dimensional space. Since \( \Phi(.) \) is not known, using the Mercer kernel function, \( K(Y,Z) \), it can be written as:
\[ \langle \Phi(Y), \Phi(Z) \rangle = K(Y, Z) \] (6.3)

We aim to generalize \( \mathcal{D}_{MKLLS_x} \) using the following properties of Mercer kernels \([74]\).

**Theorem** \([74]\):

Assume that \( K_1 \) and \( K_2 \) are defined over \( \Xi \times \Xi, \Xi \subseteq \mathbb{R}^p \) as kernel functions and \( K_3 \) is also defined as a kernel function over \( \mathbb{R}^p \times \mathbb{R}^p \):

- \( K(Y, Z) = K_1(Y, Z) + K_2(Y, Z) \) is a kernel.
- If \( \alpha \) be a positive constant, \( K(Y, Z) = \alpha K_1(Y, Z) \) is a kernel.
- \( K(Y, Z) = K_1(Y, Z)K_2(Y, Z) \) is a kernel.
- Let function \( \Psi: \Xi \rightarrow \mathbb{R}^p \), \( K(Y, Z) = K_3(\Psi(Y), \Psi(Z)) \) is a kernel.

Based on the properties of Mercer kernels, the combination of the Mercer kernels, \( K_{\text{com}} \), is introduced and the nonlinear mapping, \( \Phi(.) \), can be replaced with \( \Phi_{\text{com}}(.) \).

\[ \langle \Phi_{\text{com}}(Y), \Phi_{\text{com}}(Z) \rangle = K_{\text{com}}(Y, Z) \] (6.4)

Now the data term can be written as:

\[ \mathcal{D}_{MKLLS_x}(I, \nu_j) = \sum_{i=1}^{2} \int_{R_i} \| \Phi_{\text{com}}(I(y)) - \nu_i \|^2 \, dy \] (6.5)

In this study, the linear form is selected as the combination strategy. Therefore, \( K_{\text{com}} \) is replaced with \( K_{\text{icom}} \), which is defined as follows:

\[ K_{\text{icom}}(Y, Z) = \sum_{j=1}^{L} w_{ij} K_j(Y, Z) \]

s.t \[ \sum_{j=1}^{L} w_{ij} = 1 \] (6.6)
Using $K_{\text{com}}$, $\Phi_{\text{com}}(I(y))$ can be replaced with $\Phi_{\text{com}}(I(y))$.

Since different pixels should have different weights reflecting their influence, the truncated Gaussian function introduced in Chapter 5 is utilized. Therefore, the data term can be rewritten as follows:

$$
\mathcal{D}_{\text{MKLLS}_x}(I, v_i) = \sum_{i=1}^{\Omega_i} \int \omega(x - y) \|\Phi_{\text{com}}(I(y)) - v_i\|^2 \, dy
$$

(6.7)

Using the Lipchitz function $\varnothing(x)$ and the Heaviside function $H$, $\mathcal{D}_{\text{MKLLS}_x}$ can be rewritten as:

$$
\mathcal{D}_{\text{MKLLS}_x}(I, v_i, \varnothing(x)) = 
\sum_{i=1}^{\Omega_i} \int \omega(x - y) \|\Phi_{\text{com}}(I(y)) - v_i\|^2 M_i(\varnothing(x)) \, dy
$$

(6.8)

Now the final energy functional of MKLLS can be written as follows:

$$
E_{\text{MKLLS}} = \mathcal{D}_{\text{MKLLS}} + \gamma \mathcal{L}(\varnothing) + \mathcal{P}(\varnothing)
$$

(6.9)

where $\mathcal{D}_{\text{MKLLS}} = \int \Omega \mathcal{D}_{\text{MKLLS}_x} \, dx$. Using the $H_{\varepsilon}(z)$, the energy functional (6.9) can be approximated as follows:

$$
E_{\text{MKLLS}} = \mathcal{D}_{\text{MKLLS}_\varepsilon} + \gamma \mathcal{L}_{\varepsilon}(\varnothing) + \mathcal{P}(\varnothing)
$$

(6.10)

where $\mathcal{D}_{\text{MKLLS}_\varepsilon}$ approximates $\mathcal{D}$ and $\mathcal{L}_{\varepsilon}(\varnothing)$ denotes an estimation of $\mathcal{L}(\varnothing)$. In this chapter, $\gamma$ is set to $255 \times 255 \times 10^{-3}$ and $\varepsilon$ is set to 1 [52, 89]. For more details about $\gamma$ and $\varepsilon$ please see Chapter 4, section 4.3.
Using calculus of variations, for a fixed partition of the image domain, the $\nu_i$ can be calculated as follows:

$$
\nu_i = \frac{\int_\Omega \int_\Omega \omega(x - y) \Phi_{\text{com}}(I(x)) M_{i\epsilon}(\emptyset(x)) dy dx}{\int_\Omega \int_\Omega \omega(x - y) M_{i\epsilon}(\emptyset(x)) dy dx}
$$

(6.11)

By introducing the Lagrange term of the constraint of weights $w_j$ ($j = 1, \ldots, l$) into $D_{\text{MKLLS}_e}$, the following formulation can be obtained:

$$
\mathcal{Q} = \int_\Omega \left( \sum_{i=1}^2 \omega(x - y) \| \Phi_{\text{com}}(I(y)) - \nu_i \|^2 M_{i\epsilon}(\emptyset(y)) \right) dy + \eta \left( 1 - \sum_{j=1}^1 w_j \right)
$$

(6.12)

where $\eta$ denotes Lagrange multiplier. Taking the derivative of $\mathcal{Q}$ over $w_j$, the updating rule for $w_j$ can be obtained as follows:

$$
\frac{\partial \mathcal{Q}}{\partial w_j} = 0 (j = 1, \ldots, l) \Rightarrow w_j = \frac{1}{\sum_{j=1}^l \left( \frac{\mathcal{Q}_j}{\mathcal{Q}_z} \right)^{\frac{1}{b-1}}}
$$

(6.13)

where $\mathcal{Q}_z$ is defined as follows:

$$
\mathcal{Q}_z = \int_\Omega \left( \sum_{i=1}^2 \int_\Omega \omega(x - y) \| \Phi_z(I(x)) - \nu_i \|^2 dy \right) M_{i\epsilon}(\emptyset(x)) dx
$$

(6.14)

where $< \Phi_z(Y), \Phi_z(Z) > = K_z(Y,Z)$ for $z = 1, \ldots, l$. For the specific pixel $x \in \Omega$ and set of variables of integration $\{ a, c, x, y \in \Omega \}$, $\| \Phi_z(I(x)) - \nu_i \|^2$ can be calculated as follows:
\[
\|\Phi_z(I(x)) - \nu_1\|^2 = K_z(I(x), I(x)) - 2 \frac{\int_\Omega \int_\Omega \omega(x-y)M_{\text{le}}(\emptyset(x))K_z(I(x), I(x))dydx}{\int_\Omega \int_\Omega \omega(x-y)M_{\text{le}}(\emptyset(x))dydx} \\
+ \frac{\int_\Omega \int_\Omega \int_\Omega \int_\Omega \beta \tau \alpha d\mathbf{a} d\mathbf{c} dydx}{\int_\Omega \int_\Omega \int_\Omega \int_\Omega \beta \alpha d\mathbf{a} d\mathbf{c} dydx}
\] (6.15)

where \(\tau, \alpha, \beta\) are defined as follows:

\[
\begin{align*}
\tau &= K_z(I(\mathbf{a}), I(\mathbf{x})) \\
\alpha &= \omega(\mathbf{a} - \mathbf{c})M_{\text{le}}(\emptyset(\mathbf{c})) \\
\beta &= \omega(\mathbf{x} - \mathbf{y})M_{\text{le}}(\emptyset(\mathbf{y}))
\end{align*}
\] (6.16)

Minimization of the energy functional \(E_{\text{MKLLS}}\) is obtained by solving the following gradient descent flow equation:

\[
\frac{\partial \emptyset}{\partial t} = \delta_{\epsilon}(\emptyset) F + \left( \nabla^2 \emptyset - \text{div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) \right)
\] (6.17)

where \(F = \left[ e_{\text{MKLLS}_1}, e_{\text{MKLLS}_2} \right] + \text{div} \left( \frac{\emptyset}{|\nabla \emptyset|} \right)\), \(e_{\text{MKLLS}_1}\) and \(e_{\text{MKLLS}_2}\) are defined as:

\[
\begin{align*}
e_{\text{MKLLS}_1} &= \int_\Omega \omega(x-y)\|\Phi_{\text{le}}(I(x)) - \nu_1\|^2 dx \\
e_{\text{MKLLS}_2} &= \int_\Omega \omega(x-y)\|\Phi_{\text{le}}(I(x)) - \nu_2\|^2 dx
\end{align*}
\] (6.18)

### 6.3 Experimental results

To illustrate the effectiveness of the MKLLS, we apply MKLLS to 2D fluorescence microscopy images of cell nuclei from [23], specifically, the U20S data set (see Figure 2.4 (a)) and the NIH3T3 data set (see Figure 2.4 (b)) which have been introduced in Chapter 2, Section 2.4.2. Since the NIH3T3 data set suffered from intensity...
inhomogeneity, compared with the images in the U20S data set, automatic analysis of the NIH3T3 data set is more difficult. The proposed method is compared with OT [91], WA [90], and four level set-based methods, BLS [55], RSFE [5], and DRLSE [89], and kernel based level set approach (KLS) [102]. Table 6.1 reports the quantitative results for the different performance measures which are averaged over all images in the U20S cells. The quantitative results for NIH3T3 cells are reported in Table 6.2. The mathematical description of the utilized measures has been proposed in Chapter 2, Section 2.4.3.

Table 6.1. Quantitative results for U20S data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>U20S cells (48 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaccard</td>
</tr>
<tr>
<td>WA[90]</td>
<td>52.4</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.1</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>77.6</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>71.7</td>
</tr>
<tr>
<td>DRLSE [89]</td>
<td>72.9</td>
</tr>
<tr>
<td>ACWE[52]</td>
<td>69.3</td>
</tr>
<tr>
<td>KLS[102]</td>
<td>78.5</td>
</tr>
<tr>
<td>MKLLS</td>
<td><strong>82.2</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

In this chapter Gaussian kernel function and hyperbolic tangent kernel function are used as Mercer kernels. The standard deviation of the truncated Gaussian function is $\sigma=4$ and the neighborhood radius of the kernel function is $\rho=8$.

As can be seen from Table 6.1, the MKLLS obtains the best results for all measures for U20S data set. Also, compared with the level set based approaches, the MKLLS
obtains better results. It is apparent from Table 6.2 that the MKLLS obtains the best results for the Jaccard coefficient, MAD, Dice FP, and Dice FN. For Hausdorff distance value, the best result is obtained by WA while the result of the MKLLS is comparable to the results of other approaches. The standard deviations of LLCK results for different data sets are summarized in Table 6.3.

Table 6.2. Quantitative results for NIH3T3 data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>NIH3T3 cells (49 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaccard</td>
</tr>
<tr>
<td>WA[90]</td>
<td>45.2</td>
</tr>
<tr>
<td>OT[91]</td>
<td>47.3</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>61.2</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>62.8</td>
</tr>
<tr>
<td>DRLSE [89]</td>
<td>61.6</td>
</tr>
<tr>
<td>ACWE[52]</td>
<td>56.7</td>
</tr>
<tr>
<td>KLS[102]</td>
<td>64.1</td>
</tr>
<tr>
<td>MKLLS</td>
<td><strong>68.3</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

Table 6.3. Standard deviations of MKLLS results.

<table>
<thead>
<tr>
<th></th>
<th>MAD</th>
<th>Hausdorff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaccard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U20S cells</td>
<td>5.9</td>
<td>1.1</td>
</tr>
<tr>
<td>NIH3T3 cells</td>
<td>7.6</td>
<td>1.3</td>
</tr>
</tbody>
</table>
6.4 Summary

In this chapter, a multiple kernel local level set segmentation method based on the linear composition of multiple kernels has been developed. The proposed method utilizes a truncated Gaussian function to incorporate spatial constraints into the local data term and uses a linear composite of multiple kernels to map implicitly local image into data of a higher dimension. Our results for both the U20S data set and the NIH3T3 dataset show that the proposed approach can produce accurate segmentation results.
Chapter 7

Segmentation with Multiple-Kernel Induced Data Term

This chapter investigates an alternative approach for image segmentation using spatial information and multiple kernels mapping. Compared with MKLLS, this chapter utilizes multiple kernels to reflect the fact that practical learning problems often involve data from multiple heterogeneous or homogeneous sources. Therefore, to define the data term, we utilize the intensity of a pixel obtained from the original image and the local spatial information obtained from the processed image. The performance of the proposed method is evaluated using 2D fluorescence microscopy images of cell nuclei from four experiments which have been introduced in Chapter 2, Section 2.4.2 and compared with some state-of-the-art segmentation approaches. The proposed algorithm obtains the highest Jaccard coefficient value for all four experiments. The Jaccard coefficient values are approximately 2%, 5%, 2%, and 4% better than the state-of-the-art methods we compared with for U20S cells, NIH3T3 cells, BBBC005 data set, and Synthetic cells, respectively.
7.1 Introduction

Kernel methods utilizing multiple kernels have more flexibility on kernel selections and can work with data from multiple heterogeneous or homogeneous sources [103-113]. Specifically, in cell nuclei image segmentation, the required information, which is the properties of image pixels, could be obtained from different sources. For example, an image pixel can be described by its intensity as well as its texture property obtained from some wavelet filtering of the image [114]. Multiple-kernel methods provide us an excellent tool to fuse information from different sources [108]. It is noted that, in this chapter, we use the term “multiple kernel” in a different sense than the one used in Chapter 6. In Chapter 6, multiple kernels work with the same data and are used to derive a composite kernel, whereas here, multiple kernels are used for the purpose of information fusion.

The contribution of this chapter is as follows. A novel segmentation method named “level set segmentation with multiple kernels induced data term” (LSMKD) is proposed as a framework for cell image segmentation. The proposed approach provides us a new flexible vehicle to fuse different pixel information in image segmentation. That is, different pixel information transformed by different kernels is combined to produce a new feature.

The remainder of this chapter is organized as follows. Section 2 introduces the proposed approaches. In Section 3, experimental results are presented and analyzed using different cell types. The results of the proposed methods are also compared with previous approaches. This chapter is summarized in Section 4.
7.2 LSMKD

Using the Bayesian framework [54, 115, 116], image segmentation aims to find partitions \( \{ \Omega_i \}_{i=1}^2 \) which maximizes the posterior probability over region partitions of \( \Omega \).

\[
\{ \Omega_i \}_{i=1}^2 = \arg \max_{\Omega_i \in \Omega} P(\{ \Omega_i \}_{i=1}^2 | I) = \arg \max_{\Omega_i \in \Omega} P(I|\{ \Omega_i \}_{i=1}^2) P(\{ \Omega_i \}_{i=1}^2)
\] (7.1)

By taking the logarithm (7.1) can be rewritten as:

\[
\{ \Omega_i \}_{i=1}^2 = \arg \min_{\Omega_i \in \Omega} F(\{ \Omega_i \}_{i=1}^2)
\] (7.2)

where \( F \) is defined as follows:

\[
F(\{ \Omega_i \}_{i=1}^2) = \sum_{i=1}^2 \int_{\Omega_i} -\log P(I(x)|\Omega_i) dx - \log P(\{ \Omega_i \}_{i=1}^2)
\] (7.3)

As it has been discussed in Chapter 3, the first term, named the data term, is to measure the conformity within each region to the distribution \( P(I(x)|\Omega_i) \). The Gaussian distribution is commonly used to model the distribution \( P(I(x)|\Omega_i) \). The piecewise constant segmentation model [52-54] is a specific case of the Gaussian distribution. In the piecewise constant segmentation model, the data term is formulated as follows:

\[
D = \sum_{i=1}^2 \int_{\Omega_i} f_i(x) dx
\] (7.4)

where \( f_i(x) = -\ln P_{v_i}(I(x)) = (I(x) - v_i) \), and \( v_i \) denotes the mean intensity of \( \Omega_i \).

From Chapter 3, we know that for the LCV model, the data term, \( D_{LCV} \), consists of two terms: a global term which uses the \( D_{ACWE} \), and a local term \( D_{LACWE} \). The local
term uses the difference image $I(x) = g_q I(x) - I(x)$, where $g_q$ is an averaging filter with $q \times q$ size. $\mathcal{D}_{LCV}$ can be reformulated as follows:

$$\mathcal{D}_{LCV}(I, I, v_i, d_i) = \mathcal{D}_{ACWE}(I, v_i) + \mathcal{D}_{LACWE}(I, d_i)$$  \hspace{1cm} (7.5)

where $\mathcal{D}_{ACWE}(I, v_i)$ and $\mathcal{D}_{LACWE}(I, d_i)$ can be written as follows:

$$\mathcal{D}_{ACWE}(I, v_i) = \sum_{i=1}^{2} \int_{\Omega_i} f_i(x) \, dx$$  \hspace{1cm} (7.6)

$$\mathcal{D}_{LACWE}(I, d_i) = \sum_{i=1}^{2} \tilde{f}_i(x) \, dx$$  \hspace{1cm} (7.7)

where $\tilde{f}_i(x) = -\ln P_{d_i}(I(x)) = (I(x) - d_i)$, and $d_i$ denotes the intensity averages of the difference image. As can be seen from (7.6) and (7.7) $\mathcal{D}_{ACWE}$ and $\mathcal{D}_{LACWE}$ are formulated based on the piecewise constant model under the Gaussianity assumption. The piecewise constant model assumes that image data is linearly separable between regions. However, this assumption is not generally applicable for fluorescence microscopy images [3, 27, 32, 35, 44, 117-122]. For a nonlinearly separable segmentation problem, the Cover’s theorem [122] indicated that the image data can be transformed via a nonlinear mapping into data of a higher dimension so that piecewise constant model becomes applicable [102].

Let $\Phi(.)$ denote a nonlinear mapping from the given information channel into a higher dimensional space. Based on the Mercer’s theorem [74], any continuous, symmetric, positive semi-definite kernel function can be explained as a dot product in a high-dimensional space. Therefore, it is not required to know explicitly the mapping and a kernel function can be used. Using the Mercer kernel function, $K(Y, Z), \Phi(.)$ can be written as:
\[ < \Phi(Y), \Phi(Z) > = K(Y,Z) \]  
(7.8)

where "." denotes the dot product. Using the kernel function and intensity information which can be directly obtained from the image itself, the data term of LSMKD can be written as follows:

\[
\mathcal{D}_{\text{LSMKD}} = \sum_{i=1}^{2} \int_{\Omega_i} \| \Phi(I(x)) - v_i \|^2 \, dx
\]  
(7.9)

where \( \| \Phi(I(x)) - v_i \|^2 \) is defined as follows:

\[
\| \Phi(I(x)) - v_i \|^2 = < (\Phi(I(x)) - v_i), (\Phi(I(x)) - v_i) > \\
= < \Phi(I(x)), \Phi(I(x)) > - 2 < \Phi(I(x)), v_i > + < v_i, v_i > \\
= K(I(x), I(x)) - 2K(I(x), v_i) + K(v_i, v_i)
\]  
(7.10)

The \( \mathcal{D}_{\text{LSMKD}} \) therefore defines the region parameters in the kernel space. The local term can be introduced which uses the spatial information.

\[
\mathcal{D}_{\text{LSMKD}} = \sum_{i=1}^{2} \int_{\Omega_i} \| \Phi(I(x)) - v_i \|^2 \, dx + \sum_{i=1}^{2} \int_{\Omega_i} \| I(x) - d_i \|^2 \, dx
\]  
(7.11)

(7.11) is an extension of \( \mathcal{D}_{\text{LCV}} \), which maps the intensity data and difference image data into a much higher dimensional Hilbert space by some transform function. Let \( I(x) : \Omega \to \mathbb{R}^v \) denote the input data using different information channels (intensity information, spatial information,...). \( I(x) \) can be defined as \( I(x) = (I(x)_p, I(x)_q) \), where \( p + q = v \). It illustrates that \( I(x)_p \) contains the first \( p \) dimensions of input data and \( I(x)_q \) the remaining \( q \) dimensions. If \( K_p : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R} \) be the kernel over \( \mathbb{R}^p \times \mathbb{R}^p \), then:

\[
K_p(x, y) = K(I(x)_p, I(y)_p)
\]  
(7.12)
Let $\mathbb{R}^p$, then the function $K: \mathbb{R}^v \times \mathbb{R}^v \to \mathbb{R}$ where $K(\hat{I}(x), \hat{I}(x)) = K_p(\hat{I}(x)_p, \hat{I}(x)_p)$, can be considered as a kernel function. Let $\hat{I}(x) = (I(x), \overline{I}(x))$, $\mathcal{D}_{LSMKD}$ can be rewritten as follows:

$$
\mathcal{D}_{LSMKD} = \sum_{i=1}^{2} \int_{\Omega_i} \| \Phi_{lcom}(\hat{I}(x)) - \mathcal{O}_i \|^2 \, dx
$$

(7.12)

where $K_{lcom}(Y, Z) = \langle \Phi_{lcom}(Y), \Phi_{lcom}(Z) \rangle$ and $K_{lcom}(Y, Z)$ is defined as follows:

$$
K_{lcom}(Y, Z) = \sum_{j=1}^{1} w_j^b K_j(Y, Z)
$$

s.t $\sum_{j=1}^{1} w_j = 1$

(7.13)

where $b > 1$. It is noted that (7.12) confines the region parameters to be constructed in the kernel space. Two regions $\Omega_1$ and $\Omega_2$ can be defined by the regions separated by the Lipchitz function $\mathcal{O}$, $\Omega_1 = \{ \mathcal{O}(x) > 0 \}$ and $\Omega_2 = \{ \mathcal{O}(x) \leq 0 \}$. Using the Heaviside function $H$, $\mathcal{D}_{LSMKD}$ is rewritten as follows:

$$
\mathcal{D}_{LSMKD} = \sum_{i=1}^{2} \int_{\Omega_i} \| \Phi_{lcom}(\hat{I}(x)) - \mathcal{O}_i \|^2 M_i(\mathcal{O}(x)) \, dx
$$

(7.14)

The final energy functional can be written as:

$$
E_{LSMKD} = \mathcal{D}_{LSMKD} + \gamma \mathcal{L}(\mathcal{O}) + \mathcal{P}(\mathcal{O})
$$

(7.15)

The Heaviside function $H(z)$ is approximated using a smooth function $H_\varepsilon(z)$ where $\varepsilon$ denotes a positive constant. Therefore, the energy functional (7.15) can be approximated as follows:
\[ E_{\text{LSMKD}_\varepsilon} = \mathcal{D}_{\text{LSMKD}_\varepsilon} + \gamma \mathcal{L}_\varepsilon(\emptyset) + \mathcal{P}(\emptyset) \]  \hspace{1cm} (7.16)

where \( \mathcal{D}_{\text{LSMKD}_\varepsilon} \) approximates \( \mathcal{D} \) and \( \mathcal{L}_\varepsilon(\emptyset) \) denotes an estimation of \( \mathcal{L}(\emptyset) \).

In numerical implementation, for fixed \( \emptyset \), the variables \( \emptyset_i \) is updated as follows:

\[ \emptyset_i = \frac{\int_\Omega \Phi_{\text{com}}(\mathbb{I}(x))M_i(\emptyset(x)) \, dx}{\int_\Omega M_i(\emptyset(x)) \, dx} \]  \hspace{1cm} (7.17)

Minimization of the energy functional \( E_{\text{LSMKD}_\varepsilon} \) with respect to \( \emptyset \) is achieved by solving the following gradient descent flow equation:

\[ \frac{\partial \emptyset}{\partial t} = \delta_\varepsilon(\emptyset) \mathcal{F} + \left( \nabla^2 \emptyset - \text{div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) \right) \]  \hspace{1cm} (7.18)

where \( \mathcal{F} \) is defined as:

\[ \mathcal{F} = \left( e_{\text{LSMKD}_1} - e_{\text{LSMKD}_2} + \gamma \text{div} \left( \frac{\nabla \emptyset}{|\nabla \emptyset|} \right) \right) \]  \hspace{1cm} (7.19)

where \( e_{\text{LSMKD}_1} \) and \( e_{\text{LSMKD}_2} \) are defined as follows:

\[ \begin{aligned}
    e_{\text{LLBWIP}_1} &= \left\| \Phi_{\text{com}}(\mathbb{I}(x)) - \emptyset_1 \right\|^2 \\
    e_{\text{LLBWIP}_2} &= \left\| \Phi_{\text{com}}(\mathbb{I}(x)) - \emptyset_2 \right\|^2
\end{aligned} \]  \hspace{1cm} (7.20)

By introducing the Lagrange term of the constraint of weights \( w_j \) \( (j = 1, \ldots, l) \) into the \( \mathcal{D}_{\text{LSMKD}_\varepsilon} \), the following formulation can be obtained:

\[ \mathcal{D}_{\text{LSMKD}_\varepsilon} = \sum_{i=1}^{2} \int_{\Omega_i} \left\| \Phi_{\text{com}}\left(\mathbb{I}(x)\right) - \emptyset_i \right\|^2 M_i(\emptyset(x)) \, dx \]
\[ + \eta \left( 1 - \sum_{j=1}^{1} w_j \right) \]  \hspace{1cm} (7.21)

where \( \eta \) denotes Lagrange multiplier. Taking the derivative of (7.21) over \( w_j \), the updating rule for \( w_j \) can be obtained as follows:

\[
\frac{\partial D_{\text{LSMKD}_z}}{\partial w_j} = 0 (j = 1, ..., l) \Rightarrow w_j = \frac{1}{\sum_{z=1}^{l} \left( \frac{D_{\text{LSMKD}_z}}{D_{\text{LSMKD}_z}^i} \right)} \hspace{1cm} (7.22)
\]

where \( D_{\text{LSMKD}_z} \) is defined as follows:

\[
D_{\text{LSMKD}_z} = \sum_{i=1}^{2} \int_{\Omega} \| \Phi_z (\mathbb{I}(x)) - \mathbb{O}_i \|^2 M_{ie} (\mathbb{O}(x)) \, dx \quad (z = 1, ..., l) \hspace{1cm} (7.23)
\]

where \( \Phi_z (\mathbb{I}(x)) \) is a transform function defined by \( K_z \) and \( \| \Phi_z (\mathbb{I}(x)) - \mathbb{O}_i \|^2 \) is formulated as follows:

\[
\| \Phi_z (\mathbb{I}(x)) - \mathbb{O}_i \|^2 = K_z (\mathbb{I}(x), \mathbb{I}(x)) - \frac{2 \int_{\Omega} K_z (\mathbb{I}(x), \mathbb{I}(y)) M_{ie} (\mathbb{O}(y)) \, dy}{\int_{\Omega} M_{ie} (\mathbb{O}(y)) \, dy} \]

\[
+ \int_{\Omega} \int_{\Omega} K_z (\mathbb{I}(y), \mathbb{I}(z)) M_{ie} (\mathbb{O}(y)) M_{ie} (\mathbb{O}(z)) \frac{2 \int_{\Omega} M_{ie} (\mathbb{O}(y)) \, dy}{(\int_{\Omega} M_{ie} (\mathbb{O}(z)))^2} \hspace{1cm} (7.24)
\]

### 7.3 Experimental results

The proposed approach is applied to 2D fluorescence microscopy images of cell nuclei from four experiments which have been introduced in Chapter 2, Section 2.4.2. Region-based and contour-based measures are used to evaluate the performance of the LSMKD algorithm. For region-based measure, the Jaccard coefficient [26] as well as Dice FP and Dice FN are used. For contour-based measures, the Hausdorff distance and MAD
are used. The mathematical description of the utilized measures has been proposed in Chapter 2, Section 2.4.3.

Table 7.1. Quantitative results for U20S data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
<th>Dice FP</th>
<th>Dice FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA[2]</td>
<td>83.8</td>
<td>4.6</td>
<td>13.6</td>
<td>8.3</td>
<td>9.7</td>
</tr>
<tr>
<td>WA[90]</td>
<td>52.4</td>
<td>11.2</td>
<td>34.1</td>
<td>30.5</td>
<td>31.5</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.1</td>
<td>11.7</td>
<td>33.9</td>
<td>12.7</td>
<td>13.3</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>77.6</td>
<td>8.3</td>
<td>24.1</td>
<td>11.8</td>
<td>12.2</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>71.7</td>
<td>6.4</td>
<td>19.4</td>
<td>15.6</td>
<td>16.4</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>72.9</td>
<td>7.8</td>
<td>21.3</td>
<td>15.3</td>
<td>14.7</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td>88.7</td>
<td>4.2</td>
<td>12.8</td>
<td>5.4</td>
<td>7.1</td>
</tr>
<tr>
<td>Three-step[32]</td>
<td>88.4</td>
<td>4.7</td>
<td>13.4</td>
<td>5.3</td>
<td>5.2</td>
</tr>
<tr>
<td>LSMKD</td>
<td><strong>90.3</strong></td>
<td><strong>3.7</strong></td>
<td><strong>11.7</strong></td>
<td><strong>4.9</strong></td>
<td><strong>3.7</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

Table 7.1 shows the quantitative results for the different performance measures which are averaged over all images in the U20S cells. The quantitative results for NIH3T3 cells are reported in Table 7.2 while the quantitative results for BBBC005 cells are reported in Table 7.3. Table 7.4 reports the quantitative results for synthetic cells. As a comparison, we also reported the results of the MA [2], the OT [91], the WA [90], and level set-based methods, namely the BLS [55], RSFE [5], the DRLSE [89], two step and three step approaches [32].

As can be seen from Table 7.1, our approach obtains the best results for the Jaccard coefficient, Hausdorff distance, MAD, Dice FP, and Dice FN for the U20S data set.
Furthermore, our approach yields better results than the level set based approaches. For the Dice FP value and the Dice FN value, our approach yields significantly smaller values. LSMKD also obtains the best results for the Jaccard coefficient, Hausdorff distance, Dice FP, and Dice FN for the NIH3T3 data set shown in Table 7.2. For the MAD, the best result is obtained by Two-step approach, although the result of the proposed approach is very close.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Jaccard</th>
<th>MAD</th>
<th>Hausdorff</th>
<th>Dice FP</th>
<th>Dice FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA[2]</td>
<td>53.4</td>
<td>6.2</td>
<td>18.7</td>
<td>28.2</td>
<td>31.7</td>
</tr>
<tr>
<td>WA[90]</td>
<td>45.2</td>
<td>5.9</td>
<td>19.1</td>
<td>35.4</td>
<td>40.5</td>
</tr>
<tr>
<td>OT[91]</td>
<td>47.3</td>
<td>12.6</td>
<td>37.4</td>
<td>38.4</td>
<td>33.5</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>61.2</td>
<td>7.5</td>
<td>22.2</td>
<td>26.1</td>
<td>22.9</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>62.8</td>
<td>6.8</td>
<td>21.5</td>
<td>20.4</td>
<td>23.5</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>61.6</td>
<td>7.8</td>
<td>23.3</td>
<td>25.3</td>
<td>23.7</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td>73.9</td>
<td><strong>4.4</strong></td>
<td>14.2</td>
<td>16.4</td>
<td>13.5</td>
</tr>
<tr>
<td>Three-step[32]</td>
<td>70.8</td>
<td>5.7</td>
<td>16.4</td>
<td>15.5</td>
<td>19.7</td>
</tr>
<tr>
<td>LSMKD</td>
<td><strong>77.9</strong></td>
<td>4.9</td>
<td><strong>13.1</strong></td>
<td><strong>11.7</strong></td>
<td><strong>13.2</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

For the BBBC005 cell images results shown in Table 7.3, LSMKD yields significantly better results for the Jaccard, Hausdorff distance, and Dice FN. In particular, for the BBBC005 images, we obtain a significantly smaller Dice FN value. For the MAD and Dice FP, the best results are obtained by Two-step approach and MA, respectively, while the results of LSMKD are very close behind.
As can be seen from Table 7.4, for the synthetic cell images, the best result for the MAD value is obtained by BLS, but our approach is only slightly poorer. However, the proposed approach has a significantly smaller Dice FP value. LSMKD obtains significantly better results for the Jaccard coefficient, Dice FN, and Hausdorff distance.

Table 7.3. Quantitative results for BBBC005 data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>BBBC005 cells (17 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaccard</td>
</tr>
<tr>
<td>MA[2]</td>
<td>80.4</td>
</tr>
<tr>
<td>WA[90]</td>
<td>75.3</td>
</tr>
<tr>
<td>OT[91]</td>
<td>76.0</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>68.0</td>
</tr>
<tr>
<td>RSFE[5]</td>
<td>70.6</td>
</tr>
<tr>
<td>DRLSE[89]</td>
<td>70.4</td>
</tr>
<tr>
<td>Two-step[32]</td>
<td>85.6</td>
</tr>
<tr>
<td>Three-step[32]</td>
<td>83.2</td>
</tr>
<tr>
<td>LSMKD</td>
<td><strong>87.1</strong></td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

Table 7.4. Quantitative results for synthetic data set for the different segmentation approaches.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Synthetic cells (20 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaccard</td>
</tr>
<tr>
<td>MA[2]</td>
<td>65.2</td>
</tr>
<tr>
<td>WA[90]</td>
<td>64.5</td>
</tr>
<tr>
<td>OT[91]</td>
<td>63.2</td>
</tr>
<tr>
<td>BLS[55]</td>
<td>72.4</td>
</tr>
<tr>
<td>Method</td>
<td>Jaccard</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
</tr>
<tr>
<td>U20S cells</td>
<td>5.7</td>
</tr>
<tr>
<td>NIH3T3 cells</td>
<td>7.8</td>
</tr>
<tr>
<td>BBBC005 cells</td>
<td>4.9</td>
</tr>
<tr>
<td>Synthetic cells</td>
<td>7.3</td>
</tr>
</tbody>
</table>

The best results are indicated by bold values.

7.4 Summary

In this chapter, the LSMKD algorithm has been proposed for cell nuclei image segmentation. In LSMKD, the kernel function is constructed as a linear combination of multiple kernels. These kernels are selected for different information channels or properties of image pixels. A flexible level set based framework is proposed for selecting and combining different kernel functions. LSMKD can be considered as a new information fusion method for level set segmentation. It combines information of the image from multiple heterogeneous or homogeneous data sources in the kernel space.
Chapter 8

Summary and Conclusions

In this concluding chapter, the achievements from this study are summarized, and the significant outcomes are highlighted. The chapter concludes with several suggestions for further research.

8.1 Research accomplishments

The first outcome of this study was a new local implicit model in a variational level set formulation for fluorescence microscopy cell image segmentation where the new energy functional minimized the Bayesian classification risk for an image patch. The proposed approach was based on the Gaussianity assumption and was sensitive to the noise and outliers. Therefore, a weighting scheme was utilized to enable the pixels in each image patch to have anisotropic weights. The proposed algorithm obtains the lowest MAD measure for all four experiments. The MAD values are approximately 18%, 7%, and 12% better than the state-of-the-art methods we compared with for U20S cells, NIH3T3 cells, and Synthetic cells, respectively.

The second outcome of this thesis was a new segmentation method based on the local spatial information and correntropy-based k-means clustering in a variational level set formulation for fluorescence microscopy cell image segmentation. Since the correntropy criterion was utilized, the proposed segmentation algorithm was robust to
outliers. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 7%, and 5% better than the state-of-the-art methods we compared with for U20S cells and BBBC005 cells, respectively.

The third outcome of this study is local segmentation algorithm based on the multiple kernel and local spatial information as a framework for fluorescence microscopy cell image segmentation. In this framework, a new segmentation algorithm in a variational level set formulation based on the piecewise constant model and multiple kernels mapping was proposed where a linear combination of multiple kernels is utilized to map implicitly the original local image data into data of a higher dimension. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 5%, and 6% better than the state-of-the-art methods we compared with for U20S data set and NIH3T3 data set, respectively.

The final contribution of this research is the development of a new implicit model with multiple kernels transforming pixel information from multiple sources for image segmentation. In this approach, kernels were selected for different information channels or properties of image pixels. The proposed approach used a linear combination of multiple kernels. It can be considered as a new information fusion method for level set segmentation. The proposed algorithm obtains the highest Jaccard coefficient value for all experiments. The Jaccard coefficient values are approximately 2%, 5%, 2%, and 4% better than the state-of-the-art methods we compared with for U20S cells, NIH3T3 cells, BBBC005 data set, and Synthetic cells, respectively.
8.2 Recommendations for future work

Based on the literature review performed on existing work, and research outcomes presented in this thesis, the following gaps have been identified.

A common difficulty with many deformable models is that the energy functional to be minimized is not convex and has local minima. The local minima of deformable segmentation models often have completely wrong levels of detail and yield poor quantitative results [123]. Compared with non-uniqueness of global minimizers, the local minima can be considered as a much more serious drawback due to the fact that global minimizers of a given model are usually all reasonable solutions. Therefore, a segmentation framework with convex energy functional which can utilize the spatial information is currently demanded in the literature.

Many solution techniques for implicit segmentation models are based on gradient descent method. An initial contour is deformed in the steepest (gradient) descent of the energy using the equations of motion which are derived using the Euler-Lagrange method and the condition that the first variation of the energy functional should vanish at an optimum. The use of a gradient descent search commonly leads to problems with convergence to local optima and slow convergence in general. There are many alternatives to gradient descent such as subdivision schemes, stochastic methods such as the Monte-Carlo family, and heuristic optimization methods. Utilizing these optimization methods would be beneficial to obtain good segmentation performance.

In this thesis, we proposed implicit models based on the multiple kernel learning for information fusion. However, the local distributions of different information channels can differ significantly across data space. Therefore, finding a globally optimal combination of multiple kernels might be difficult. It would be interesting to propose an
implicit model with localized multiple kernel method instead of using a uniform combinational kernel.
Bibliography


[8] L. Wang, C. Li, Q. Sun, D. Xia, and C.-Y. Kao, "Active contours driven by local and global intensity fitting energy with application to brain MR image


[16] Available: [http://dgenotebook.ucr.edu/courseBook/E5_A_DNA_mPing.pdf](http://dgenotebook.ucr.edu/courseBook/E5_A_DNA_mPing.pdf)


Transform-Based Marker Extraction and Contour Parameterization," Biomedical

1838-1842, 2005.

[29] W. Xiaoxu, H. Weijun, D. Metaxas, R. Mathew, and E. White, "CELL
SEGMENTATION AND TRACKING USING TEXTURE-ADAPTIVE

with active contour models," Machine Vision and Applications, vol. 23, pp. 91-
109, 2012.

[31] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed:
algorithms based on Hamilton-Jacobi formulations," Journal of computational

fluorescence microscopy images using level sets and convex energy


