2011

Gaussian Mixture Model based Spatial Information Concept for Image Segmentation

Thanh Nguyen
University of Windsor

Gaussian Mixture Model based Spatial Information Concept for Image Segmentation

by

Thanh Minh Nguyen

A Dissertation
Submitted to the Faculty of Graduate Studies
through Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

2011

Author's Declaration of Originality

I hereby declare that I am the sole author of this thesis. I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Segmentation of images has found widespread applications in image recognition systems. Over the last two decades, there has been a growing research interest in model-based technique. In this technique, standard Gaussian mixture model (GMM) is a well-known method for image segmentation. The model assumes a common prior distribution, which independently generates the pixel labels. In addition, the spatial relationship between neighboring pixels is not taken into account of the standard GMM. For this reason, its segmentation result is sensitive to noise. To reduce the sensitivity of the segmented result with respect to noise, Markov Random Field (MRF) models provide a powerful way to account for spatial dependencies between image pixels. However, their main drawback is that they are computationally expensive to implement.

Based on these considerations, in the first part of this thesis (Chapter 4), we propose an extension of the standard GMM for image segmentation, which utilizes a novel approach to incorporate the spatial relationships between neighboring pixels into the standard GMM. The proposed model is easy to implement and compared with the existing MRF models, requires lesser number of parameters. We also propose a new method to estimate the model

parameters in order to minimize the higher bound on the data negative log-likelihood, based on the gradient method. Experimental results obtained on noisy synthetic and real world grayscale images demonstrate the robustness, accuracy and effectiveness of the proposed model in image segmentation.

In the final part of this thesis (Chapter 5), another way to incorporate spatial information between the neighboring pixels into the GMM based on MRF is proposed. In comparison to other mixture models that are complex and computationally expensive, the proposed method is robust and fast to implement. In mixture models based on MRF, the M-step of the EM algorithm cannot be directly applied to the prior distribution for maximization of the log-likelihood with respect to the corresponding parameters. Compared with these models, our proposed method directly applies the EM algorithm to optimize the parameters, which makes it much simpler. Finally, our approach is used to segment many images with excellent results.

To my beloved Dad and Mom

&

My beloved wife, Dang Nguyen Thanh Yen

Acknowledgements

I am deeply indebted to my thesis advisor, Prof. Jonathan Wu, for his informed guidance, understanding, support and encouragement through the years. His depth and vision in this area are truly admirable, and he has been an inspiration to me in many ways.

I am grateful to my committee members Prof. Narayan Kar, Prof. Huapeng Wu, Prof. Dan Wu, and Prof. Peter X. Liu for reviewing my thesis and making many excellent suggestions to improve its substance and presentation. Their openness, encouragement and interest in my research were very important to me.

Thanks also go to Andria Ballo and Shelby Marchand from Department of Electrical and Computer Engineering for their reminders and assistance in various administrative tasks.

I am grateful that during the past four years, I get to know many other talented graduate students: Siddhant Ahuja, Mohammed Golam Sarwer, Baradarani Aryaz, Rashid Minhas, AbdulAdeel Mohammed, Dibyendu Mukherjeey, and Ashirbani Saha. I want to thank them for their help, support, interest, and valuable hints. I would also like to thank other members

in our lab for encouragements and the joys and tears that we shared.

I thank my parents who always stand beside me. I could not complete this long journey without their never-ending love, support and prayer. I cannot find any word to express my deep appreciation to them. I just want to tell them how much I love them. I love you, Dad and Mom, and thanks for always being there for me.

Finally, I thank my best friend and beloved wife Dang Nguyen Thanh Yen, who always takes care of many things around me since I came here so that I could concentrate on this thesis. Her love has been a strong energy for my entire study.

This research has been supported in part by the Canada Research Chair Program, AUTO21 NCE, the NSERC Discovery grant, and OGS.

Thanh Minh Nguyen

The University of Windsor

August 2011

Table of Contents

Author's Declaration of Originality iii

Abstract iv

Dedication vi

Acknowledgements vii

List of Figures xiii

List of Tables xx

Symbols and Abbreviations xxi

1 Introduction 1

1.1 General Introduction 1

1.2 Thesis Overview 7

2 Standard Finite Mixture Model for Image Segmentation 10

2.1 Probability Distributions 10

2.1.1 The Gaussian Distribution 10
2.1.2 The Student's-t distribution 12
2.2 Maximum likelihood for the Gaussian 14
2.3 Standard Finite Mixture Model 16
2.3.1 Gaussian Mixture Model 16
2.3.2 Student's-t Mixture Model 19
2.4 The Expectation Maximization (EM) Algorithm 22
2.4.1 EM Algorithm for the Gaussian Mixture Model 22
2.4.2 Relation between EM and K-means 25
2.5 Gradient-Based Optimization Techniques 27
2.6 Image Segmentation Evaluation 30
2.7 Conclusions 31

3 Gaussian Mixture Model based Markov Random Field 33
3.1 Introduction 33
3.2 Gaussian Mixture Model based MRF for the Pixel Labels . . . 34
3.2.1 Markov Random Field Theory 34
3.2.2 Hidden Markov models 36
3.3 Gaussian Mixture Model based MRF for the Priors of the Pixel Labels 39
3.4 Conclusions 44

4 An Extension of the Standard Mixture Model for Image Segmentation 45
4.1 Introduction 45

4.2 Proposed Method 46
4.3 Parameter Learning 50
4.4 Experiments 52
4.4.1 Synthetic Images 54
4.4.2 Natural Images 58
4.5 Conclusions 65

5 Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation 66

5.1 Introduction 66
5.2 Proposed Method 70
5.3 Experiments 74
5.3.1 Segmentation of Synthetic Images 75
5.3.2 Segmentation of Grayscale Natural Images 81
5.3.3 Segmentation of Colored Images 83
5.4 Conclusions 87

6 Conclusions 89

6.1 Summary of Main Contributions 89
6.2 Future Directions 91

References 94

Appendix A 110

Publications 112

Vita Auctoris **115**

List of Figures

2.1 Plot of the Gaussian distribution showing the mean μ and standard deviation σ. 11

2.2 Plot of Student's-t distribution for $\mu = 0$ and $\Lambda = 1$ for various values of v. The limit $v \to \infty$ corresponds to a Gaussian distribution with mean μ and precision Λ. 13

2.3 The graphical representation of a Gaussian mixture model for a set of N pixel x_i. 17

2.4 Synthetic image, (a): original image, (128x128 image) (b): Corrupted original image with Gaussian noise (0 mean, 0.02 variance), (c): standard GMM. 18

2.5 Real world image (321x481 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.02 variance), (c): standard GMM 18

2.6 Synthetic image, (a): original image, (128x128 image) (b): Corrupted original image with Gaussian noise (0 mean, 0. 005 variance), (c): standard GMM, (c): standard SMM. 21

2.7 EM algorithm for the mixture of gaussians, (a): The original 2D point set with the initial condition, (b): Result of GMM with this initial condition. 24

3.1 Synthetic image, (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.05 variance), (c): standard GMM, (d): SIMF [63]. 38

3.2 Synthetic image, (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.01 variance), (c): standard GMM (time = 5.5s), (d): SVFMM [72] (time = 246.1s). . 43

4.1 The first experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.05 variance), (c): K-means (MCR = 17.315%), (d): standard GMM (MCR = 33.982%), (e): SVFMM (MCR = 13.671%), (f): NEM (MCR = 17.034%), (g) ICM (MCR = 9.605%), (h): MODEF (MCR = 7.781%), (i): SIMF (MCR = 7.725%), (k): MEANF (MCR = 7.721%), (l) HMRF-FCM (MCR = 0.823%), (m): The proposed method (MCR = 0.653%). 54

4.2 The second experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with mixed noise (salt and pepper noise (0.03%) + Gaussian noise (0 mean, 0.01 variance)), (c): K-means (MCR = 24.609%), (d): standard GMM (MCR = 40.209%), (e): SVFMM (MCR = 22.338%), (f): ICM (MCR = 22.001%), (g): MODEF (MCR = 12.744%), (h): SIMF (MCR = 13.307%), (i): MEANF (MCR = 12.200%), (k) FGFCM (MCR = 5.562%), (l) HMRF-FCM (MCR = 3.692%), (m): The proposed method (MCR = 2.789%). 56

4.3 Affect of the simple competitive selection, (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.03 variance), (c): MODEF (MCR = 3.512%), (d): MODEF with the simple competitive selection (MCR = 3.117%), (e): proposed method without the simple competitive selection (MCR = 0.199%), (f): proposed method with the simple competitive selection (MCR = 0.197%). 58

4.4 Images from the Berkeleys grayscale image segmentation dataset, (a): 135069, (b): 124084, (c): 58060, (d): 353013 with Gaussian noise (0 mean, 0.001 variance), (e): 239007, (f): 46076, (g): 15088 with Gaussian noise (0 mean, 0.005 variance), (h): 374067 with Gaussian noise (0 mean, 0.01 variance), (i): 302003 with Gaussian noise (0 mean, 0.01 variance). 59

4.5 Image segmentation results obtained by employing the proposed method, (a): 135069, (b): 124084, (c): 58060, (d): 353013 with Gaussian noise (0 mean, 0.001 variance), (e): 239007, (f): 46076, (g): 15088 with Gaussian noise (0 mean, 0.005 variance), (h): 374067 with Gaussian noise (0 mean, 0.01 variance), (i): 302003 with Gaussian noise (0 mean, 0.01 variance). 60

4.6 Grayscale image segmentation (55067), (a): original image, (b): K-means (PR = 0.879), (c): standard GMM (PR = 0.843), (d): SVFMM (PR = 0.882), (e): ICM (PR = 0.880), (f): MODEF (PR = 0.882), (g): MEANF (PR = 0.881), (h): FGFCM (PR = 0.879), (i): HMRF-FCM (PR = 0.887), (k): The proposed method (PR = 0.891). 62

4.7 Noisy grayscale image segmentation (24063), (a): original image, (b): corrupted original image with Gaussian noise (0 mean, 0.005 variance), (c): K-means (PR = 0.778), (d): standard GMM (PR = 0.765), (e): SVFMM (PR = 0.787), (f): MODEF (PR = 0.814), (g): MEANF (PR = 0.818), (h): FGFCM (PR = 0.796), (i): HMRF-FCM (PR = 0.815), (k): The proposed method (PR = 0.826). 63

4.8 Computational cost (in seconds) comparison, (a): original image, (b): K-means (0.7 sec), (c): standard GMM (36 sec), (d): ICM (169 sec), (e): MODEF (390 sec), (f): SIMF (432 sec), (g): MEANF (445 sec), (h): The proposed method (61 sec). . 64

4.9 Minimization Progress of the negative log-likelihood function of the proposed algorithm, for the final experiment. 64

5.1 The first experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.03 variance), (c): K-means, (d): Standard GMM (MCR = 41.67%), (e): SVFMM (MCR = 23.28%), (f): CA–SVFMM (MCR = 20.29%), (g): ICM (MCR = 20.23%), (h): SIMF (MCR = 3.83%), (i) MEANF (MCR = 3.55%), (j): Proposed method (MCR = 1.13%). 76

5.2 Maximization progress of the log-likelihood of the proposed method of the first experiment. 77

5.3 The second experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.05 variance), (c): K-means, (d): Standard GMM (MCR = 35.02%), (e): SVFMM (MCR = 12.01%), (f): CA–SVFMM (MCR = 11.16%), (g): ICM (MCR = 7.65%), (h): SIMF (MCR = 5.65%), (i) MEANF (MCR = 5.94%), (j): Proposed method (MCR = 1.05%). 78

5.4 The third experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.03 variance), (c): K-means, (d): Standard GMM (MCR = 30.31%), (e): SVFMM (MCR = 18.11%), (f): CA-SVFMM (MCR = 17.73%), (g): ICM (MCR = 5.90%), (h): SIMF (MCR = 2.86%), (i) MEANF (MCR = 2.70%), (j): Proposed method (MCR = 0.21%). 80

5.5 The fourth experiment (256x256 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.1 variance), (c): Proposed method in Chapter 4 (MCR = 0.10%, time = 8.9s), (j): Proposed method in Chapter 4 (MCR = 0.22%, time = 3.7s). 81

5.6 Grayscale natural image segmentation (80099), (a): original image, (b): Standard GMM, (c): SVFMM, (d): CA–SVFMM, (e): SIMF, (f) MEANF, (g): Proposed method, (h): Maximization progress of the log-likelihood of the proposed method of this experiment. 82

5.7 Grayscale natural image segmentation (86016), (a): original image, (b): Standard GMM, (c): SVFMM, (d): CA–SVFMM, (e): SIMF, (f) MEANF, (g): Proposed method, (h): Maximization progress of the log-likelihood of the proposed method. 83

5.8 Grayscale natural image segmentation (374067), (a): original image, (b): Standard GMM, (c): SVFMM, (d): SIMF, (e): Proposed method. 84

5.9 Color image segmentation (310007), (a): original image, (b): Standard GMM, (c): SVFMM, (d): CA–SVFMM, (e): SIMF, (f) MEANF, (g): Proposed method, (h): Maximization progress of the log-likelihood of the proposed method. 85

5.10 Color image segmentation (388016), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.0015 variance), (c): Standard GMM, (d): SVFMM, (e): CA–SVFMM, (f): ICM, (g) MEANF, (h): Proposed method. 86

5.11 (first row): original image, (second row): SVFMM, (third row): MEANF, (last row): Proposed method. 88

List of Tables

4.1 Comparison of the proposed method with other methods in term of MCR (%), for the first experiment. 55

4.2 Comparison of the proposed method with other methods in term of MCR (%), for the second experiment. 57

4.3 Comparison of image segmentation results on Berkeley's grayscale image segmentation dataset: Probabilistic Rand (PR) Index. . 61

5.1 Computational cost (in seconds) comparison for the synthetic image in the first experiment. 76

5.2 Computational cost (in seconds) comparison for the synthetic image in the second experiment. 79

5.3 Comparison of the proposed method with other methods in term of MCR (%), for the third experiment, in the presence of varying levels of noise. 79

5.4 Comparison of image segmentation results on Berkeley's color image segmentation dataset: Probabilistic Rand (PR) Index. . 87

Symbols and Abbreviations

x_i	:	The i-th pixel of an image
N	:	The number of pixels in an image
Ω_j	:	The j-th label in an image
K	:	The number of labels in an image
$\mathcal{N}_i$	:	The neighborhood pixels of the i-th pixel
$\Phi(\mathrm{x}_i\|\Theta_j)$	:	The Gaussian function
μ_j	:	The mean of the Gaussian function $\Phi(\mathrm{x}_i\|\Theta_j)$
σ_j	:	The standard deviation of the Gaussian function
	:	$\Phi(\mathrm{x}_i\|\Theta_j)$, for the case of a single real-valued variable x_i
Σ_j	:	The covariance of the Gaussian function $\Phi(\mathrm{x}_i\|\Theta_j)$,
	:	for the case of a D-dimensional vector x_i
π_j	:	The prior probability of all pixel belonging to the label Ω_j
π_{ij}	:	The prior probability of the pixel x_i belonging
	:	to the label Ω_j
z_{ij}	:	The posterior probability of the pixel x_i belonging
	:	to the label Ω_j

$f(\mathrm{x}_i\|\Pi,\Theta)$	:	The density function at an observation x_i
$p(\mathrm{X}\|\Pi,\Theta)$	:	The joint conditional density
$L(\Theta,\Pi\|\mathrm{X})$	:	The log-likelihood function
$J(\Theta,\Pi\|\mathrm{X})$	:	The negative log-likelihood function
$H(\Pi,\Theta\|\mathrm{X})$	:	The objective function
$E(\Theta,\Pi)$	:	The error function
GMM	:	Gaussian mixture model
SMM	:	Student's-t mixture model
MRF	:	Markov random field
EM	:	Expectation maximization
SVFMM	:	Spatially variant finite mixture model
CA-SVFMM	:	Class adaptive spatially variant finite mixture model
NEM	:	Neighborhood expectation maximization
ICM	:	Iterated conditional model
MODEF	:	Mode field algorithm
SIMF	:	Simulated field algorithm
MEANF	:	Mean field algorithm
FGFCM	:	Fast generalized fuzzy c-means
HMRF-FCM	:	Hidden Markov random field based fuzzy c-means
MCR	:	The misclassification ratio
PR	:	The probabilistic rand index
Dice_j	:	Dice similarity coefficient for the label Ω_j

Chapter 1

Introduction

1.1 General Introduction

In order to analyse the content of an image, it is often useful to construct a simpler representation of multiple segments. And the process to partition an image into non-overlapping regions that humans can easily separate is called image segmentation. In an image, various features can be used for segmentation process. These might be colour information that is used to create histograms, or information about the pixels that indicate boundaries or texture information.

Segmentation is an important step of low level vision. An accurately segmented image provides detailed information about the objects present in an image and their respective boundaries. There are many applications of segmentation. For example, in a vision guided tool tracker system [1], [2], [3], the robot needs to track the appropriate components in automotive manufacturing

environments, thereby increasing the productivity and profitability of automotive manufacturing enterprises and the global competitiveness. In the field of medical imaging [4], [5], [6], [7], segmentation plays an important role. Accurate medical image segmentation provides additional information that helps to prepare treatment scheme and to evaluate therapeutic effect. The applications of segmentation vary from the detection of synthetic aperture radar images [8], [9], video analysis [10], to magnetic resonance imaging (MRI) [11], and object detection [12], [13], [14] etc. In all these areas, the quality of the segmented output affects on the quality of the final output largely. However, automated segmentation [15] is still a very challenging research topic, due to overlapping intensities and low contrast in images, as well as noise perturbation.

Many previous works have been proposed for image segmentation, in particular by the method of threshold [16], [17], [18]. However, thresholding is significantly susceptible to low resolution, low contrast and signal to noise ratio. As for some part of the image, high intensity variation may correspond to edges of interest, while the other part may require high low variation. The selection of the threshold is very crucial. A bad choice of threshold [19] leads to a poor quality of the segmentation. Adaptive thresholding [20], [21], [22] often is taken as a solution to this. However, it cannot eliminate the problem of threshold selection [23].

In order to avoid the above-mentioned disadvantages, an artificial neural network [24], [25], [26] is applied for image segmentation. In [27], the authors clustered feature vectors extracted from an image using a neural network which

minimized the distance between the feature vectors. Although this approach worked well in the examples shown, it led to sub-optimal image segmentation. This is because the pixels in general are spatially correlated and the approach presented in this method did not incorporate any spatial information.

Many algorithms have been developed for image segmentation including graph-based methods [28], [29], mean shift based methods [30], [31], histogram-based methods [32], multi-scale segmentation [33], and clustering methods [34], [35], [36]. In clustering methods, K-means [37], [86] and fuzzy c-means [38] are two well-known methods that have been widely used for segmenting an image due to their simplicity and ease of implementation. However, one of their main drawbacks [39]–[42] is that these two methods ignore the spatial constraints in an image.

During the last decades, much attention has been given to model-based techniques [43], [44], [45], [100]–[105] to model the uncertainty in a probabilistic manner. In model-based techniques, standard Gaussian mixture model (GMM) [46]–[49] is a well-known method. It is a flexible and powerful statistical modeling tool for multivariate data. Many researchers have used it to study a number of key problems in the area of image segmentation [50], [51]. In standard GMM, each pixel x_i is considered to be a random variable whose possibility density function $\Phi(\mathrm{x}_i|\Theta_j)$ is a Gaussian function. The model assumes a common prior distribution π_j, which independently generates the pixel labels. In order to estimate the model parameters, expectation maximization (EM) algorithm [52]–[57] is employed to maximize the log-likelihood of the

given data set. The main advantage of the standard GMM is that it is easy to implement and requires a small number of parameters. The log-likelihood function that is used to estimate the parameters is inherently simple. However, one of the main drawbacks of this model is that the prior distribution π_j has no dependence on the pixel index i. One of the other problems is that the spatial relationships between the neighboring pixels are not taken into its account [58]. Although the standard GMM is a well known and simple method for image segmentation, its segmentation result is thus sensitive to noise, varying illumination and other environmental factors such as wind, rain or camera shaking.

In order to reduce the segmentation sensitivity to noise, mixture models with Markov random field (MRF) have been employed for pixel labels [59], [62], [63], [64], [66]. The most important distinction is that in standard GMM, a common prior distribution π_j for all pixels x_i is evaluated, whereas, in these approaches, the prior distribution π_{ij} varies for every pixel x_i corresponding to each label Ω_j and depends on the neighboring pixels and their corresponding parameters [67]. This prior distribution π_{ij} is a probability. Although these approaches can lead to an improved segmentation quality, they lack enough robustness with respect to noise. In addition, the computational cost of the MRF based methods remains quite high.

To incorporate the spatial relationships in a given image, several researchers have suggested the GMM model based on MRF [58], [68], [69], [70], [71], [72], [73], where an MRF models the joint distribution of the priors of each

pixel label, instead of the joint distribution of the pixel labels as in [59], [63], [64], [66]. These models work well for noisy image segmentation; however, in order to accurately evaluate the influence of the neighboring pixel labels during the learning step, the algorithm becomes complex and computationally expensive. In order to maximize the log-likelihood with respect to the parameters in [58], [72], [73], the M-step of EM algorithm [58], [72] cannot be applied directly to the prior distribution π_{ij}. Therefore, various approximations have been introduced in order to tackle this problem. For example, the MAP algorithm in [58] cannot evaluate the prior distribution π_{ij} in a closed form, and thus the gradient projection algorithm was proposed to implement the M-step. In [72], [73], another method based on a closed form update equation was used to implement the M-step, and estimate the parameters. As compared to standard GMM based methods, the computational cost of these methods remains high. However, in addition to increased complexity, the final segmented image lacks adequate robustness to noise.

Based on these considerations, in the first part of this thesis, we propose an extension of the standard GMM [82], [83] for image segmentation, which utilizes a novel approach to incorporate the spatial relationships between neighboring pixels into the standard GMM. The proposed model is similar to the standard GMM and thus easy to implement, with the main difference that the prior distribution of each label Ω_j is different for each pixel x_i and depends on its neighboring pixels. A new way to properly account for the relationship of neighboring pixels is introduced. In addition, the proposed method does not

require as many parameters as compared to the models based on MRF. To estimate the unknown parameters of the pixel's prior distributions, as well as the parameters of the distribution itself, instead of using the EM algorithm, we use the gradient method to minimize a higher bound on the data negative log-likelihood. The proposed method has been applied for segmenting synthetic and real world grayscale images. The performance of the proposed model is compared with other methods based on standard GMM and MRF models, there by demonstrating its robustness, accuracy and effectiveness.

In the final part of this thesis, a new mixture model for image segmentation [84] is presented, which differs from the above methods in the following manner. Firstly, our proposed method incorporates spatial relationships amongst neighboring pixels in a simpler metric based on MRF. Therefore, the proposed method is fast and easy to implement, compared with other mixture models that are complex and computationally expensive. Generally, in abovementioned models based MRF, the M-step of the EM algorithm cannot be directly applied for the maximization of the log-likelihood with respect to the parameters. In our proposed method, we can directly apply the EM algorithm to optimize the parameters, which makes it simpler. Finally, the proposed model is quite robust with respect to noise, more accurate and effective as compared to other GMM based methods.

1.2 Thesis Overview

In the Chapter 2, the first group of model-based techniques is described beginning with the using of standard GMM to solve the fully unsupervised segmentation problem. The advantages and disadvantages of the standard GMM are then discussed. Next, in order to estimate the parameters of the model, various techniques based on maximizing their likelihood are described, beginning with the EM algorithm, then continuing with the gradient-based optimization techniques. The criteria evaluation for unsupervised segmentation algorithms is addressed before concluding with some observations regarding the relevance of the reviewed literature to the direction of research presented in our thesis.

In Chapter 3, we describe the second group of model-based techniques for unsupervised segmentation. In order to take into account the spatial correlation between the neighboring pixels and reduce the sensitivity of the segmentation result with respect to noise, the mixture models presented here is based on the MRF for modeling and processing image data. In the second group of model-based techniques, the mixture models based on MRF can be divided into two types. In the first type, mixture models with MRF have been employed for pixel labels. In order to take into account the spatial correlation, an MRF model in the second type is used to model the joint distribution of the priors of each pixel label, instead of the joint distribution of the pixel labels as in first type. For each type of mixture models in this chapter, the limitations and advantages are also described. Moreover, the main differences between the first and the second group of model-based techniques are discussed. The

most important distinction is that in standard GMM of the first group, a common prior distribution for all pixels is evaluated, whereas, in the approaches of the second group, the prior distribution varies for every pixel corresponding to each label and depends on the neighboring pixels and their corresponding parameters.

Chapter 4 presents new unsupervised segmentation algorithms [82], [83]. In this chapter, we propose an extension of the standard GMM for image segmentation, which utilizes a novel approach to incorporate the spatial relationships between neighboring pixels into the standard GMM. The proposed model is easy to implement and compared with MRF models, requires fewer parameters. We also propose a new method to estimate the model parameters in order to minimize the higher bound on the data negative log-likelihood, based on the gradient method. Results are presented before conclusions are drawn.

Chapter 5 describes a new way to incorporate spatial information between the neighboring pixels into the GMM based on MRF [84]. In comparison to other mixture models that are complex and computationally expensive, the proposed method is fast and easy to implement. In mixture models based on MRF, the M-step of the EM algorithm cannot be directly applied to the prior distribution for maximization of the log-likelihood with respect to the corresponding parameters. Compared with these models, our proposed method directly applies the EM algorithm to optimize the parameters, which makes it much simpler. Finally, experimental results obtained by employing the pro-

posed method on many synthetic and real-world grayscale and colored images demonstrate its robustness, accuracy and effectiveness, as compared with other mixture models.

Chapter 6 concludes the thesis with suggestions for further work.

Chapter 2

Standard Finite Mixture Model for Image Segmentation

2.1 Probability Distributions

2.1.1 The Gaussian Distribution

In probability theory, one of the most important distributions for continuous variables is Gaussian distribution. It is historically called the law of errors and is considered the most popular probability distribution in practice, and is used throughout statistics. For the case of a single real-valued variable x, the Gaussian distribution has its own mean μ and standard deviation σ and is defined by:

$$\Phi(\mathrm{x}|\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}\right) \tag{2.1}$$

Where $\Theta = \{\mu, \sigma\}$. The graph of the Gaussian distribution [80] is shown in Figure 2.1.

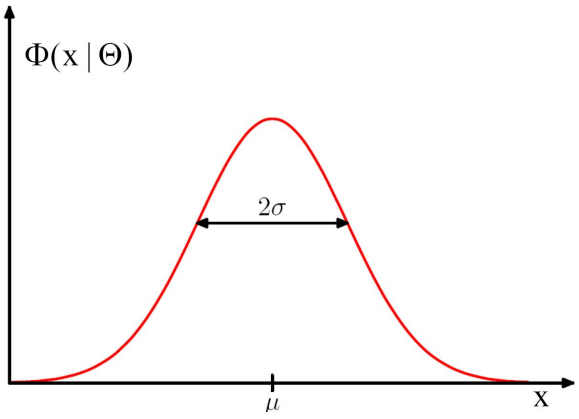

Figure 2.1: Plot of the Gaussian distribution showing the mean μ and standard deviation σ.

As shown in Eq.(2.1), we see that the Gaussian distribution satisfies the two requirements for a valid probability density.

$$\Phi(\mathrm{x}|\Theta) > 0 \tag{2.2}$$

And,

$$\int_{-\infty}^{\infty} \Phi(\mathrm{x}|\Theta) d\mathrm{x} = 1 \tag{2.3}$$

Within the Gaussian distribution defined by Eq.(2.1), the average value of x is

$$\mathbb{E}\left[\mathrm{x}\right] = \int_{-\infty}^{\infty} \Phi(\mathrm{x}|\Theta)\mathrm{x} d\mathrm{x} = \mu \tag{2.4}$$

And the variance of x is given by

$$\mathrm{var}\left[\mathrm{x}^2\right] = \mathbb{E}\left[\mathrm{x}^2\right] - \mathbb{E}[\mathrm{x}]^2 = \int_{-\infty}^{\infty} \Phi(\mathrm{x}|\Theta)\mathrm{x}^2 d\mathrm{x} - \mu^2 = \sigma^2 \tag{2.5}$$

For the case of a D-dimensional vector x, each Gaussian distribution $\Phi(\mathrm{x}|\Theta)$ can be written in the form:

$$\Phi(\mathrm{x}|\Theta) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathrm{x}-\mu)^T \Sigma^{-1}(\mathrm{x}-\mu)\right\} \tag{2.6}$$

where $\Theta = (\mu, \Sigma)$. The D-dimensional vector μ is the mean, the $D\mathrm{x}D$ matrix Σ is the covariance, and $|\Sigma|$ denotes the determinant of Σ.

2.1.2 The Student's-t distribution

Student's-t distribution is a continuous probability distribution that is heavily tailed than Gaussian. Hence, it is more prone to producing values that fall far from its mean. Student's-t distribution plays an important role in a number of widely used statistical analysis and is used to estimate the mean of a normally distributed population in situations where the sample size is small. It has proven to be quite effective for image segmentation [106].

Student's-t distribution is symmetric and bell-shaped, like the normal distribution, meaning that it has its own parameters $\Theta_j = \{\mu, \Lambda, v\}$ with mean μ, precision (inverse covariance) Λ and degree of freedom v. For the case of a D-dimensional vector x, it has the probability density function.

$$S(\mathrm{x}|\Theta) = \frac{\Gamma(v/2 + D/2)|\Lambda|^{1/2}}{\Gamma(v/2)(v\pi)^{D/2}} \left(1 + \frac{\Delta^2}{v}\right)^{-(v+D)/2} \tag{2.7}$$

where, $\Gamma(\cdot)$ is the Gamma function [80]:

$$\Gamma(y) = \int_0^\infty t^{y-1} e^{-t} dt \tag{2.8}$$

and, Δ^2 is the squared Mahalanobis distance from pixel x to mean μ.

$$\Delta^2 = (\mathrm{x} - \mu)^{\mathrm{T}} \Lambda (\mathrm{x} - \mu) \tag{2.9}$$

The degree of freedom v is illustrated in Figure 2.2. For the particular case of $v = 1$, the Student's-t distribution reduces to the Cauchy distribution [80], while in the limit $v \to \infty$ the Student's-t distribution becomes a Gaussian with mean μ and precision Λ. Eq.(2.7) is the multivariate form of Student's-t

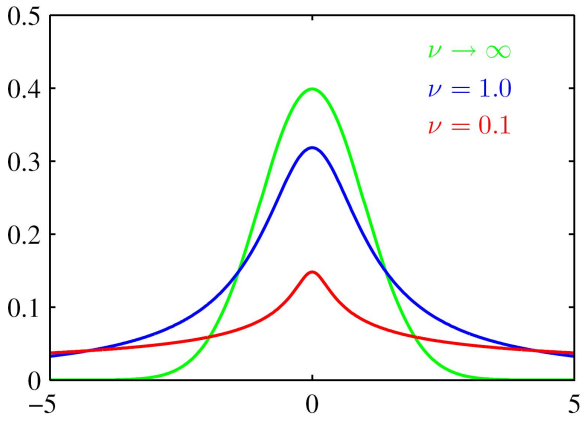

Figure 2.2: Plot of Student's-t distribution for $\mu = 0$ and $\Lambda = 1$ for various values of v. The limit $v \to \infty$ corresponds to a Gaussian distribution with mean μ and precision Λ.

distribution and satisfies the following properties

$$\mathbb{E}\left[\mathrm{x}\right] = \int_{-\infty}^{\infty} S(\mathrm{x}|\Theta)\mathrm{x}d\mathrm{x} = \mu \tag{2.10}$$

And the variance of x is given by

$$\mathrm{var}\left[\mathrm{x}^2\right] = \mathbb{E}\left[\mathrm{x}^2\right] - \mathbb{E}[\mathrm{x}]^2 = \int_{-\infty}^{\infty} S(\mathrm{x}|\Theta)\mathrm{x}^2 d\mathrm{x} - \mu^2 = \frac{v}{v-2}\Lambda^{-1} \tag{2.11}$$

2.2 Maximum likelihood for the Gaussian

Let x_i denote an observation. Given a data set $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)$ in which the observations x_i are assumed to be drawn independently from a Gaussian distribution, we can estimate the parameters of the Gaussian distribution by maximum likelihood [80]. The log-likelihood function is given by

$$L(\Theta|\mathrm{X}) = \frac{ND}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}\frac{(\mathrm{x}_i - \mu)^2}{\sigma^2} \tag{2.12}$$

where $\Theta = \{\mu, \sigma\}$. After some manipulation, we see that the likelihood function depends on the data set only through the two quantities

$$\sum_{i=1}^{N}\mathrm{x}_i \text{ and } \sum_{i=1}^{N}\mathrm{x}_i^2 \tag{2.13}$$

The next objective is to optimize the parameter set $\Theta = \{\mu, \sigma\}$ in order to maximize the log-likelihood function in Eq.(2.12). Let us now consider the derivation of the function $L(\Theta|\mathrm{X})$ with the means μ, we have:

$$\frac{\partial L(\Theta|\mathrm{X})}{\partial \mu} = \sum_{i=1}^{N}\frac{\mathrm{x}_i - \mu}{\sigma^2} \tag{2.14}$$

and setting this derivative to zero, we obtain the solution for the maximum likelihood estimate of the mean at the $(t+1)$ step. The mean of the observed set of data points is given by

$$\mu^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{x}_i \tag{2.15}$$

The maximization of log-likelihood function $L(\Theta|\mathrm{X})$ in Eq.(2.12) with respect to σ is rather more involved. Setting the derivative of the function in Eq.(2.12)

with σ at the $(t+1)$ iteration step, we have:

$$\frac{\partial L(\Theta|\mathrm{X})}{\partial \sigma} = \sum_{i=1}^{N} \left(-\frac{1}{\sigma} + \frac{(\mathrm{x}_i - \mu)^2}{\sigma^3} \right) \tag{2.16}$$

The solution of $\partial L(\Theta|\mathrm{X})/\partial\sigma = 0$ yields the minimizer of σ at the $(t+1)$ step. The result is as expected and takes the form

$$[\sigma^2]^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} (\mathrm{x}_i - \mu^{(t+1)})^2 \tag{2.17}$$

Note that the solution for $\mu^{(t+1)}$ in Eq.(2.15) does not depend on $\sigma^{(t+1)}$, and so we can first evaluate $\mu^{(t+1)}$ and then use this to evaluate $\sigma^{(t+1)}$. If we evaluate the expectations of the maximum likelihood solutions under the true distribution in Eq.(2.15) and Eq.(2.17), we obtain the following results

$$\mathbb{E}\left[\mu^{(t+1)}\right] = \mu \tag{2.18}$$

and,

$$\mathbb{E}\left[[\sigma^2]^{(t+1)}\right] = \frac{N-1}{N}\sigma^2 \tag{2.19}$$

As shown in Eq.(2.18), the expectation of the maximum likelihood estimate for the mean is equal to the true mean. However, the maximum likelihood estimate for the covariance has an expectation that is less than the true value. We can correct this bias by defining a different estimator $[\tilde{\sigma}^2]^{(t+1)}$ given by

$$[\tilde{\sigma}^2]^{(t+1)} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathrm{x}_i - \mu^{(t+1)})^2 \tag{2.20}$$

From Eq.(2.19) and (2.20), the expectation of $[\tilde{\sigma}^2]^{(t+1)}$ is equal to σ^2.

2.3 Standard Finite Mixture Model

2.3.1 Gaussian Mixture Model

Over the last few years, much attention has been given to the standard GMM. An advantage of the standard GMM is that it requires a small amount of parameters for learning. Another advantage is that these parameters can be efficiently estimated by adopting the EM algorithm to maximize the log-likelihood function.

Let x_i; i=1,2,...,N; denote the observation at the i-th pixel of an image. Labels are denoted by Ω_1,Ω_1,...,Ω_K. Consider the problem of estimating the posterior probability of x_i belonging to label Ω_j. If we assume that x_i is drawn independently from the distribution, then the standard GMM [46], [75] assumes that the density function at an observation x_i is given by:

$$f(\mathrm{x}_i|\Pi,\Theta) = \sum_{j=1}^{K} \pi_j \Phi(\mathrm{x}_i|\Theta_j) \tag{2.21}$$

The graphical representation of a Gaussian mixture model for a set of N pixel x_i is shown in Figure 2.3. Where $\Pi = \{\pi_1, \pi_2, ..., \pi_K\}$, and π_j is the prior distribution of the pixel x_i belonging to the label Ω_j, which satisfies the constraints:

$$0 \leq \pi_j \leq 1 \text{ and } \sum_{j=1}^{K} \pi_j = 1 \tag{2.22}$$

Each Gaussian distribution $\Phi(\mathrm{x}_i|\Theta_j)$ is called a component of the mixture. For the case of a single real-valued variable x_i, the Gaussian distribution

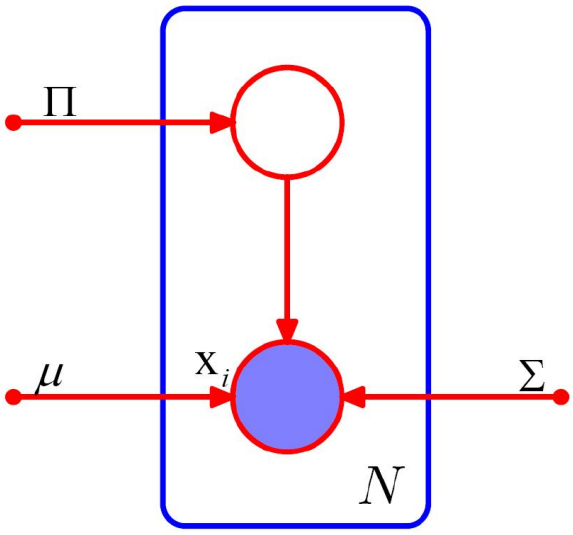

Figure 2.3: The graphical representation of a Gaussian mixture model for a set of N pixel x_i.

has its own mean μ_j and covariance σ_j and is defined by:

$$\Phi(\mathrm{x}_i|\Theta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\mathrm{x}_i - \mu_j)^2}{2\sigma_j^2}\right) \tag{2.23}$$

where $\Theta_j = \{\mu_j, \sigma_j\}$. The observation x_i in Eq.(2.21) is modeled as statistically independent. And the joint conditional density [58], [80] of the data set $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)$ can be modeled as:

$$p(\mathrm{X}|\Pi, \Theta) = \prod_{i=1}^{N} f(\mathrm{x}_i|\Pi, \Theta) = \prod_{i=1}^{N} \left[\sum_{j=1}^{K} \pi_j \Phi(\mathrm{x}_i|\Theta_j)\right] \tag{2.24}$$

Given the joint conditional density from Eq.(2.24), the log-likelihood function of the standard GMM [80] is given by:

$$L(\Theta, \Pi|\mathrm{X}) = \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_j \Phi(\mathrm{x}_i|\Theta_j)\right\} \tag{2.25}$$

Where $\Theta = \{\Theta_j\}$; j=1,2,...,K. As can be seen from the likelihood function in

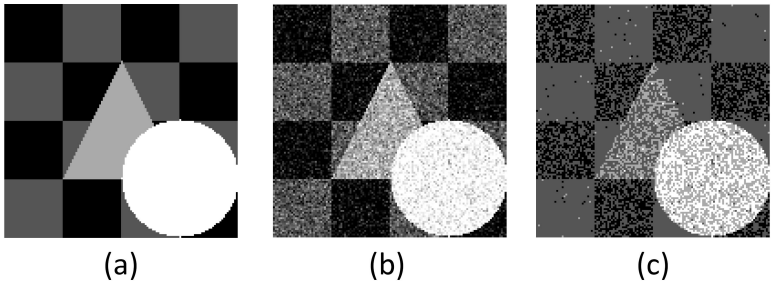

Figure 2.4: Synthetic image, (a): original image, (128x128 image) (b): Corrupted original image with Gaussian noise (0 mean, 0.02 variance), (c): standard GMM.

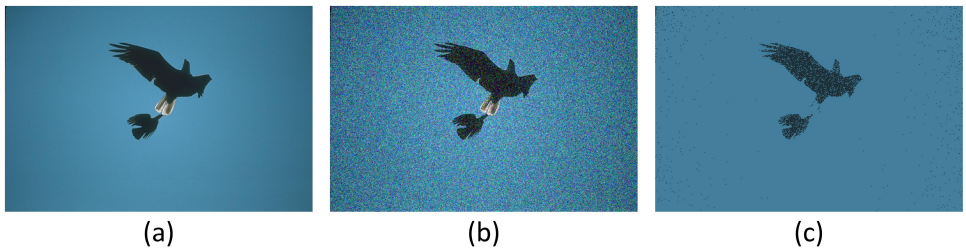

Figure 2.5: Real world image (321x481 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.02 variance), (c): standard GMM

Eq.(2.25), one of the biggest advantages of the standard GMM is that it has a simple form, and requires a small number of parameters.

However, the main drawback is that we cannot assign the same weight for every pixel belonging to the label Ω_j, as the pixels in the image vary in their intensity values and locations. Another limitation of the standard GMM is that the pixel x_i is considered to be an independent sample. Therefore, it does not take into account the spatial correlation between the neighboring pixels in the decision process. Segmentation is extremely sensitive to noise and illumination. The original image in Figure 2.4(a) is corrupted with a Gaussian noise (0 mean, 0.02 variance). The objective is to segment the image

in Figure 2.4(b) into four labels. Figure 2.4(c) show the segmentation results for standard GMM. We can see that some details are lost in the segmented image. Another image from the Berkeleys image segmentation dataset [98], as shown in Figure 2.5(a), is used in next experiment. The image shown in Figure 2.5(b) is derived from the original image by corrupting it with Gaussian noise (0 mean, 0.02 variance). The objective is to segment the noisy image into two labels. As can be seen, the segmentation accuracy of standard GMM method, along the object boundaries is quite poor.

2.3.2 Student's-t Mixture Model

To improve the robustness of the algorithm to outliers, Student's-t distribution has been used. The main advantage of the Student's-t distribution is that it is heavily tailed than Gaussian, and hence finite mixture model of the longertailed multivariate Student's-t distribution provides a much more robust approach to the standard GMM. It has proven to be quite effective for image segmentation [106]. In order to partition an image consisting of N pixels into K labels, standard Student's-t mixture model (SMM) assumes that each pixel x_i is independent of the label Ω_j. The density function at a pixel x_i is given by:

$$f(\mathrm{x}_i|\Pi,\Theta) = \sum_{j=1}^{K} \pi_j S(\mathrm{x}_i|\Theta_j) \tag{2.26}$$

where, $\Pi = \{\pi_j\}$; j=(1,2,...,K); is the set of prior distributions modeling the probability that pixel x_i is in label Ω_j, which satisfies the constraints in Eq.(2.22).

Each Student's-t distribution $S(\mathrm{x}_i|\Theta_j)$, called a component of the mixture, has its own parameters $\Theta_j = \{\mu_j, \Lambda_j, v_j\}$. The Student's-t distribution $S(\mathrm{x}_i|\Theta_j)$ is given by:

$$S(\mathrm{x}_i|\Theta_j) = \frac{\Gamma(v_j/2 + D/2)|\Lambda_j|^{1/2}}{\Gamma(v_j/2)(v_j\pi)^{D/2}}\left(1 + \frac{\Delta_j^2}{v_j}\right)^{-(v_j+D)/2} \tag{2.27}$$

where, $\Gamma(\cdot)$ is the Gamma function. And, Δ^2 is the squared Mahalanobis distance from pixel x_i to mean μ_j.

$$\Delta^2 = (\mathrm{x}_i - \mu_j)^{\mathrm{T}}\Lambda_j(\mathrm{x}_i - \mu_j) \tag{2.28}$$

The joint conditional density of the data set $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)$ is modeled as:

$$p(\mathrm{X}|\Pi, \Theta) = \prod_{i=1}^{N} f(\mathrm{x}_i|\Pi, \Theta) = \prod_{i=1}^{N}\sum_{j=1}^{K} \pi_j S(\mathrm{x}_i|\Theta_j) \tag{2.29}$$

Then, the log-likelihood function of the standard SMM is given by the following identity:

$$L(\Theta, \Pi|\mathrm{X}) = \log p(\mathrm{X}|\Pi, \Theta) = \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_j S(\mathrm{x}_i|\Theta_j)\right\} \tag{2.30}$$

Unfortunately, there is no closed form solution for maximizing the log-likelihood under a Student's-t distribution. To overcome this problem, the Student's-t distribution in previous SMM models [80] is represented as an infinite mixture of scaled Gaussians. In particular, we can write the Student's-t distribution in the form.

$$S(\mathrm{x}_i|\Theta_j) = \int_0^{\infty} \Phi(\mathrm{x}_i|\mu_j, u_j\Lambda_j)\mathcal{G}(u_j|v_j/2, v_j/2)du_j \tag{2.31}$$

where, $\Phi(\mathrm{x}_i|\mu_j, u_j\Lambda_j)$ denotes the Gaussian distribution, and $\mathcal{G}(u_j|v_j/2, v_j/2)$ is the Gamma distribution. As apparent, the representation of the Student's-t distribution as an infinite mixture of scaled Gaussians in Eq.(2.31) will correspond to an increase in complexity.

As shown from the log-likelihood function in Eq.(2.30), the pixel x_i in SMM is regarded as the same as x_i in GMM. Each pixel x_i is considered independent of its neighbors. The spatial correlation between the neighboring pixels is not taken into account in the decision process. Moreover, the prior distribution π_j does not depend on the pixel index and has the same value for all pixels.

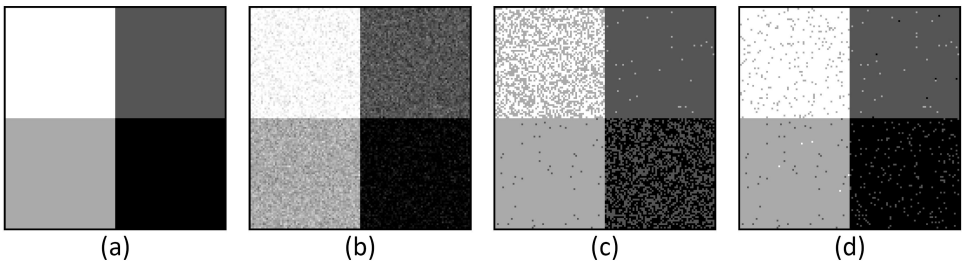

Figure 2.6: Synthetic image, (a): original image, (128x128 image) (b): Corrupted original image with Gaussian noise (0 mean, 0. 005 variance), (c): standard GMM, (c): standard SMM.

In Figure 2.6, we illustrate the performances of the standard GMM and SMM for image segmentation. An image of size 128x128 with four labels, as shown in Figure 2.6(a), is used in this example. Each square box in this image has a size of 64x64 pixel and has the same luminance value [0, 1/3, 2/3, 1]. The segmentation results of standard GMM, and standard SMM are shown Figure 2.6(c)–(d), respectively. In this example, compared with GMM

method, the standard SMM demonstrates a higher degree of robustness with respect to noise. However, as we see in Figure 2.6(d), the effect of noise on the result of standard SMM is still very high.

2.4 The Expectation Maximization (EM) Algorithm

2.4.1 EM Algorithm for the Gaussian Mixture Model

In order to maximize the likelihood function given in Eq.(2.25), we need to determine the parameters of the GMM. Various techniques [48], [76] have been previously developed to determine these parameters, based on maximizing their likelihood $L(\Theta, \Pi|\mathrm{X})$ in Eq.(2.25), for a given data set. In [46], the well-known EM algorithm is used to approximate the maximum likelihood.

Let us begin by setting the derivatives of $L(\Theta, \Pi|\mathrm{X})$ in Eq.(2.25) with respect to the means μ_j of the Gaussian components to zero. We obtain

$$\frac{\partial L(\Theta, \Pi|\mathrm{X})}{\partial \mu_j} = -\sum_{i=1}^{N} \underbrace{\frac{\pi_j \Phi(\mathrm{x}_i|\Theta_j)}{\sum_{l=1}^{K} \pi_l \Phi(\mathrm{x}_i|\Theta_l)}}_{z_{ij}} \frac{\mathrm{x}_i - \mu_j}{\sigma_j^2} = 0 \tag{2.32}$$

where we have made use of the form Eq.(2.23) for the Gaussian distribution. Note that the posterior probabilities z_{ij} appear naturally on the right-hand side:

$$z_{ij}^{(t)} = \frac{\pi_j^{(t)} \Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{l=1}^{K} \pi_l^{(t)} \Phi(\mathrm{x}_i|\Theta_l^{(t)})} \tag{2.33}$$

where t indicates the iteration step. The solution of $\partial L(\Theta, \Pi|\mathrm{X})/\partial \mu_j = 0$ yields the minimum of μ_j at the $(t+1)$ iteration step:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} \mathrm{x}_i}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{2.34}$$

If we set the derivative of $L(\Theta, \Pi|\mathrm{X})$ in Eq.(2.25) with respect to σ_j, and follow a similar line of reasoning, we obtain

$$[\sigma_j^2]^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} (\mathrm{x}_i - \mu_j^{(t+1)})^2}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{2.35}$$

Finally, we maximize $L(\Theta, \Pi|\mathrm{X})$ with respect to the prior distribution π_j. Here we must take account of the constraint in Eq.(2.22), which requires the prior distribution π_j to sum to one. This can be achieved by using a Lagrange multiplier η and maximizing the following quantity:

$$\frac{\partial}{\partial \pi_j} \left[L(\Theta, \Pi|\mathrm{X}) - \eta \left(\sum_{j=1}^{K} \pi_j - 1 \right) \right] = 0 \tag{2.36}$$

which gives

$$\sum_{i=1}^{N} \frac{\Phi(\mathrm{x}_i|\Theta_j)}{\sum_{l=1}^{K} \pi_l \Phi(\mathrm{x}_i|\Theta_l)} - \eta = 0 \tag{2.37}$$

If we now multiply both sides by π_j and use the constraint in Eq.(2.22), we find $\eta = N$. Using this to eliminate η and rearranging we obtain

$$\pi_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} z_{ij}^{(t)} \tag{2.38}$$

We summarize the EM algorithm for Gaussian mixture model below:

Step 1: Initialize the parameters $\Xi = \{\Theta, \Pi\} = \{\mu_j, \sigma_j, \pi_j\}$: the means μ_j,

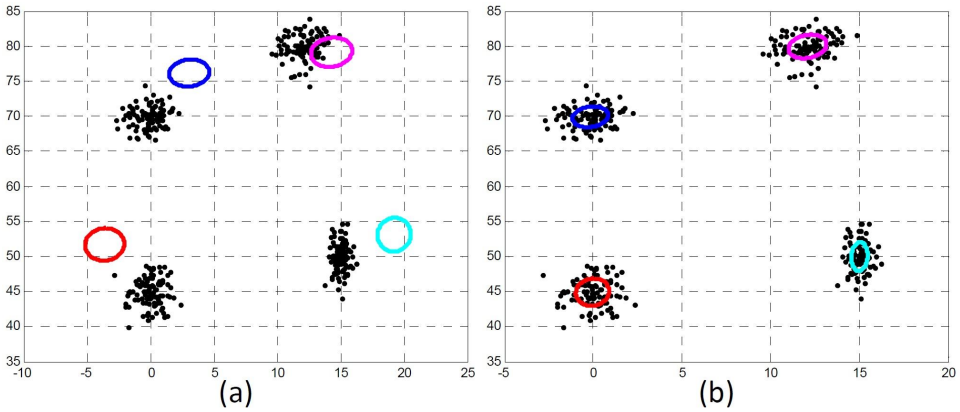

Figure 2.7: EM algorithm for the mixture of gaussians, (a): The original 2D point set with the initial condition, (b): Result of GMM with this initial condition.

covariance values σ_j and prior distributions π_j.

Step 2 (E-step): Evaluate the values z_{ij} in Eq.(2.33) using the current parameter values.

Step 3 (M-step): Re-estimate the parameters $\Xi = \{\Theta, \Pi\} = \{\mu_j, \sigma_j, \pi_j\}$.

+ Update the means μ_j by using Eq.(2.34).

+ Update covariance values σ_j by using Eq.(2.35).

+ Update prior distributions π_j by using Eq.(2.38).

Step 4: Evaluate the log-likelihood $L(\Theta, \Pi|\mathrm{X})$ in Eq.(2.25) and check the convergence of either the log-likelihood function, or the parameter values. If the convergence criterion is not satisfied, then go to step 2.

After optimizing the parameters of the GMM and determining the posterior probability z_{ij}, Eq.(2.33) is used to assign labels to each pixel in the image. The performance of EM algorithm for Gaussian mixture model is shown in Figure 2.7. In this example, four hundred simulated points are generated from

four multivariate Gaussian distributions. Each component has one hundred data points. In Figure 2.7(a), we show the initial condition for EM algorithm. As shown in Figure 2.7(b), GMM is likely to successfully classify the data points.

2.4.2 Relation between EM and K-means

According to the K-means [86], each pixel x_i in an image belongs to just one label. It is based on the minimization of the following objective function:

$$H = \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij} ||\mathrm{x}_i - \mu_j||^2 \tag{2.39}$$

The objective function in Eq.(2.39) represents the sum of the squares of the distances of each pixel to its assigned vector μ_j. The binary indicator variable r_{ij} is expressed as

$$r_{ij} = \begin{cases} 1 & \text{if } j = \arg\min_k ||\mathrm{x}_i - \mu_k||^2 \\ 0 & \text{otherwise} \end{cases} \tag{2.40}$$

The binary indicator variable r_{ij} in Eq.(2.40) describes which of the K labels the pixel x_i is assigned to. So that if data point x_i is assigned to label Ω_j then r_{ij}=1, and r_{ij}=0 for $j \neq k$. In Eq.(2.39), the term $||\mathrm{x}_i - \mu_j||^2$ expresses the similarity between the data and the mean. The optimum is reached when the mean μ_j of the lable Ω_j is found such that the objective function H in Eq.(2.39) is minimized.

Now consider the optimization of the μ_j with the r_{ij} held fixed. The objective function H is a quadratic function of μ_j. The gradient of the function

H with respect to parameters μ_j is given by:

$$\frac{\partial H}{\partial \mu_j} = -2 \sum_{i=1}^{N} r_{ij}(\mathrm{x}_i - \mu_j) \tag{2.41}$$

The solution of $\partial H / \partial \mu_j = 0$ yields the minimum of μ_j at the $(t+1)$ iteration step:

$$\mu_j = \frac{\sum_{i=1}^{N} r_{ij} \mathrm{x}_i}{\sum_{i=1}^{N} r_{ij}} \tag{2.42}$$

As an illustration in Eq.(2.42), we can see that one of the main advantages of K-means method is that it is very simple and easy to implement. Many researchers have used it in studying a number of key problems in image segmentation. However, from the objective function in Eq.(2.39), the pixel x_i in K-means is considered an independent sample, and thus this method does not take into account the spatial correlation between the neighboring pixels in the decision process. For that reason, the segmentation result of this method is very sensitive to noise.

Comparing the mathematical expressions of K-means with the EM algorithm for Gaussian mixtures, we see that there is a close similarity [79], [80]. As shown in Eq.(2.34), the EM algorithm makes a soft assignment based on the posterior probabilities z_{ij}. Whereas the K-means algorithm performs a hard assignment of data points to clusters based on the binary indicator variables r_{ij}, in which each data point is associated uniquely with just one label as shown in Eq.(2.42). In fact, we see that the K-means algorithm can be derived as a particular limit of EM for Gaussian mixture as follows.

Consider a Gaussian mixture in which the covariance matrices of the mixture components are given by $\varepsilon\mathbf{I}$, where $\mathbf{I}$ is the identity matrix

$$\Phi(\mathrm{x}_i|\Theta_j) = \frac{1}{\sqrt{2\pi\varepsilon}} \exp\left(-\frac{||\mathrm{x}_i - \mu_j||^2}{2\varepsilon}\right) \tag{2.43}$$

Applying the EM algorithm in subsection 2.4.1, the posterior probabilities z_{ij} for a particular data point x_i, are given by

$$z_{ij} = \frac{\pi_j \exp\{-||\mathrm{x}_i - \mu_j||^2/2\varepsilon\}}{\sum_{l=1}^{K} \pi_l \exp\{-||\mathrm{x}_i - \mu_l||^2/2\varepsilon\}} \tag{2.44}$$

If we consider the limit $\varepsilon \to 0$, we see that $z_{ij} \to r_{ij}$, where r_{ij} is defined by Eq.(2.40). And the EM estimation equation for the mean μ_j, given by Eq.(2.34), then reduces to the K-means result in Eq.(2.42). Note that the K-means algorithm only estimates the means but not the covariances of the labels. Finally, in the limit $\varepsilon \to 0$, the expected complete data log-likelihood [80], is given by

$$\mathbb{E}\left[L(\Theta, \Pi|\mathrm{X})\right] \to \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij}||\mathrm{x}_i - \mu_j||^2 + \text{const} \tag{2.45}$$

From Eq.(2.45), we observe that K-means is a special case of the EM algorithm for Gaussian mixtures.

2.5 Gradient-Based Optimization Techniques

In order to estimate the parameters of a Gaussian mixture model, we will consider another technique based on the gradient method [79] in this subsection.

First of all, the prior probabilities π_j corresponding to the label Ω_j are chosen:

$$\pi_j = \frac{\exp(\gamma_j)}{\sum_{k=1}^{K} \exp(\gamma_k)} \tag{2.46}$$

The transformation given by Eq.(2.46) is called the softmax function, or normalized exponential, and ensures that, for $-\infty \leq \gamma_j \leq \infty$, the constraints in Eq.(2.22) are satisfied as required for probabilities.

Given the density function in Eq.(2.21) and the prior probabilities π_j in Eq.(2.46), we need to optimize the parameter set $\Xi = \{\Theta, \Pi\} = \{\mu_j, \sigma_j, \gamma_j\}$ in order to maximize the log-likelihood function in Eq.(2.25). Since the logarithm is a monotonically increasing function, it is more convenient to consider the negative logarithm of the likelihood function [80], [82], [83] as an error function:

$$J(\Theta, \Pi|\mathrm{X}) = -L(\Theta, \Pi|\mathrm{X}) = -\sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_j \Phi(\mathrm{x}_i|\Theta_j) \right\} \tag{2.47}$$

Applying the complete data [79], [80], [82], [83], minimizing the negative log-likelihood function in Eq.(2.47), is equivalent to minimizing the error function $E(\Xi^{(t)}|\Xi^{(t+1)})$:

$$E(\Xi^{(t)}|\Xi^{(t+1)}) = -\sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \log \left\{ \pi_{ij}^{(t+1)} \Phi(\mathrm{x}_i|\Theta_j^{(t+1)}) \right\} \tag{2.48}$$

where

$$z_{ij}^{(t)} = \frac{\pi_j^{(t)} \Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{l=1}^{K} \pi_l^{(t)} \Phi(\mathrm{x}_i|\Theta_l^{(t)})}$$

is the posterior probability at the iteration of the current step as shown in Eq.(2.33).

To minimize this error function, we apply the gradient descent algorithm [79] to adjust the parameters $\Xi = \{\Theta, \Pi\} = \{\mu_j, \sigma_j, \gamma_j\}$. The change in the parameters is then given by:

$$\Xi_{new} = \Xi_{old} - \eta \nabla E\left(\Xi_{old}\right) \tag{2.49}$$

where, $\nabla E(\Xi) = (\partial E/\partial \mu_j, \partial E/\partial \sigma_j, \partial E/\partial \gamma_j)$, η is the learning rate and its value is sufficiently small. The gradient of this error function $E(\Xi)$ with respect to parameters μ_j is given by:

$$\frac{\partial E}{\partial \mu_j} = -\sum_{i=1}^{N} z_{ij}^{(t)} \frac{\mathrm{x}_i - \mu_j}{\sigma_j^2} \tag{2.50}$$

Similarly, the derivative of $E(\Xi)$ with respect to σ_j is given by

$$\frac{\partial E}{\partial \sigma_j} = -\sum_{i=1}^{N} z_{ij}^{(t)} \left(-\frac{1}{\sigma_j} + \frac{\left(\mathrm{x}_i - \mu_j\right)^2}{\sigma_j^3} \right) \tag{2.51}$$

The derivative of $E(\Xi)$ with respect to γ_j is expressed as:

$$\frac{\partial E}{\partial \gamma_j} = -\sum_{i=1}^{N} \left(z_{ij}^{(t)} - \pi_j^{(t)} \right) \tag{2.52}$$

We summarize the gradient-based optimization techniques for Gaussian mixture model below:

Step 1: Initialize the parameters $\Xi = \{\mu_j, \sigma_j, \gamma_j\}$: the means μ_j, covariance values σ_j and the value of γ_j.

Step 2: Evaluate the values z_{ij} in Eq.(2.33) using the current parameter values.

Step 3: Re-estimate the parameters $\Xi = \{\mu_j, \sigma_j, \gamma_j\}$ by using Eq.(2.49).

Step 4: Evaluate the log-likelihood $L(\Theta, \Pi|\mathrm{X})$ in Eq.(2.25) and check the convergence of either the log-likelihood function, or the parameter values. If the convergence criterion is not satisfied, then go to step 2.

2.6 Image Segmentation Evaluation

Once the parameter-learning phase is complete, in order to assign labels to each pixel, the posterior probability z_{ij} is used. For each pixel x_i, given the posterior probability z_{ij} for all labels, in order to segment an image consisting of N pixels into K labels, a determination is made, whereby the pixelx_i is assigned to the label with the largest posterior probability:

$$\mathrm{x}_i \in \Omega_j : \text{IF } z_{ij} \geq z_{ik};\ j,k = 1,2,...,K \tag{2.53}$$

In order to evaluate the segmentation performance quantitatively, the misclassification ratio (MCR) [91] is employed:

$$\text{MCR} = \frac{\text{number of misclassfied pixels}}{\text{total number ofpixels}} \times 100 \tag{2.54}$$

The value of MCR is in the [0–100] range, where lower values indicate better segmentation results.

Another technique to obtain an objective performance evaluation is to adopt the probabilistic rand (PR) index [93]. For each image, the multiple ground truths are available and are denoted as $\mathrm{G} = \{G_1, G_2, ..., G_M\}$. The segmentation map under evaluation is denoted as G_{eval}. The PR index is given by:

$$\text{PR}(\mathrm{G}, \mathrm{G}_{eval}) = \frac{2}{M(M-1)} \sum_i \sum_{j>i} [c_{ij}p_{ij} + (1-c_{ij})(1-p_{ij})] \tag{2.55}$$

where M is the number of image pixels. p_{ij} is the ground truth probability that pixels i and j belong to the same segment. If pixels i and j belong to the same segment in G_{eval}, the value of c_{ij} is one. Otherwise, its value is zero.

The PR index takes a value in the interval [0–1]. A score of zero indicates a bad segmentation where every pixel pair in the test image has the opposite relationship as every pair in the ground truth segmentation. Otherwise, a score of one indicates a good result where every pixel pair in the test image has the same relationship as every pair in the ground truth segmentation.

In order to quantify the overlap between the segmented image and the ground truth for the label Ω_j, the Dice similarity coefficient [94] is used:

$$\text{Dice}_j = \frac{2V_{ab}^j}{(V_a^j + V_b^j)} \times 100 \tag{2.56}$$

Where, V_{ab}^j denotes the number of pixels that are assigned to label Ω_j by both the segmented image and ground truth. The number of pixels assigned to Ω_j by the segmented image and the ground truth are denoted by V_a^j and V_b^j, respectively. The Dice index attains the value in the [0–100] range, where higher values indicate better segmentation results.

2.7 Conclusions

In this chapter, various criteria evaluations for unsupervised segmentation algorithms are addressed. Besides that, the standard GMM for image segmentation is presented. This mixture model is a well-known method that has been widely used as a tool for image segmentation. Its success is attributed to the fact that the model parameters can be efficiently estimated by adopting various techniques such as EM algorithm, or gradient-based optimization techniques. Other advantages are its simplicity and ease of implementation.

However, the major disadvantage of GMM is that the model assumes that each pixel is independent of its neighbors. It is well known that pixels in an image are similar in some sense and cannot be classified consistently based on feature attributes alone. Thus, the segmentation result of GMM is extremely sensitive to noise.

Chapter 3

Gaussian Mixture Model based Markov Random Field

3.1 Introduction

As mentioned in Chapter 2, a major shortcoming of standard GMM is that it does not take into account the spatial dependencies in the image. Moreover, it does not use the prior knowledge that adjacent pixels most likely belong to the same cluster. In this family of model-based techniques, prior probabilities [82] of label membership are considered constant for every pixel of an image. Thus, the performance of these methods is too sensitive to noise and image contrast levels.

A possible approach to overcome this problem is to impose spatial smoothness constraints to incorporate the spatial relationships between neighboring pixels [59], [60], [61]. Several mixture models based on MRF for pixel

labels are proposed in [59]–[64], [89], [90]. According to these approaches, prior probabilities are based on MRF to capture spatial information. The primary advantage of this family of mixture models is that it incorporates spatial information. Hence, it improves segmentation results, particularly when an image is corrupted by high levels of noise.

Another family of mixture models based on MRF for pixel label priors have been successfully applied to image segmentation [58], [72]–[74], [100]. Instead of imposing the smoothness constraint on the pixel label as in the above category, however, these methods aim to impose the smoothness constraint on the contextual mixing proportions. Their primary disadvantage, however, lies in its additional training complexity. The M-step of the EM algorithm in [72]–[74] cannot evaluate the prior distribution in a closed form, which therefore corresponds to an increase in the algorithm's complexity. In [58] the gradient projection step was proposed to implement the M-step. Another reparatory projection step based on a closed form update equation was introduced [72] to guarantee that the prior probabilities are positive and sum to one.

3.2 Gaussian Mixture Model based MRF for the Pixel Labels

3.2.1 Markov Random Field Theory

Markov random field (MRF) is a probability theory that provides a stochastic mathematical framework for analyzing the spatial constraint in an image. Let

$\mathcal{S}$, $\mathcal{S} = (1, 2, ..., N)$, index a discrete set of N sites (pixels). And $\mathcal{L}$, $\mathcal{L} = (1, 2, ..., K)$ is a set of label. For every site $i \in \mathcal{S}$, we consider a finite space $\mathcal{Z}_i$ of states z_i, such as $\mathcal{Z}_i = (\mathrm{z}_i, \mathrm{z}_i \in \mathcal{L})$. The space of the configurations of the state values of the considered sites set is denoted by the product space.

$$\mathcal{Z} = \prod_{i=1}^{N} \mathcal{Z}_i \tag{3.1}$$

Then, the $p(\mathrm{Z})$ is a random field, if the following condition is satisfied

$$p(\mathrm{Z}) > 0, \ \mathrm{Z} \in \mathcal{Z} \tag{3.2}$$

Now, let us denote the neighborhood of site i as $\mathcal{N}_i$, $i \notin \mathcal{N}_i$ and $i \in \mathcal{N}_j \Leftrightarrow j \in \mathcal{N}_i$. The neighborhood system on $\mathcal{S}$ is defined as $\mathcal{N} = (\mathcal{N}_i, i \in \mathcal{S})$. Then, the previously considered random field $p(\mathrm{Z})$ is an MRF with respect to a neighborhood system $\mathcal{N}$ [91], [92] if and only if

$$p(\mathrm{z}_i|\mathrm{z}_{\mathcal{S}-\{i\}}) = p(\mathrm{z}_i|\mathrm{z}_{\mathcal{N}_i}) \tag{3.3}$$

Hammersley-Clifford theorem is proposed in [95], [96] to establish the equivalence of MRF and Gibbs random field. According to this theorem, a Gibbs distribution is equivalently characterized by a MRF and vice versa. Thus, an MRF given in Eq.(3.3) is rewritten as:

$$p(\mathrm{Z}|\beta) = W^{-1} \exp\left(-U(\mathrm{Z}|\beta)\right) \tag{3.4}$$

where, W is a normalizing constant called the partition function

$$W(\beta) = \sum_{\mathrm{z} \in \mathcal{Z}} \exp(-U(\mathrm{Z}|\beta)) \tag{3.5}$$

and $U(\mathrm{Z}|\beta)$ is an energy function. This energy function is a sum of clique potentials over all possible cliques of the form.

$$U(\mathrm{Z}|\beta) = \sum_{c \in C} V_c(\mathrm{Z}|\beta) \tag{3.6}$$

V_c stands for the clique potential associated with the clique c. And C is a label of subsets of the sites that contains sites that are all neighbors, and are known as cliques. β is a parameter of the clique potentials known as the inverse supercritical temperature. The computation of the term W in Eq.(3.5) involves all possible realizations Z of the MRF which is hardly ever feasible, in terms of computational requirements. To overcome this problem, an approximation of the likelihood in Eq.(3.4) is the pseudo-likelihood introduced by Besag [97] and defined as

$$p(\mathrm{Z}|\beta) = \prod_{j=1}^{N} p(\mathrm{z}_i|\mathrm{z}_{\mathcal{N}_i}; \beta) \tag{3.7}$$

where, each term in the product is to compute

$$p(\mathrm{z}_i|\mathrm{z}_{\mathcal{N}_i}; \beta) = \frac{\exp\left(-\sum_{ci} V_c(\mathrm{Z}|\beta)\right)}{\sum_{\mathrm{z}_i} \exp\left(-\sum_{ci} V_c(\mathrm{Z}|\beta)\right)} \tag{3.8}$$

The probability distribution in Eq.(3.8) is used to obtain estimates of a Markov random field parameters.

3.2.2 Hidden Markov models

Let us consider the problem of segmenting an image with N pixels, $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)$, into K labels. In this model, the observations X are conditionally independent given Z. According to the GMM based MRF for the pixel

label [59], [63], [64], [65], the density function at an observation x_i is given by:

$$f(\mathrm{x}_i|\Pi,\Theta) = \sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j) \tag{3.9}$$

and the prior distributions π_{ij} are defined by:

$$\pi_{ij} = p(\mathrm{z}_i|\mathrm{z}_{\mathcal{N}_i};\beta) \tag{3.10}$$

where, the prior π_{ij} is different for each pixel i and depends on the neighbors of the pixel. For more details, please refer to [64], [89]. Thus far, the problem has focused on how to estimate the parameters to maximize the following log-likelihood function:

$$L(\Theta,\Pi|\mathrm{X}) = \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j)\right\} \tag{3.11}$$

The iterative EM algorithm for estimating the parameters of the component densities is applied to estimate the parameters. The conditional expectation values z_{ij} of the hidden variables is computed as follows:

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)}\Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{k=1}^{K} \pi_{ik}^{(t)}\Phi(\mathrm{x}_i|\Theta_k^{(t)})} \tag{3.12}$$

The estimates of the means μ_j and covariance matrices Σ_j yield:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)}\mathrm{x}_i}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{3.13}$$

and,

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)}(\mathrm{x}_i-\mu_j)(\mathrm{x}_i-\mu_j)^T}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{3.14}$$

Finally, the estimate of the inverse temperature parameter β yields

$$\beta^{(t+1)} = \arg\max_{\beta} \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \log p(\mathrm{z}_i = j | \mathrm{z}_{\mathcal{N}_i}^{(t)}; \beta) \tag{3.15}$$

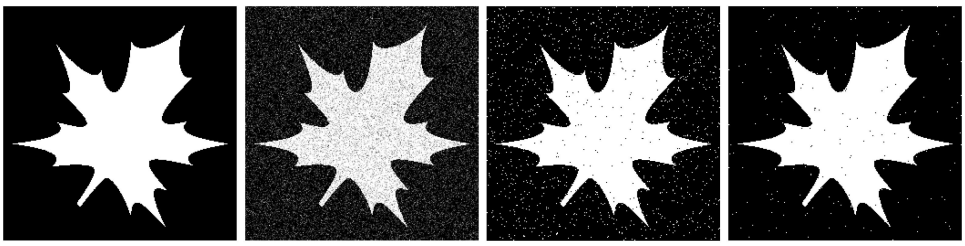

Figure 3.1: Synthetic image, (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.05 variance), (c): standard GMM, (d): SIMF [63].

From the log-likelihood function in Eq.(3.11), we can see that the main advantage of the GMM based MRF for the pixel labels is that it incorporates spatial dependencies between pixels. Compared to the standard GMM, this approach works well in noisy image segmentation. A synthetic image, shown in Figure 3.1(a), was used to test the effectiveness of this model. The objective is to segment the image into two labels. The image shown in Figure 3.1(b) is made from the original image by corrupting with Gaussian noise. In Figure 3.1(c) and Figure 3.1(d), we present the segmentation results obtained by employing standard GMM and SIMF [63] methods, respectively. As can be easily seen, the effect of noise on the performance of the SIMF is much less as compared to the standard GMM.

3.3 Gaussian Mixture Model based MRF for the Priors of the Pixel Labels

In order to reduce the sensitivity of the segmentation result with respect to noise, several researchers have suggested modifications to incorporate the local spatial interactions between the neighboring pixels. In the models in [58], [72]–[74]. the pixel label priors are treated as random variables forming an MRF have been presented in. In [58], the authors proposed a spatially variant finite mixture model (SVFMM) for image segmentation. The model assumes that the density function at an observation x_i is given by:

$$f(\mathrm{x}_i|\Pi,\Theta) = \sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j) \tag{3.16}$$

where the Gaussian distribution $\Phi(\mathrm{x}_i|\Theta_j)$ is the same as Eq.(2.23). The prior distribution π_{ij} of the pixel x_i belonging to the label Ω_j should satisfy the following constraints:

$$0 \leq \pi_{ij} \leq 1 \text{ and } \sum_{j=1}^{K} \pi_{ij} = 1 \tag{3.17}$$

Note that the observation x_i in Eq.(3.16) is modeled as statistically independent of the label Ω_j. The joint conditional density [72] of the data set $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)$ can be modeled as:

$$p(\mathrm{X}|\Pi,\Theta) = \prod_{i=1}^{N} f(\mathrm{x}_i|\Pi,\Theta) = \prod_{i=1}^{N} \left[\sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j) \right] \tag{3.18}$$

Since the observation x_i is considered to be independent given the pixel label, the spatial correlation between the neighboring pixels is not taken into account.

As a result, the segmented image is sensitive to noise, varying illumination and other environmental factors such as wind, rain or camera movements. To overcome this problem, MRF distribution [99] is applied to incorporate the spatial correlation amongst label values:

$$p(\Pi) = W^{-1} \exp\left\{-\frac{1}{T}U(\Pi)\right\} \tag{3.19}$$

where, W is a normalizing constant, T is a temperature constant. And $U(\Pi)$ is the smoothing prior. The posterior probability density function given by Bayes' rules can be written as:

$$p(\Pi, \Theta|\mathrm{X}) \propto p(\mathrm{X}|\Pi, \Theta)p(\Pi) \tag{3.20}$$

By incorporating Eq.(3.20), the log-likelihood function can be derived as:

$$\begin{aligned} L(\Pi, \Theta|\mathrm{X}) &= \log\left(p(\Pi, \Theta|\mathrm{X})\right) \\ &= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j)\right\} + \log p(\Pi) \\ &= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j)\right\} - \log W - \frac{1}{T}U(\Pi) \end{aligned} \tag{3.21}$$

Depending on the type of energy $U(\Pi)$ selected, we can have different kinds of models. In SVFMM method in [58], the value of T is set to one (T=1), and the Gibbs function for the priors $p(\Pi)$ is given by:

$$p(\Pi) = \frac{1}{Z}\exp\left(-U(\Pi)\right); \text{ where} : U(\Pi) = \beta \sum_{i=1}^{N}\sum_{j=1}^{K}\sum_{m\in\mathcal{N}_i} \left(\pi_{ij} - \pi_{mj}\right)^2 \tag{3.22}$$

where, Π is the parameter set; $\Pi = \{\pi_{ij}\}$; i=1,2,...,N; j=1,2,...,K. And β is a scalar. The log-likelihood function (ignore the constant) is given by:

$$\begin{aligned} L(\Pi, \Theta|\mathrm{X}) &= \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j) \right\} + \log p(\Pi) \\ &= \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j) \right\} - \beta \sum_{i=1}^{N} \sum_{j=1}^{K} \sum_{m \in \mathcal{N}_i} (\pi_{ij} - \pi_{mj})^2 \end{aligned} \tag{3.23}$$

Compared to the log-likelihood function of the standard GMM in Eq.(2.25), the log-likelihood function in Eq.(3.23) is quite complex. In order to maximize this likelihood with respect to the parameters $\Xi = \{\Theta, \Pi\} = (\mu_j, \Sigma_j, \pi_{ij})$, an iterative EM algorithm is adopted. Application of the complete data condition in [58], maximizing the log-likelihood function $L(\Pi, \Theta|\mathrm{X})$ in Eq.(3.23) will lead to an increase in the value of the objective function $H(\Pi, \Theta|\mathrm{X})$.

$$H(\Pi, \Theta|\mathrm{X}) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij} + \log \Phi(\mathrm{x}_i|\Theta_j) \right\} - \beta \sum_{i=1}^{N} \sum_{j=1}^{K} \sum_{m \in \mathcal{N}_i} (\pi_{ij} - \pi_{mj})^2 \tag{3.24}$$

where the conditional expectation values z_{ij} of the hidden variables can be computed as follows:

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{k=1}^{K} \pi_{ik}^{(t)} \Phi(\mathrm{x}_i|\Theta_k^{(t)})} \tag{3.25}$$

Let us now consider the derivation of the function $H(\Pi, \Theta|\mathrm{X})$ with the means μ_j at the (t+1) iteration step. We have:

$$\frac{\partial H}{\partial \mu_j} = \sum_{i=1}^{N} z_{ij}^{(t)} \left[-\frac{1}{2} (2\Sigma_j^{-1} \mu_j - 2\Sigma_j^{-1} \mathrm{x}_i) \right] \tag{3.26}$$

The solution of $\partial J/\partial \mu_j = 0$ yields the minimizer of μ_j at the $(t+1)$ step:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} \mathrm{x}_i}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{3.27}$$

Thus, setting the derivative of the function in Eq.(3.24) with respect to Σ_j^{-1} at the $(t+1)$ iteration step we have:

$$\frac{\partial H}{\partial \Sigma_j^{-1}} = \sum_{i=1}^{N} z_{ij}^{(t)} \left[\frac{1}{2}\Sigma_j - \frac{1}{2}(\mathrm{x}_i - \mu_j)(\mathrm{x}_i - \mu_j)^T \right] \tag{3.28}$$

and, equating it to zero yields:

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} (\mathrm{x}_i - \mu_j)(\mathrm{x}_i - \mu_j)^T}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{3.29}$$

However, due to the complexity of the log-likelihood function in Eq.(3.23), the M-step of EM algorithm cannot evaluate the prior distribution π_{ij} in a closed form. In order to maximize objective function $H(\Pi, \Theta|\mathrm{X})$ with respect π_{ij} [58], [72], we set its derivative equal to zero and obtain the following quadratic expression:

$$4\beta N_i \left(\pi_{ij}^{(t+1)}\right)^2 - 4\beta \pi_{ij}^{(t+1)} \sum_{m \in \mathcal{N}_i} \pi_{mj} - z_{ij}^{(t)} = 0 \tag{3.30}$$

where N_i stands for the set of neighbors falling in a window around the pixel x_i. The above equation has two roots:

$$\pi_{ij}^{(t+1)} = \frac{\sum_{m \in \mathcal{N}_i} \pi_{mj} \pm \sqrt{\left(\sum_{m \in \mathcal{N}_i} \pi_{mj}\right)^2 + \frac{N_i}{\beta} z_{ij}^{(t)}}}{2N_i} \tag{3.31}$$

We select the root with the positive sign + since it yields $\pi_{ij} \geq 0$. The above equation provides a straightforward update for the values of label parameters

π_{ij} of each pixel i at the M-step of every EM iteration. However, note that the prior distribution π_{ij} should satisfy the constraints in Eq.(3.17). Therefore, the algorithm becomes even more computationally complex. In [58], [72], [73], a large amount of computational power is utilized to solve the constrained optimization problem of the prior distribution π_{ij}.

To illustrate the computational cost of this approach, an image (245x245 image resolution) with three labels as shown in Figure 3.2(a) is used. The image shown in Figure 3.2(b) is made from the original image by corrupting with Gaussian noise. All methods are initialized with the same initial condition and are performed on a PC (Core i3 with 4GB RAM) until convergence by using MATLAB in the Windows environment. As shown in Figure 3.2(d), although, SVFMM [72] demonstrates a higher degree of robustness with respect to noise, it is still low in terms of the speed. Note that SVFMM takes 246.1 seconds to segment this image. Compared to SVFMM, standard GMM is fast (5.5 seconds). However, the segmentation accuracy of standard GMM is quite poor as shown in Figure 3.2(c).

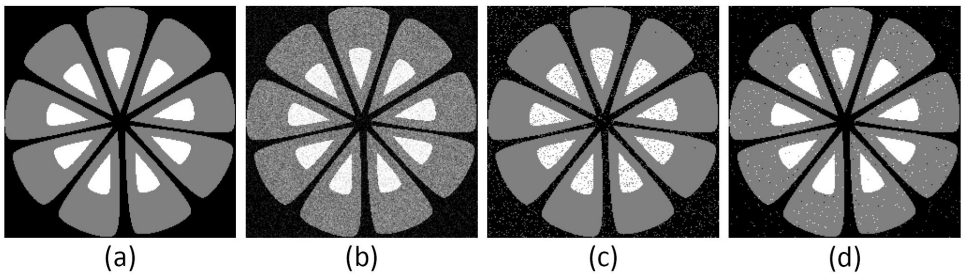

Figure 3.2: Synthetic image, (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.01 variance), (c): standard GMM (time = 5.5s), (d): SVFMM [72] (time = 246.1s).

Another limitation of this model, as mentioned in [58], is that it requires a greater number of parameters compared to the standard GMM. In order to segment an image consisting of N pixels into K labels, we have to deal with Kx(2+N) parameters (K parameters of μ_j, K parameters of Σ_j and NK parameters of π_{ij}). This implies that the larger the image, the more the number of parameters that we have to estimate.

3.4 Conclusions

In this chapter, mixture models based on the Markov random fields are presented. Compared with the standards GMM, the major difference is that instead of using the common prior distribution π_j for all pixels, the prior distribution π_{ij} of the mixture models based on the Markov random fields are different for each pixel and depends on the neighbors of the pixel of interest and the corresponding parameters. The spatial relationship between neighboring pixels is taken into account. Although the effect of noise on the final segmentation result is reduced, these mixture models lack enough robustness with respect to noise. Besides that, they are too complex, and require a large number of parameters compared to the standard GMM.

Chapter 4

An Extension of the Standard Mixture Model for Image Segmentation

4.1 Introduction

The key concept in the field of image segmentation is the imposing of spatial smoothness. Indeed, methods that do not utilize smoothing for image segmentation, such as standard GMM, standard SMM, K-means lead to very poor results. In the standard GMM method, it can be easily seen that the spatial relationship between neighboring pixels is not taken into account. For this reason, although the standard GMM is a well known and simple method for image segmentation, its segmentation result is sensitive to noise, varying illumination and other environmental factors such as wind, rain or camera

shaking.

In this chapter, we propose a new model based on the standard GMM that applies to the image classification problem. Our approach differs from those discussed above by the following statements. Firstly, a unique approach accounting for the relationship amongst neighboring pixels is presented. The proposed model is quite similar to the standard GMM and thus, is easy to implement. Secondly, compared to the standard GMM, the main difference in the proposed method is that the prior distribution π_{ij} of each label Ω_j is different for each pixel x_i and depends on its neighboring pixels. Thirdly, compared to the models based on MRF, the proposed model is simple and requires fewer parameters.

The rest of this chapter is organized as follows. In section 4.2, we describe the details of the proposed algorithm. Learning algorithms for the proposed system are presented in section 4.3. In section 4.4, we present the experimental results and conclude with a discussion in section 4.5.

4.2 Proposed Method

First, we define a function that represents the weight of each i-th pixel for each label Ω_j.

$$\xi_j(\mathrm{x}_i) = \exp\left(-\frac{(\mathrm{x}_i - c_j)^2}{2b_j^2}\right) \tag{4.1}$$

where c_j and b_j; j=1,2,...,K; are parameters whose optimal values can be obtained by utilizing the methodology presented in the following section. For the neighborhood $\mathcal{N}_i$ of the i-th pixel, the weight function for each label Ω_j is

defined as:

$$\vartheta_j(\mathrm{x}_i) = \left[\sum_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2} \right) \right]^{\alpha} \tag{4.2}$$

where, α is a parameter, and $\mathcal{N}_i$ is the neighborhood of the i-th pixel (a 5x5 window is used in this method). Next, we propose a novel approach to incorporate the spatial relationships between neighboring pixels into the prior probability distribution π_{ij}. This prior distribution has different values for each pixel corresponding to each label Ω_j in the image, given by:

$$\pi_{ij} = \frac{\vartheta_j(\mathrm{x}_i)}{\sum_{k=1}^{K} \vartheta_k(\mathrm{x}_i)} \tag{4.3}$$

The prior probability π_{ij} in Eq.(4.3) is computed subject to the constraints $0 \leq \pi_{ij} \leq 1$ and $\sum_{j=1}^{K} \pi_{ij} = 1$. The density function at an observation x_i is given by:

$$f(\mathrm{x}_i|\Pi, \Theta) = \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j) \tag{4.4}$$

The log-likelihood function is given by:

$$L(\Theta, \Pi|\mathrm{X}) = \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j) \right\} \tag{4.5}$$

When $\pi_{ij} = \pi_{cj}$; $\forall i, c = 1, 2, ..., N$; the log-likelihood function $L(\Theta, \Pi|\mathrm{X})$ in Eq.(4.5) is the same as that given in Eq.(2.25). Therefore, the standard GMM is a special case of the proposed method. Now, if we compare the log-likelihood function of the proposed method in Eq.(4.5) with the log-likelihood function in Eq.(3.11), we can see that it is very similar to the Gaussian mixture model based MRF for the pixel labels in Section 3.2 with two main differences. First, while the above-mentioned methods introduce a mean-field approximation of

the MRF for π_{ij}, we propose a new way to incorporate the spatial relationships between neighboring pixels into the prior distribution π_{ij}, as shown in Eq.(4.3). Considering the formulae in Eq.(4.2) and Eq.(4.3), it can be easily seen that the prior distribution π_{ij} in the proposed method acts like a mean filter. For that reason, the image segmentation result obtained by employing the proposed method is robust with respect to noise. We also propose a new method to estimate the model parameters in order to minimize the higher bound on the data negative log-likelihood [83], [82], [7], [77], [78], based on the gradient method that offers a closed form M-step, with computational complexity similar to that of the M-step for Gaussian mixture model.

The next objective is to optimize the parameter set $\Xi = \{\Theta, \Pi\} = \{\mu_j, \sigma_j, c_j, b_j, \alpha\}$ to maximize the log-likelihood function in Eq.(4.5). Since the logarithm is a monotonically increasing function, it is more convenient to consider the negative logarithm of the likelihood function [79], [80] as an error function:

$$J(\Theta, \Pi|\mathrm{X}) = -L(\Theta, \Pi|\mathrm{X}) = -\sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j) \right\} \tag{4.6}$$

Applying the complete data in [46], with proper replacements for old parameter (at t iteration step) values with the new ones (at t+1 iteration step) in Eq.(4.6), the change in the error function can be expressed as:

$$J(\Theta^{(t+1)}, \Pi^{(t+1)}|\mathrm{X}) - J(\Theta^{(t)}, \Pi^{(t)}|\mathrm{X}) = -\sum_{i=1}^{N} \log \left\{ \frac{\sum_{j=1}^{K} \pi_{ij}^{(t+1)} \Phi(\mathrm{x}_i|\Theta_j^{(t+1)})}{\sum_{k=1}^{K} \pi_{ik}^{(t)} \Phi(\mathrm{x}_i|\Theta_k^{(t)})} \times \frac{z_{ij}^{(t)}}{z_{ij}^{(t)}} \right\} \tag{4.7}$$

Note, that $z_{ij}^{(t)}$, as shown in Eq.(2.33), always satisfies the conditions: $z_{ij}^{(t)} \geq 0$ and $\sum_{j=1}^{K} z_{ij}^{(t)} = 1$. We can now apply the Jensen's inequality [81] which states that, given a set of numbers $\lambda_j \geq 0$ and $\sum_{j=1}^{K} \lambda_j = 1$, we have:

$$\log\left(\sum_{j=1}^{K} \lambda_j y_j\right) \geq \sum_{j=1}^{K} \lambda_j \log\left(y_j\right) \tag{4.8}$$

From Eq.(4.8), the change in error function in Eq.(4.7) is given by:

$$J(\Theta^{(t+1)}, \Pi^{(t+1)}|\mathrm{X}) - J(\Theta^{(t)}, \Pi^{(t)}|\mathrm{X}) \leq -\sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \log\left\{\frac{\pi_{ij}^{(t+1)}\Phi(\mathrm{x}_i|\Theta_j^{(t+1)})}{z_{ij}^{(t)}\sum\limits_{k=1}^{K}\pi_{ik}^{(t)}\Phi(\mathrm{x}_i|\Theta_k^{(t+1)})}\right\} \tag{4.9}$$

Thus, we have to minimize the log-likelihood function with respect to the new parameters (at the t+1 iteration step). Therefore, we can drop the terms that depend only on the old parameters (at the t iteration step). The change in the error function can be written in the form:

$$E(\Theta^{(t)}, \Pi^{(t)}|\Theta^{(t+1)}, \Pi^{(t+1)}) = -\sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \log\left\{\pi_{ij}^{(t+1)}\Phi(\mathrm{x}_i|\Theta_j^{(t+1)})\right\} \tag{4.10}$$

The E in Eq.(4.10) can be regarded as an error function. Therefore, maximizing the likelihood L in Eq.(4.5) is then equivalent to minimizing E in Eq.(4.10). The minimization of the error function E with respect to the parameters $\Xi = \{\Theta, \Pi\} = \{\mu_j, \sigma_j, c_j, b_j, \alpha\}$; j=1,2,...,K; will be discussed in detail in the next part. In order to assign labels to each pixel, the posterior probability z_{ij} is used:

$$z_{ij} = \frac{\pi_{ij}\Phi(\mathrm{x}_i|\Theta_j)}{\sum\limits_{k=1}^{K}\pi_{ik}\Phi(\mathrm{x}_i|\Theta_k)} \tag{4.11}$$

After labeling each pixel using Eq.(4.11), a simple competitive selection is carried out for each pixel, in order to remove the remaining noise. If the i-th pixel belongs to label j; j=1,2,...,K; and all its neighborhood pixels $\mathcal{N}_i$ belong to the label k, $k \neq j$ and k=1,2,...,K, then the i-th pixel is set to the label k. It is worth noticing that we only use one type of selection criteria in this method: if all the neighbors of a given pixel are assigned to a specific label, the considered pixel is also assigned to the same label. The effect of this simple competitive selection employed in the proposed method is shown in the section containing the experimental results.

4.3 Parameter Learning

Thus far, the discussion has focused on probability estimation used to determine the label Ω_j to which the pixel x_i should be assigned. To generalize the posterior probability z_{ij}, we need to adjust the parameters $\Xi = \{\Theta, \Pi\} = (\mu_j, \sigma_j, c_j, b_j, \alpha)$; j=1,2,...,K; to minimize the error function E in Eq.(4.10), corresponding to maximizing log-likelihood function L in Eq.(4.5). Note, that the total number of parameters required for the proposed method is only $4K+1$ (K parameters of μ_j, K parameters of σ_j, K parameters of c_j, K parameters of b_j and 1 parameter of α), which is less than the number of parameters in the models based on MRF mentioned in the section 3.3.

In this part, instead of utilizing EM algorithm, we employ the gradient method [82], [83], [79], [80] for adjusting the parameters to minimize the error function E. The proposed algorithm can be summarized as follows:

Step 1: Initialize the parameters Ξ by the following sub-steps:

+ Use K-means to initialize the mean μ_j, and covariance σ_j. Then, select $c_j = \mu_j$, and $b_j = \sigma_j$. In this method, the initial value of α is set to 14.

+ While $t \leq T$, (T=5 in this method), repeat the following sub-steps:

  - Calculate the Gaussian distribution $\Phi(\mathrm{x}_i|\Theta_j^{(t)})$ from Eq.(2.23) and the prior probability distribution $\pi_{ij}^{(t)}$ from Eq.(4.3). Next, calculate the posterior probability $z_{ij}^{(t)}$, given in Eq.(4.11):

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)}\Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{k=1}^{K}\pi_{ik}^{(t)}\Phi(\mathrm{x}_i|\Theta_k^{(t)})}$$

  - Update the parameters μ_j and σ_j using Eq.(2.34) and (2.35):

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)}\mathrm{x}_i}{\sum_{i=1}^{N} z_{ij}^{(t)}}; \quad [\sigma_j^2]^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)}\left(\mathrm{x}_i - \mu_j^{(t+1)}\right)^2}{\sum_{i=1}^{N} z_{ij}^{(t)}}$$

  - Set $c_j = \mu_j$, and $b_j = \sigma_j$.

  - Increase t by 1.

After finishing Step 1, we obtain the initial values of the parameters $\Xi^{(t)} = (\mu_j, \sigma_j, c_j, b_j, \alpha)$.

Step 2: Evaluate $z_{ij}^{(t)}$ given by Eq.(4.11):

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)}\Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{k=1}^{K}\pi_{ik}^{(t)}\Phi(\mathrm{x}_i|\Theta_k^{(t)})}$$

where, $\Phi^{(t)}(\mathrm{x}_i|\Theta_j)$ and $\pi_{ij}^{(t)}$ are calculated from Eq.(2.23) and Eq.(4.3), respectively.

Step 3: Update parameters $\Xi = \{\Theta, \Pi\} = (\mu_j, \sigma_j, c_j, b_j, \alpha)$ to obtain the new parameters $\Theta^{(t+1)}$, which can then be calculated and updated by using the gradient method [82], [83], [85].

$$\Xi^{(t+1)} = \Xi^{(t)} - \eta \nabla E\left(\Xi^{(t)}\right) \tag{4.12}$$

where, $\nabla E\left(\Xi^{(t)}\right) = [\partial E/\partial \mu_j, \partial E/\partial \sigma_j, \partial E/\partial c_j, \partial E/\partial b_j, \partial E/\partial \alpha]$. Details of the formulae $\partial E/\partial \mu_j$, $\partial E/\partial \sigma_j$, $\partial E/\partial c_j$, $\partial E/\partial b_j$ and $\partial E/\partial \alpha$ used to update the parameters μ_j, σ_j, c_j, b_j and α are given in the Appendix A. η is the learning rate and its value is sufficiently small. In this method, we have selected $\eta = 10^{-5}$.

Step 4: Check for convergence of either the negative log-likelihood function, or the parameter values. If the convergence criterion is not satisfied, then set $\Xi^{(t)} = \Xi^{(t+1)}$, and return to step 2.

4.4 Experiments

In this section, the performance of the proposed algorithm is compared to the K-means [86], standard GMM [75], SVFMM [72], neighborhood expectation maximization (NEM) [48], [87], iterated conditional model (ICM) [88], mode field (MODEF), SIMF, MEANF [63] and [64], fast generalized fuzzy

c-means (FGFCM) [35], and hidden markov random field based fuzzy c-means (HMRF-FCM) [89]. The source code for the SVFMM algorithm can be downloaded from http://www.cs.uoi.gr/∼kblekas/sw/MAPsegmentation.html. Parameter β in SVFMM algorithms is assigned a value of 0.1. This method was implemented in the MATLAB environment. For the NEM, ICM, MODEF, SIMF and MEANF methods, we used a software implementation developed by the authors [63], [64] (for the Windows environment) publicly available at http://spacem3.gforge.inria.fr/. Standard GMM, SVFMM, NEM, ICM, MODEF, SIMF and MEANF methods are initialized by the K-means algorithm similar to the initialization of the proposed algorithm. SVFMM methods use a first order neighborhood system, while ICM, MODEF , SIMF and MEANF methods use a second order (8-neighbor) neighborhood system. For these methods, the standard isotropic Potts model is used and the temperature value β is heuristically optimized. For the FGFCM method, we have selected λ_s=3 and λ_g=6. The optimally selected value of the fuzziness parameter λ in the HMRF-FCM is set to five. These methods were run until convergence of the iteration steps.

To test the robustness of the proposed method, different types of noise with varying levels have been introduced to each image. The proposed method was implemented and tested on synthetic and real-world images. For the synthetic images, to compare the results obtained, the misclassification ratio (MCR) [91], as shown in Eq.(2.27), has been used, which is the number of misclassified pixels divided by the total number of pixels. For natural real-

world images, we employ the probabilistic rand (PR) index [93], as shown in Eq.(2.28), to compare the results of these methods. The proposed method was implemented and tested on a PC (Pentium 4, running at 3 GHz with 1GB of RAM) in MATLAB compiled visual C environment. The algorithms were tested using synthetic and real-world images.

4.4.1 Synthetic Images

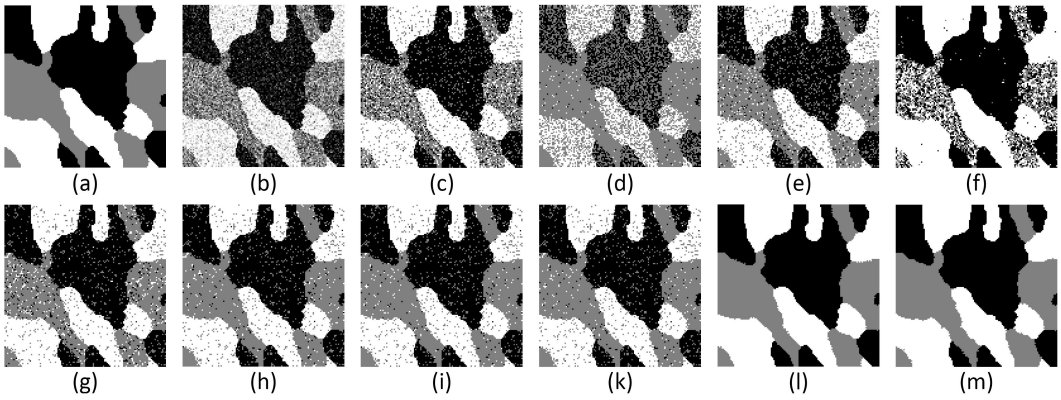

Figure 4.1: The first experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.05 variance), (c): K-means (MCR = 17.315%), (d): standard GMM (MCR = 33.982%), (e): SVFMM (MCR = 13.671%), (f): NEM (MCR = 17.034%), (g) ICM (MCR = 9.605%), (h): MODEF (MCR = 7.781%), (i): SIMF (MCR = 7.725%), (k): MEANF (MCR = 7.721%), (l) HMRF-FCM (MCR = 0.823%), (m): The proposed method (MCR = 0.653%).

In the first experiment, a synthetic image (128x128 image resolution) similar to the one used in [91], as shown in Figure 4.1(a), was used to compare the performance of the proposed algorithm with others. The image has three labels (K=3) with luminance values [0, 0.5, 1]. The image shown in Figure 4.1(b) is obtained by corrupting the original image with Gaussian noise (0

Table 4.1: Comparison of the proposed method with other methods in term of MCR (%), for the first experiment.

Methods	Gaussian noise (0 mean, var)			
	var=0.02	var=0.03	var=0.04	var=0.05
K-means	5.212	9.985	13.721	17.315
Standard GMM	29.974	31.332	32.445	33.982
SVFMM	4.168	7.385	10.546	13.671
NEM	3.584	10.101	15.130	17.034
ICM	0.385	1.025	2.386	9.605
MODEF	0.598	1.837	3.757	7.781
SIMF	0.586	1.831	3.661	7.725
MEANF	0.585	1.800	3.659	7.721
HMRF-FCM	0.177	0.391	0.493	0.823
The proposed method	0.122	0.290	0.415	0.653

mean, 0.05 variance). Figure 4.1(c) presents the segmentation result obtained by employing K-means algorithm. This result is used in the initialization step for all the remaining methods. From Figure 4.1(d) to Figure 4.1(l), we present the segmentation results obtained by employing standard GMM, SVFMM, NEM, ICM, MODEF, SIMF, MEANF and HMRF-FCM, respectively. As can be seen, the accuracy of the standard GMM method is poor compared to SVFMM in Figure 4.1(e). However, the SVFMM method requires a large number of parameters in the process of estimation ($K(2+N)= 49158$ parameters in this example). In Figure 4.1(l), the HMRF-FCM algorithm reduces the effect of noise significantly and can segment the image well. However, the proposed method in Figure 4.1(m) can segment the image with a better result. The results obtained with varying levels of noise are presented in Table 4.1. As can be seen, the proposed method has a lower MCR compared with the

other methods.

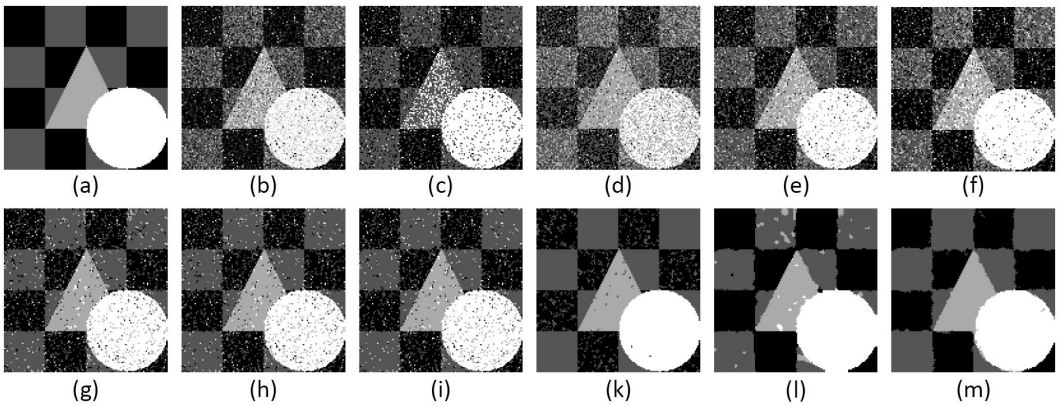

Figure 4.2: The second experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with mixed noise (salt and pepper noise (0.03%) + Gaussian noise (0 mean, 0.01 variance)), (c): K-means (MCR = 24.609%), (d): standard GMM (MCR = 40.209%), (e): SVFMM (MCR = 22.338%), (f): ICM (MCR = 22.001%), (g): MODEF (MCR = 12.744%), (h): SIMF (MCR = 13.307%), (i): MEANF (MCR = 12.200%), (k) FGFCM (MCR = 5.562%), (l) HMRF-FCM (MCR = 3.692%), (m): The proposed method (MCR = 2.789%).

Figure 4.2(a) shows the synthetic image (128x128 image resolution) which consists of four labels (K=4) with luminance values [0, 1/3, 2/3, 1], used in the second experiment. The image shown in Figure 4.2(b) is obtained by corrupting the original image with mixed noise. First corrupting the original image with salt & pepper noise (noise=0.03), and then adding Gaussian noise (0 mean, 0.01 variance). Figure 4.2(d)–(l), we present the segmentation results of standard GMM, SVFMM, ICM, MODEF, SIMF MEANF, FGFCM and HMRF-FCM, respectively. Amongst these methods, FGFCM and HMRF-FCM classify the image with the lowest MCR. However, as compared with the other methods, the segmentation result obtained by employing the proposed

Table 4.2: Comparison of the proposed method with other methods in term of MCR (%), for the second experiment.

Methods Methods	Gaussian noise (0 mean, var)		Mixed Noise: Salt & Pepper Noise (sp) + Gaussian Noise (0 mean, var)	
	var=0.01	var=0.03	sp=0.03 var=0.01	sp=0.03 var=0.03
K-means	7.489	22.845	9.565	24.609
Standard GMM	27.624	37.634	29.290	40.209
SVFMM	4.687	19.860	6.347	22.338
ICM	0.463	17.126	3.118	22.001
MODEF	0.512	5.120	3.088	12.744
SIMF	0.511	4.235	3.125	13.307
MEANF	0.494	4.095	3.082	12.200
FGFCM	0.262	2.777	0.676	5.562
HMRF-FCM	0.225	2.032	0.643	3.692
Proposed method	0.189	1.007	0.469	2.789

method, as shown in Figure 4.2(l), demonstrates a higher degree of robustness with respect to the given level of noise. Note that the initialization for all algorithms was carried out by using K-means algorithm as shown in Figure 4.2(c). Table 4.2 contains the results obtained for all methods, for varying levels of noise. As can be easily seen, the proposed method outperforms other methods, with a lower MCR.

In the third experiment, an image with four labels (K=4) with luminance values [0, 1/3, 2/3, 1] as shown in Figure 4.3(a) was used to test the effect of simple competitive selection used in the proposed method. We also added this selection step to MODEF method to determine its effect on the result. The image shown in Figure 4.3(b) is obtained by corrupting the orig-

inal image with Gaussian noise (0 mean, 0.03 variance). The results for the MODEF method, and the proposed method, without and with the simple competitive selection are shown in Figure 4.3(c), Figure 4.3(d), Figure 4.3(e) and Figure 4.3(f), respectively. It can be easily seen that even though by employing simple competitive selection, there is a slight decrease in the MCR for both MODEF and the proposed methods, its main effect is mostly qualitative in nature. Thus, the significant quantitative difference between the MCRs for both these methods can be attributed to the higher degree of robustness of the proposed method with respect to noise.

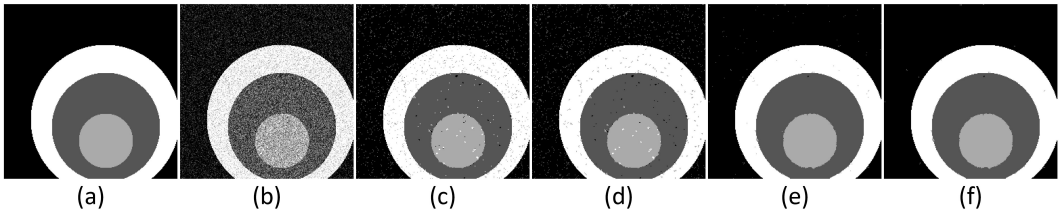

Figure 4.3: Affect of the simple competitive selection, (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.03 variance), (c): MODEF (MCR = 3.512%), (d): MODEF with the simple competitive selection (MCR = 3.117%), (e): proposed method without the simple competitive selection (MCR = 0.199%), (f): proposed method with the simple competitive selection (MCR = 0.197%).

4.4.2 Natural Images

Analysis of real-world outdoor scenes is a challenging problem for image segmentation. It is quite hard to come up with a good model for objects such as flowers, trees, birds etc. In this set of experiments, we compare the performance of various algorithms on real world images obtained from the Berkeley's

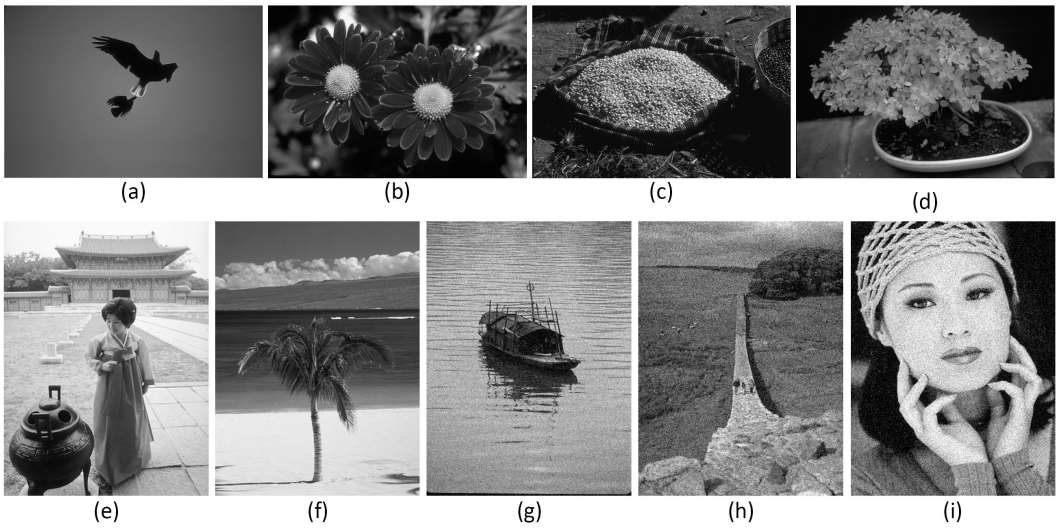

Figure 4.4: Images from the Berkeleys grayscale image segmentation dataset, (a): 135069, (b): 124084, (c): 58060, (d): 353013 with Gaussian noise (0 mean, 0.001 variance), (e): 239007, (f): 46076, (g): 15088 with Gaussian noise (0 mean, 0.005 variance), (h): 374067 with Gaussian noise (0 mean, 0.01 variance), (i): 302003 with Gaussian noise (0 mean, 0.01 variance).

image segmentation dataset [98]. This set of grayscale images includes natural Images along with their ground truth segmentation results provided by human subjects. PR index, previously introduced in [93], was used to obtain the performance evaluation results in this subsection. It contains values that are between 0 and 1, with values closer to 1 indicating a good result.

In the first experiment, a set of real world grayscale images with and without artificial noise were used to evaluate the performance of the proposed method against K-means, standard GMM, MODEF, SIMF, MEANF, FGFCM, HMRF-FCM methods. Table 4.3 contains the cumulative results obtained for all methods, for the given set of real world images. As can be easily seen, on average, the proposed method outperforms other methods with a

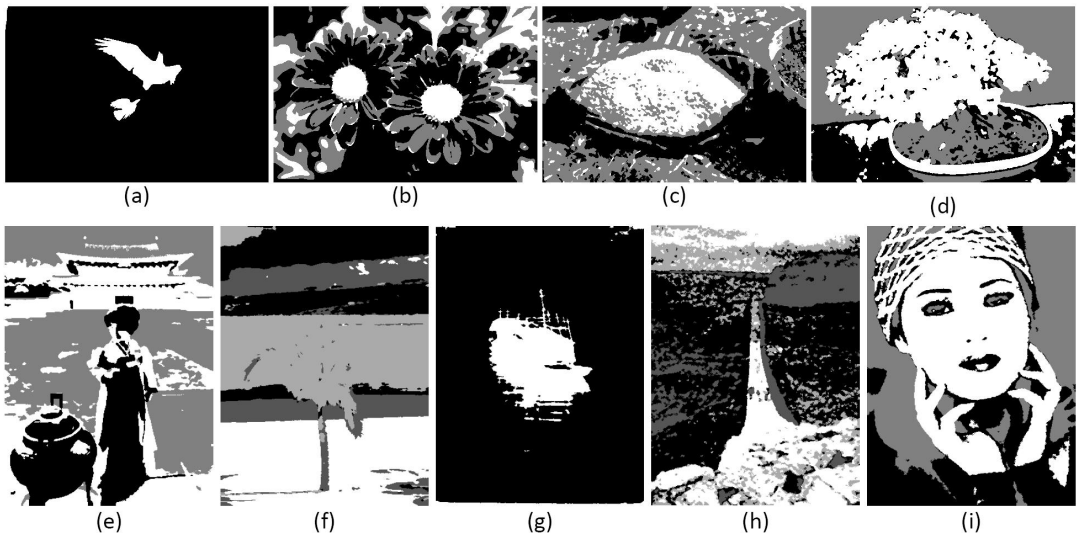

Figure 4.5: Image segmentation results obtained by employing the proposed method, (a): 135069, (b): 124084, (c): 58060, (d): 353013 with Gaussian noise (0 mean, 0.001 variance), (e): 239007, (f): 46076, (g): 15088 with Gaussian noise (0 mean, 0.005 variance), (h): 374067 with Gaussian noise (0 mean, 0.01 variance), (i): 302003 with Gaussian noise (0 mean, 0.01 variance).

higher PR. Figure 4.5 presents the segmentation results obtained by employing the proposed method for the given set of images shown in Figure 4.4.

In the second experiment, we tried to segment the real-world grayscale image (481x321) as shown in Figure 4.6(a) into four labels (K=4): "tree", "near mountain", "far mountain", "sky". Images in Figure 4.6(b), Figure 4.6(c), Figure 4.6(e) and Figure 4.6(h) show the result obtained by using the K-means, standard GMM, ICM and FGFCM methods. As can be seen, the accuracy of segmentation for these methods is quite poor. The sharp edge between the "near mountain" and the "far mountain", and between the "far mountain" and the "sky" is lost because of noise. Extraction accuracies of the SVFMM, MODEF and MEANF methods are shown in Figure 4.6(d), Figure

Table 4.3: Comparison of image segmentation results on Berkeley's grayscale image segmentation dataset: Probabilistic Rand (PR) Index.

Image	Noise (var)	labels	Sta. GMM	SIMF	MEANF	FGFCM	HMRF-FCM	Our method
135069	-	2	0.982	0.983	0.983	0.549	0.984	0.985
124084	-	3	0.499	0.514	0.513	0.521	0.526	0.558
69020	-	3	0.486	0.555	0.556	0.554	0.559	0.605
12003	-	3	0.515	0.614	0.615	0.605	0.618	0.623
58060	-	3	0.568	0.618	0.611	0.571	0.615	0.622
239007	-	3	0.652	0.662	0.660	0.655	0.668	0.671
46076	-	4	0.812	0.826	0.824	0.807	0.826	0.828
55067	-	3	0.843	0.881	0.881	0.879	0.888	0.891
353013	0.001	3	0.680	0.736	0.722	0.687	0.740	0.742
310007	0.005	7	0.628	0.765	0.764	0.648	0.776	0.784
61060	0.005	3	0.586	0.678	0.682	0.632	0.679	0.681
15088	0.005	3	0.844	0.860	0.867	0.844	0.869	0.871
24063	0.005	3	0.765	0.815	0.818	0.796	0.815	0.826
374067	0.01	4	0.646	0.706	0.704	0.673	0.705	0.708
302003	0.01	3	0.686	0.718	0.717	0.710	0.715	0.727
Mean	-	-	0.679	0.729	0.728	0.675	0.732	0.741

4.6(f) and Figure 4.6(g). In the left hand side of the results, looking closely in the "tree" area, it can be seen that there is a small portion of pixels that have been misclassified. HMRF-FCM methods in Figure 4.6(i) can produce a better segmentation. The sharp edge between mountains is clearly defined, and there was no missing region as in the MEANF method. However, the proposed method in Figure 4.6(k), can better classify with more detail along the sharp edge between "near mountain" and "far mountain", as compared with HMRF-FCM methods.

In order to further test the accuracy and determine the efficiency of the

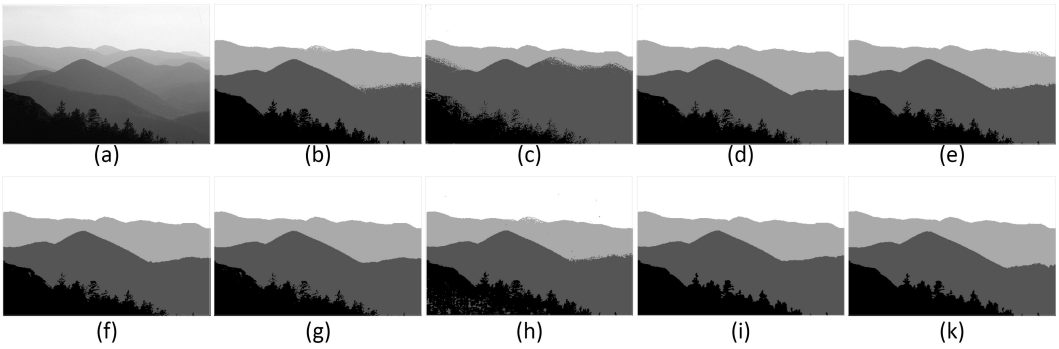

Figure 4.6: Grayscale image segmentation (55067), (a): original image, (b): K-means (PR = 0.879), (c): standard GMM (PR = 0.843), (d): SVFMM (PR = 0.882), (e): ICM (PR = 0.880), (f): MODEF (PR = 0.882), (g): MEANF (PR = 0.881), (h): FGFCM (PR = 0.879), (i): HMRF-FCM (PR = 0.887), (k): The proposed method (PR = 0.891).

proposed algorithm, another real word grayscale image was used. The image shown in Figure 4.7(b) is obtained by corrupting the original image in Figure 4.7(a) with Gaussian noise (0 mean, 0.005 variance). Figure 4.7(c)–(k), present the segmentation results obtained by employing K-means, standard GMM, SVFMM, MODEF, MEANF, FGFCM, HMRF-FCM and the proposed method, respectively. From visual inspection of the results, K-means, standard GMM, SVFMM methods are unable to segment the image successfully. FGFCM method demonstrates a better performance compared to the standard GMM method. However, the effect of noise on the final segmented image is still quite high. MODEF, MEANF, HMRF-FCM algorithms can produce a better segmentation. However, compared with these algorithms, the effect of noise in the top right hand side of the final segmented image, obtained by employing the proposed method is far less. Moreover, we also notice that the details of the balcony are better preserved by the proposed method compared

with MODEF, MEANF, HMRF-FCM methods.

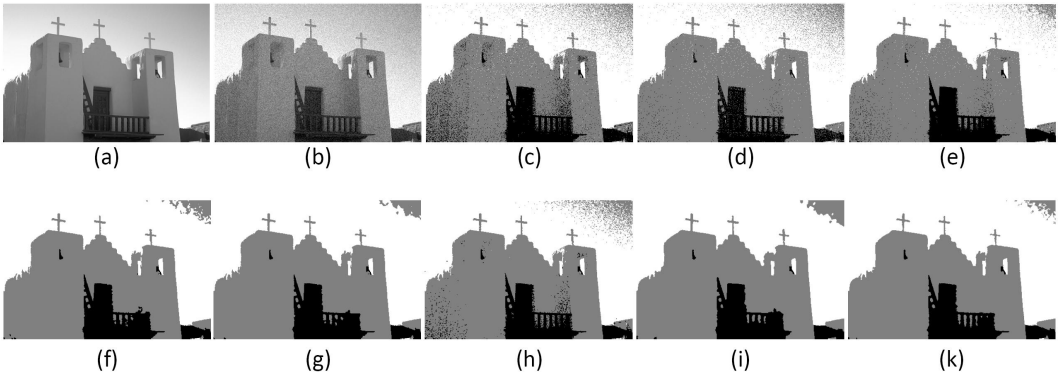

Figure 4.7: Noisy grayscale image segmentation (24063), (a): original image, (b): corrupted original image with Gaussian noise (0 mean, 0.005 variance), (c): K-means (PR = 0.778), (d): standard GMM (PR = 0.765), (e): SVFMM (PR = 0.787), (f): MODEF (PR = 0.814), (g): MEANF (PR = 0.818), (h): FGFCM (PR = 0.796), (i): HMRF-FCM (PR = 0.815), (k): The proposed method (PR = 0.826).

In the final experiment, a real world image (481x321 image resolution), as shown in Figure 4.8(a), is used for evaluating the computational cost. The objective is to segment this image into two labels (K=2). The K-means method was implemented in Visual C environment. For the GMM, ICM, MODEF, SIMF, MEANF methods, we used the software implementation developed by the authors [7], [14] (for Windows environment, C++ Language) publicly available at http://spacem3.gforge.inria.fr/. The proposed method was implemented in MATLAB compiled visual C environment. All experiments were performed on a PC (Pentium 4, running at 3 GHz with 1GB of RAM). These methods were run until convergence. The segmentation results obtained by employing K-means, standard GMM, ICM, MODEF, SIMF, MEANF and the proposed method are shown in Figure 9(b)–(h), respectively.

Once more, the proposed method demonstrates robustness with respect to noise, yielding a better segmentation result. K-means is the quickest method taking 0.7 seconds, while the slowest is MEANF taking 445 seconds. The proposed method takes 61 seconds and converges after 8 iterations, as shown in Figure 4.9. However, as shown Figure 4.8, the proposed method performs quite well in segmenting the image into two labels.

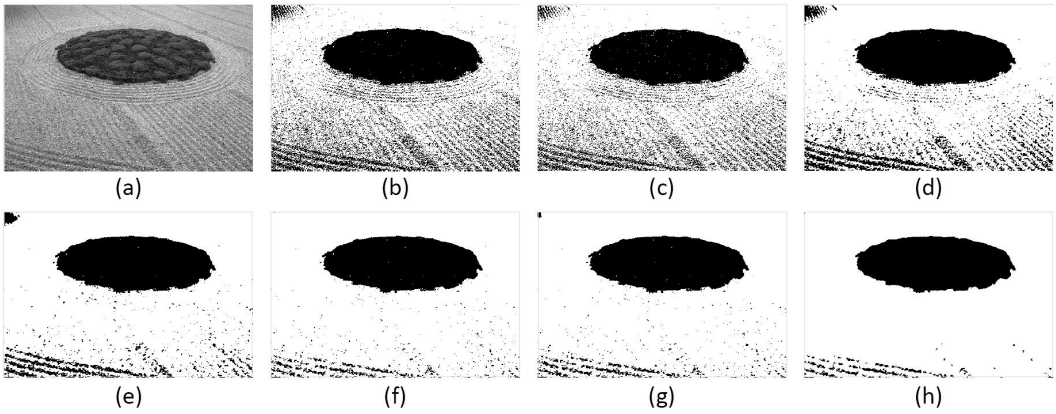

Figure 4.8: Computational cost (in seconds) comparison, (a): original image, (b): K-means (0.7 sec), (c): standard GMM (36 sec), (d): ICM (169 sec), (e): MODEF (390 sec), (f): SIMF (432 sec), (g): MEANF (445 sec), (h): The proposed method (61 sec).

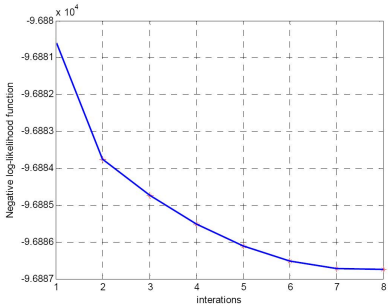

Figure 4.9: Minimization Progress of the negative log-likelihood function of the proposed algorithm, for the final experiment.

4.5 Conclusions

In this chapter, we have presented an extension of the standard GMM for grayscale image segmentation. We have proposed a novel approach to incorporate the spatial relationships between neighboring pixels into the standard GMM model. Differing from the standard GMM, the prior distribution in the proposed model is different for each pixel and depends on the neighbors of the pixel and their corresponding parameters. The proposed model is easy to implement and requires fewer parameters compared to the models based on MRF. We have presented a new way to estimate the unknown parameters of the proposed model, based on the gradient method. The proposed method has been tested with many noisy synthetic and real world images, thereby demonstrating the excellent performance of the proposed model in segmenting the grayscale images.

Chapter 5

Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation

5.1 Introduction

In this chapter, a new mixture model for image segmentation is presented. We propose a new way to incorporate spatial information between the neighboring pixels into the Gaussian mixture model (GMM) based on Markov random field (MRF). In comparison to other mixture models that are complex and computationally expensive, the proposed method is fast and easy to implement. In mixture models based on MRF, the M-step of the expectation maximization (EM) algorithm cannot be directly applied to the prior distribution π_{ij} for maximization of the log-likelihood with respect to the corresponding pa-

rameters. Compared with these models, our proposed method directly applies the EM algorithm to optimize the parameters, which makes it much simpler. Finally, experimental results obtained by employing the proposed method on many synthetic and real-world grayscale and colored images demonstrate its robustness, accuracy and effectiveness, as compared with other mixture models.

In order to partition an image consisting of N pixels into K labels, GMM [58] assumes that each observation x_i is considered independent of the label Ω_j. The density function $f(\mathrm{x}_i|\Pi,\Theta)$ at an observation x_i in Eq.(3.16) is rewritten by:

$$f(\mathrm{x}_i|\Pi,\Theta) = \sum_{j=1}^{K} \pi_{ij}\Phi(\mathrm{x}_i|\Theta_j)$$

where $\Pi = \{\pi_{ij}\}$; i=1,2,...,N; j=1,2,...,K; is the set of prior distributions modeling the probability that pixel x_i is in label Ω_j, which satisfies the constraints in Eq.(3.17). And, $\Phi(\mathrm{x}_i|\Theta_j)$ is the Gaussian distribution, called a component of the mixture.

For the case of a D-dimensional vector x_i, each Gaussian distribution $\Phi(\mathrm{x}_i|\Theta_j)$ in Eq.(2.6) can be written in the form:

$$\Phi(\mathrm{x}_i|\Theta_j) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_j|^{1/2}} \exp\left\{-\frac{1}{2}(\mathrm{x}_i-\mu_j)^T\Sigma_j^{-1}(\mathrm{x}_i-\mu_j)\right\} \tag{5.1}$$

where $\Theta = \{\Theta_j\}$ and $\Theta_j = \{\mu_j, \Sigma_j\}$; i=1,2,...,N. The D-dimensional vector μ_j is the mean, the DxD matrix Σ_j is the covariance, and $|\Sigma_j|$ denotes the determinant of Σ_j. Note that the observation x_i is modeled as statistically independent, the joint conditional density in Eq.(3.18) of the data set

$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_N)$ can be modeled as:

$$p(\mathrm{X}|\Pi, \Theta) = \prod_{i=1}^{N} f(\mathrm{x}_i|\Pi, \Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j)$$

Since the observation x_i is considered to be independent given the pixel label, the spatial correlation between the neighboring pixels is not taken into account. As a result, the segmented image is sensitive to noise. To overcome this problem, MRF distribution [99], as shown in Eq.(3.19), is applied to incorporate the spatial correlation amongst label values:

$$p(\Pi) = W^{-1} \exp\left\{-\frac{1}{T} U(\Pi)\right\}$$

By using the Eq.(3.21), the log-likelihood function can be derived as:

$$\begin{aligned} L(\Theta, \Pi|\mathrm{X}) &= \log\left(p(\Pi, \Theta|\mathrm{X})\right) \\ &= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j)\right\} + \log p(\Pi) \\ &= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j)\right\} - \log W - \frac{1}{T} U(\Pi) \end{aligned}$$

Depending on the type of energy $U(\Pi)$, we can have different kinds of models. In the Bayesian auto-logistic model [99], the function $U(\Pi)$ is chosen to incorporate the spatial correlation, as shown:

$$U(\Pi) = \sum_{i=1}^{N} \sum_{j=1}^{K} \alpha_{ij} \pi_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{K} \sum_{m \in \mathcal{N}_i} \beta_{ijm} \pi_{ij} \pi_{mj} \tag{5.2}$$

where α_{ij} and β_{ijm} form the parameter set. Although this model involves many parameters, its segmentation is not sufficiently robust to noise.

Other mixture models based on MRF for pixel labeling have been successfully applied to image segmentation [58], [72], [73], [100] and different ways

are adopted to select the energy $U(\Pi)$. In [58], $U(\Pi)$ is given by:

$$U(\Pi) = \beta \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} \sum_{j=1}^{K} (\pi_{ij} - \pi_{mj})^2 \tag{5.3}$$

while, in another MRF model based method [72], $U(\Pi)$ is given by:

$$U(\Pi) = \beta \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} \left[1 + \left(\sum_{j=1}^{K} (\pi_{ij} - \pi_{mj})^2 \right)^{-1} \right]^{-1} \tag{5.4}$$

where, β in Eq.(5.3) and Eq.(5.4) represent a constant value.

In [73], spatial information is taken into account, and $U(\Pi)$ is given as:

$$U(\Pi) = \sum_{i=1}^{N} \sum_{j=1}^{K} \sum_{s=1}^{S} \left[\frac{1}{2} \log \beta_{js}^2 - \frac{1}{2} \frac{\left(\sum\limits_{m \in \partial_i} (\pi_{ij} - \pi_{mj}) \right)^2}{\beta_{js}^2} \right] \tag{5.5}$$

where, S is the total number of the considered directions. In the general case, S is equal to four (S=4: horizontal, vertical and 2 diagonal directions). β_{js} in Eq.(5.5) is a variable parameter.

As shown in Eq.(5.3), Eq.(5.4), and Eq.(5.5), the incorporation of local information adds complexity. In order to maximize the likelihood with respect to the parameters Π and Θ, an iterative EM algorithm can be applied. However, due to the complexity of the log-likelihood function, the M-step of EM algorithm cannot be applied directly to the prior distribution π_{ij}. Note, that the prior distribution π_{ij} should satisfy the constraints in Eq.(3.17). Thus, the resulting algorithms are computationally complex and utilize large amounts of computational power to solve the constrained optimization problem of the prior distribution π_{ij}. For details regarding the maximization of the log-likelihood function, we refer the readers to [58], [72], [73], [100].

5.2 Proposed Method

Various mixture models differ based on the way they derive the strength of the smoothing prior $U(\Pi)$. In [99], given in Eq.(5.2), the smoothing prior $U(\Pi)$ has a simple form, thus, it is easy to optimize the parameter set $\{\Pi, \Theta\}$ to maximize the log-likelihood function. However, one of its main drawbacks is that the segmentation result is not robust to noise. Models in [58], [72], [73] represented by Eq.(5.3)–(5.5), make use of a complex smoothing prior. Their primary disadvantage lies in its additional training complexity. The M-step of the EM algorithm cannot be applied directly to the prior distribution, which therefore corresponds to an increase in the algorithms' complexity. In order to overcome these disadvantages, we introduce a novel factor G_{ij} defined as:

$$G_{ij}^{(t)} = \exp\left[\frac{\beta}{2N_i}\sum_{m\in\mathcal{N}_i}(z_{mj}^{(t)} + \pi_{mj}^{(t)})\right] \tag{5.6}$$

where, z_{mj} is the posterior probability. β is the temperature value that controls the smoothing prior. In this method, it has been set to 12 (β =12). ∂_i is the neighborhood of the i-th pixel, including itself. A square window of size 5x5 is used in this method. N_i is the number of neighboring pixels around the pixel x_i in this window. By taking a closer look at Eq.(5.6), it can be visualized that the factor G_{ij} in the proposed method acts as a mean filter. The goal is to simply replace each pixel value with the average value of its neighbors, including itself. The main advantage of G_{ij} is the ease of implementation, and incorporation of the spatial relationships amongst neighboring pixels in a simpler metric.

Next, we propose a novel approach to incorporate the spatial information into the smoothing prior. The new smoothing prior $U(\Pi)$ is given by:

$$U(\Pi) = -\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \tag{5.7}$$

where, t indicates the iteration step. The MRF distribution $p(\Pi)$ is given by:

$$p(\Pi) = W^{-1} \exp\left\{ \frac{1}{T} \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \right\} \tag{5.8}$$

Given the MRF distribution $p(\Pi)$, the log-likelihood function is written in the form.

$$L(\Theta, \Pi|\mathrm{X}) = \sum_{i=1}^{N} \log\left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(\mathrm{x}_i|\Theta_j) \right\} - \log W + \frac{1}{T} \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \tag{5.9}$$

Application of the complete data condition in [46], [58], maximizing the log-likelihood function $L(\Theta, \Pi|\mathrm{X})$ in Eq.(5.9) will lead to an increase in the value of the objective function $H(\Theta, \Pi|\mathrm{X})$.

$$\begin{aligned} H(\Theta, \Pi|\mathrm{X}) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} + \log \Phi(\mathrm{x}_i|\Theta_j^{(t+1)}) \right\} \\ - \log W + \frac{1}{T} \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \end{aligned} \tag{5.10}$$

The conditional expectation values z_{ij} of the hidden variables, as shown in Eq.(2.33), is rewritten as follows:

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \Phi(\mathrm{x}_i|\Theta_j^{(t)})}{\sum_{k=1}^{K} \pi_{ik}^{(t)} \Phi(\mathrm{x}_i|\Theta_k^{(t)})}$$

The next objective is to optimize the parameter set $\{\Pi, \Theta\}$ in order to maximize the objective function $H(\Theta, \Pi|\mathrm{X})$ in Eq.(5.10). Similar to the MRF

based methods [58], [72], [73], [100], W and T in Eq.(5.10) are set equal to one (W=1, T=1). From Eq.(5.10), the new objective function is given by:

$$H(\Theta,\Pi|\mathrm{X}) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} + \log \Phi(\mathrm{x}_i|\Theta_j^{(t+1)}) \right\} + \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \tag{5.11}$$

From Eq.(5.1), the function in Eq.(5.11) can be rewritten as:

$$\begin{aligned} H(\Theta,\Pi|\mathrm{X}) = & \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} - \frac{D}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_j^{(t+1)}| \right\} \\ & + \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \left\{ -\frac{1}{2}(\mathrm{x}_i - \mu_j^{(t+1)})^T \Sigma_j^{-1(t+1)} (\mathrm{x}_i - \mu_j^{(t+1)}) \right\} \\ & + \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \end{aligned} \tag{5.12}$$

To maximize this function, the EM algorithm [52], [56] is applied. Let us now consider the derivation of the function $H(\Theta,\Pi|\mathrm{X})$ with the means μ_j at the (t+1) iteration step. We have:

$$\frac{\partial H}{\partial \mu_j^{(t+1)}} = \sum_{i=1}^{N} z_{ij}^{(t)} \left[-\frac{1}{2}(2\Sigma_j^{-1(t+1)}\mu_j^{(t+1)} - 2\Sigma_j^{-1(t+1)}\mathrm{x}_i) \right] \tag{5.13}$$

The solution of $\partial H/\partial \mu_j = 0$ yields the minimizer of μ_j at the (t+1) step:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} \mathrm{x}_i}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{5.14}$$

Thus, setting the derivative of the function $H(\Theta,\Pi|\mathrm{X})$ in Eq.(5.11) with respect to Σ_j^{-1} at the (t+1) iteration step we have:

$$\frac{\partial H}{\partial \Sigma_j^{-1(t+1)}} = \sum_{i=1}^{N} z_{ij}^{(t)} \left[\frac{1}{2}\Sigma_j^{(t+1)} - \frac{1}{2}(\mathrm{x}_i - \mu_j^{(t+1)})(\mathrm{x}_i - \mu_j^{(t+1)})^T \right] \tag{5.15}$$

and, equating it to zero yields:

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} (\mathrm{x}_i - \mu_j^{(t+1)})(\mathrm{x}_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{5.16}$$

An important consideration is that the prior distribution should satisfy the constraints in Eq.(3.17). In order to enforce these constraints, we use the Lagrange's multiplier η_i for each data point:

$$\frac{\partial}{\partial \pi_{ij}^{(t+1)}} \left[H - \sum_{i=1}^{N} \eta_i \left(\sum_{j=1}^{K} \pi_{ij}^{(t+1)} - 1 \right) \right] = 0 \tag{5.17}$$

Eq.(5.17) can be rewritten in the following form:

$$\frac{z_{ij}^{(t)}}{\pi_{ij}^{(t+1)}} + \frac{G_{ij}^{(t)}}{\pi_{ij}^{(t+1)}} - \eta_i = 0 \tag{5.18}$$

The constraint $\sum_{j=1}^{K} \pi_{ij} = 1$ enables the Lagrange multiplier η_i to satisfy the following condition:

$$\eta_i = 1 + \sum_{j=1}^{K} G_{ij}^{(t)} \tag{5.19}$$

The necessary condition for determining the prior distribution π_{ij} at the $(t+1)$ iteration step becomes:

$$\pi_{ij}^{(t+1)} = \frac{z_{ij}^{(t)} + G_{ij}^{(t)}}{\sum_{k=1}^{K} \left(z_{ik}^{(t)} + G_{ik}^{(t)} \right)} \tag{5.20}$$

So far, the discussion has focused on estimating $\{\Pi, \Theta\}$ of the model, in order to assign a label Ω_j to the pixel x_i. The various steps of the proposed mixture model incorporating spatial information based on MRF can be summarized as follows:

Step 1: Initialize the parameters $\{\Pi, \Theta\}$: the means μ_j, covariance values Σ_j and prior distributions π_{ij}.

Step 2 (E-step):

+ Evaluate the values z_{ij} in Eq.(2.33) using the current parameter values.

+ Update the factor G_{ij} by using Eq.(5.6).

Step 3 (M-step): Re-estimate the parameters $\{\Pi, \Theta\}$.

+ Update the means μ_j by using Eq.(5.14).

+ Update covariance values Σ_j by using Eq.(5.16).

+ Update prior distributions π_{ij} by using Eq.(5.20).

Step 4: Evaluate the log-likelihood in Eq.(5.9) and check the convergence of either the log-likelihood function, or the parameter values. If the convergence criterion is not satisfied, then go to step 2.

Once the parameter-learning phase is complete, every pixel x_i is assigned to the label with the largest posterior probability z_{ij} by using Eq.(2.53). In the next section, we will demonstrate the robustness, accuracy and effectiveness of the proposed model, as compared with other GMM based approaches.

5.3 Experiments

In this section, the performance of the proposed algorithm is compared with algorithms based on the K-means algorithm, standard GMM, and mixture models based on MRF such as ICM, SIMF and MEANF. We also compare the results with SVFMM, and the class-adaptive spatially variant finite mixture

model (CA-SVFMM) [73]. For ICM, SIMF and MEANF methods, a standard isotropic Potts model with a second order (8-neighbor) neighborhood system is utilized, where the temperature value β is automatically optimized. All these methods were run until convergence. Parameter β in SVFMM algorithms is manually set a value of 0.1 based on user experience. This method was implemented in the MATLAB environment. The proposed method was implemented and tested on both synthetic and real-world images. For synthetic images, in order to compare the results obtained, MCR has been used, which is given by the number of misclassified pixels divided by the total number of pixels. All these methods were implemented and tested on a PC (Pentium 4, running at 3 GHz with 1GB of RAM).

5.3.1 Segmentation of Synthetic Images

In the first experiment, a synthetic image (128x128 image resolution) as shown in Figure 5.1(a), was used to compare the performance of the proposed algorithm with others. The image has four labels with luminance values [0, 1/3, 2/3, 1]. The image shown in Figure 5.1(b) was obtained by corrupting the original image with Gaussian noise (0 mean, 0.03 variance). Figure 5.1(c) presents the segmentation result obtained by employing the K-means algorithm, which is used during the initialization step for all the remaining methods. In Figure 5.1(d) to Figure 5.1(i), we present the segmentation results obtained by employing standard GMM, SVFMM, CA-SVFMM, ICM, SIMF and MEANF methods respectively. Under given conditions the segmentation accuracy of

the standard GMM method is quite poor. SIMF (MCR = 3.83%) in Figure 5.1(h) and MEANF (MCR = 3.55%) in Figure 5.1(i) demonstrate better classification, and segment the image well. However, the proposed method in Figure 5.1(j) segments the image better with the lowest MCR = 1.13%.

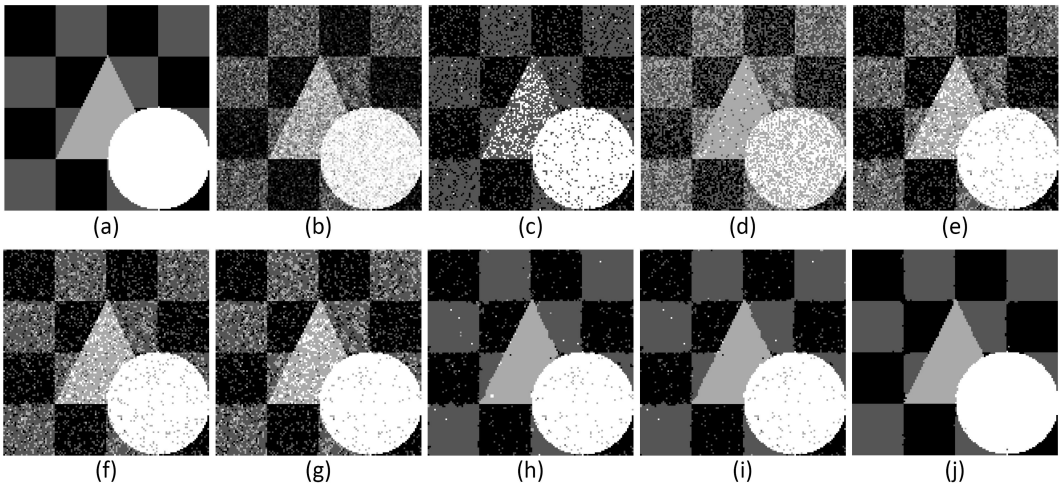

Figure 5.1: The first experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.03 variance), (c): K-means, (d): Standard GMM (MCR = 41.67%), (e): SVFMM (MCR = 23.28%), (f): CA–SVFMM (MCR = 20.29%), (g): ICM (MCR = 20.23%), (h): SIMF (MCR = 3.83%), (i) MEANF (MCR = 3.55%), (j): Proposed method (MCR = 1.13%).

Table 5.1: Computational cost (in seconds) comparison for the synthetic image in the first experiment.

Methods	Sta. GMM	SV FMM	CA–S VFMM	ICM	SIMF	MEA NF	our method
Time (sec)	3.2	80.3	218.1	85.4	163.2	121.8	4.9
MCR (%)	41.67	23.28	20.29	20.23	3.83	3.55	1.13

Moreover, due to the inherent simplicity of the proposed algorithm, it has low computational cost. In this experiment, for ICM, SIMF and MEANF

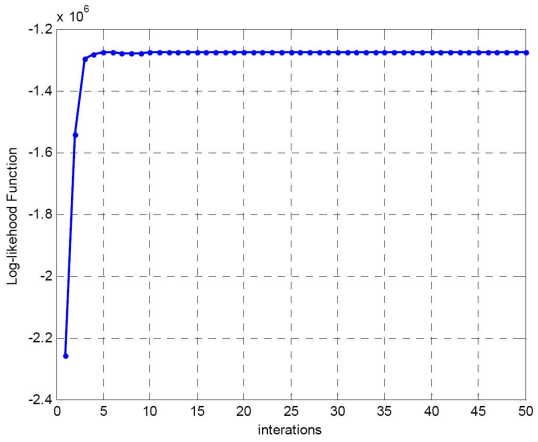

Figure 5.2: Maximization progress of the log-likelihood of the proposed method of the first experiment.

methods, we used the software implementation (for Linux environment, C++ Language) publicly available at http://mistis.inrialpes.fr/software/SEMMS.html. The source code for the SVFMM algorithm written in MATLAB environment can be downloaded from http://www.cs.uoi.gr/∼kblekas/sw/MAP segmentation.html. Standard GMM, CA-SVFMM and the proposed method were implemented using MATLAB in the Windows environment. All experiments were performed on a PC (Pentium 4, running at 3 GHz with 1GB of RAM) until convergence. Table 5.1 lists the computation time and the misclassification ratio for each of the aforementioned methods. As shown in Table 5.1, standard GMM takes the least amount of time for segmentation, while the slowest one is CA-SVFMM. Although the proposed method comes second (convergence after 4.9 seconds and 50 iterations, as shown in Figure 5.2) in terms of the speed, it has the lowest MCR, and demonstrates a higher degree of robustness with respect to noise.

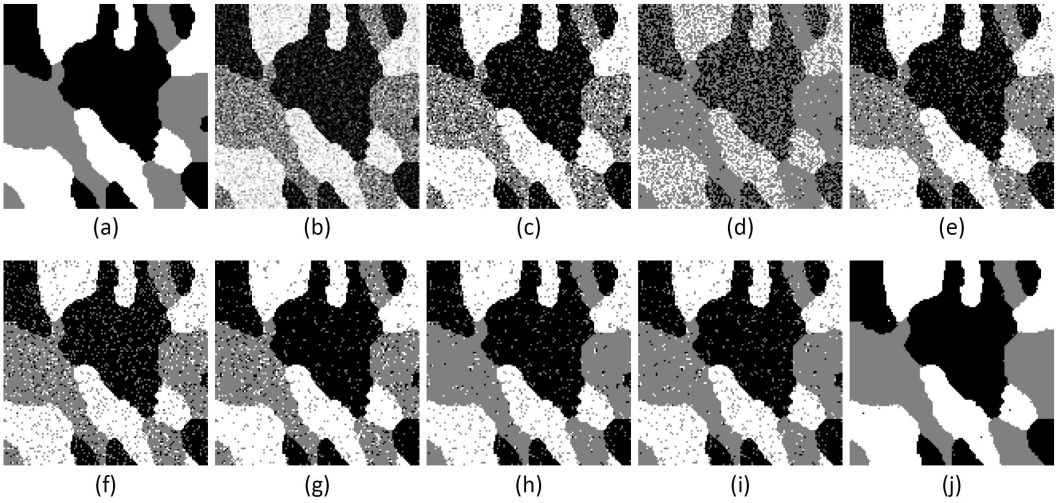

Figure 5.3: The second experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.05 variance), (c): K-means, (d): Standard GMM (MCR = 35.02%), (e): SVFMM (MCR = 12.01%), (f): CA–SVFMM (MCR = 11.16%), (g): ICM (MCR = 7.65%), (h): SIMF (MCR = 5.65%), (i) MEANF (MCR = 5.94%), (j): Proposed method (MCR = 1.05%).

In Figure 5.3, we show the segmentation results of a synthetic image corrupted by Gaussian noise. The original image with luminance values [0, 0.5, 1] is shown in Figure 5.3(a). The image in Figure 5.3(b) is obtained by adding Gaussian noise (0 mean, 0.05 variance) to the original image. Figure 5.3(d) to Figure 5.3(i), show the segmentation results for standard GMM, SVFMM, CA-SVFMM, ICM, SIMF, MEANF and the proposed method, respectively. Amongst these methods, the proposed method demonstrates a higher degree of robustness to a specified noise level. All algorithms were initialized using the K-means algorithm, as shown in Figure 5.3(c). Results in Table 5.2 confirm that the proposed method is quite fast with the lowest MCR.

The effect of varying noise level on the performance of different methods

Table 5.2: Computational cost (in seconds) comparison for the synthetic image in the second experiment.

Methods	Sta. GMM	SV FMM	CA–S VFMM	ICM	SIMF	MEA NF	our method
Time (sec)	1.2	117.1	147.1	78.3	108.2	97.7	3.1
MCR (%)	35.02	12.01	11.16	7.65	5.65	5.94	1.05

Table 5.3: Comparison of the proposed method with other methods in term of MCR (%), for the third experiment, in the presence of varying levels of noise.

Methods	Gaussian Noise (0 mean, var)		Mixed Noise: Salt & Pepper Noise (sp) + Gaussian Noise (0 mean, var)	
	var=0.03	var=0.05	sp=0.01, var=0.03	sp=0.02, var=0.05
Standard GMM	30.31	37.20	31.23	38.22
SVFMM	18.11	25.56	18.88	28.40
CA-SVFMM	17.73	25.49	18.25	28.58
ICM	5.90	27.32	11.65	29.88
SIMF	2.86	19.35	8.16	22.03
MEANF	2.70	17.65	7.14	21.94
Proposed method	0.21	0.36	0.35	0.42

is evaluated in the third experiment. A synthetic image with luminance values [0, 1/3, 2/3, 1] from Figure 5.4(a) is used. This orginal image is corrupted with varying levels of noise. The goal is to segment the corrupted image into four lables. The results are presented in Table 5.3. ICM method works well when an image is corrupted by low level of noise. However, the segmentation results are poor when the noise level is increased. The effect of noise on the performance of SIMF and MEANF methods is much less when we compared to

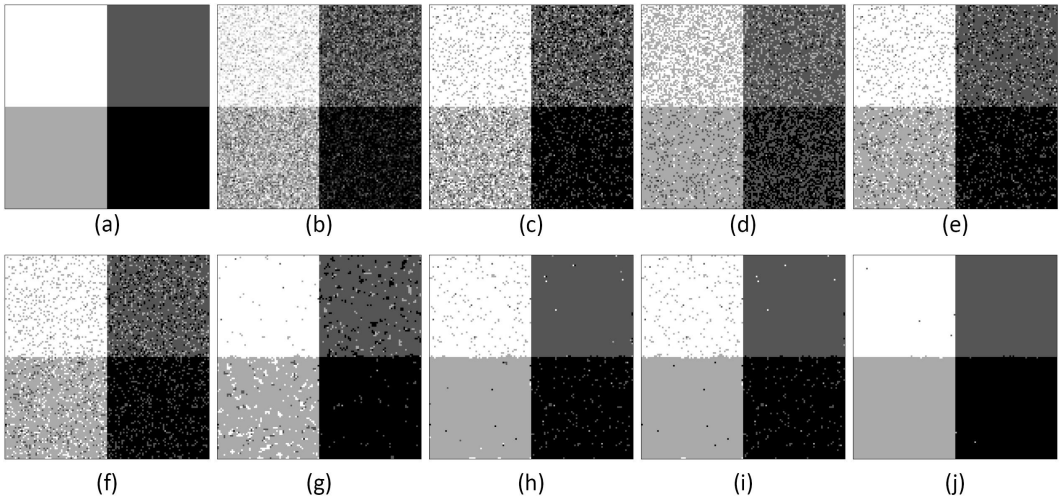

Figure 5.4: The third experiment (128x128 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.03 variance), (c): K-means, (d): Standard GMM (MCR = 30.31%), (e): SVFMM (MCR = 18.11%), (f): CA-SVFMM (MCR = 17.73%), (g): ICM (MCR = 5.90%), (h): SIMF (MCR = 2.86%), (i) MEANF (MCR = 2.70%), (j): Proposed method (MCR = 0.21%).

ICM. However, comparing to the proposed method, both SIMF and MEANF lose most of the sharpness and details in the segmented image as shown in Figure 5.4. In order to further test the accuracy of the proposed algorithm. The original image is corrupted with varying levels of mixed noise. Firstly, the original image is corrupted with salt & pepper noise, and later Gaussian noise is added. As shown in Table 5.3, the proposed method demonstrates a higher degree of robustness with respect to the given noise level.

In order to further test the accuracy of the proposed methods in Chapter 4 and Chapter 5, we generated a different image that contains two labels with luminance values (0, 1). In this experiment, the original image in Figure 5.5(a) is corrupted with a Gaussian noise (0 mean, 0.1 variance). Figure

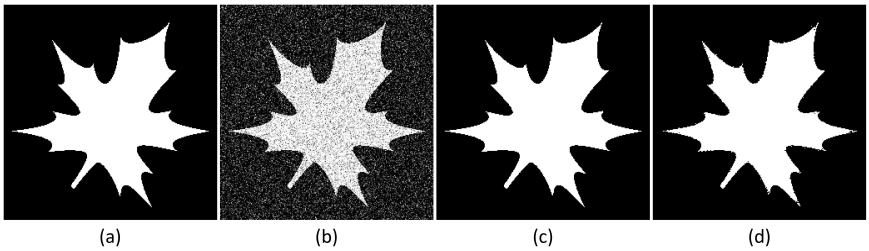

Figure 5.5: The fourth experiment (256x256 image resolution), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.1 variance), (c): Proposed method in Chapter 4 (MCR = 0.10%, time = 8.9s), (j): Proposed method in Chapter 4 (MCR = 0.22%, time = 3.7s).

5.5(c)-(d) show the segmentation the proposed methods in Chapter 4 and Chapter 5, respectively. We can see that the accuracy obtained by employing the proposed method in Chapter 4 (MCR=0.10%) is lower compared to the proposed method in Chapter 5. However, our method in Chapter 4 has high computational cost (8.9 seconds). As shown in Figure 5.5(d), the proposed method in Chapter 5 takes only 3.7 seconds.

5.3.2 Segmentation of Grayscale Natural Images

A real-world grayscale image from the Berkeley's image segmentation dataset [98] was used to compare the proposed algorithm with other algorithms. The objective is to segment the image into two labels: "buffalo" and "water". The main difficulty in this experiment is that the effect of noise on the label "water" is high. Figure 5.6(c) to Figure 5.6(g) show the results obtained by implementing SVFMM, CA-SVFMM, SIMF, MEANF, and the proposed method. The initialization for all these algorithms was carried out by using the standard GMM algorithm, as shown in Figure 5.6(b). As can be seen, the segmenta-

tion accuracy of SVFMM, CA-SVFMM, SIMF, MEANF methods, along the object boundaries is quite poor. In this experiment, only the proposed was successfully able to segment the image into "buffalo" and "water" regions. The maximization progress of the log-likelihood of the proposed methodis shown in Figure 5.6(h).

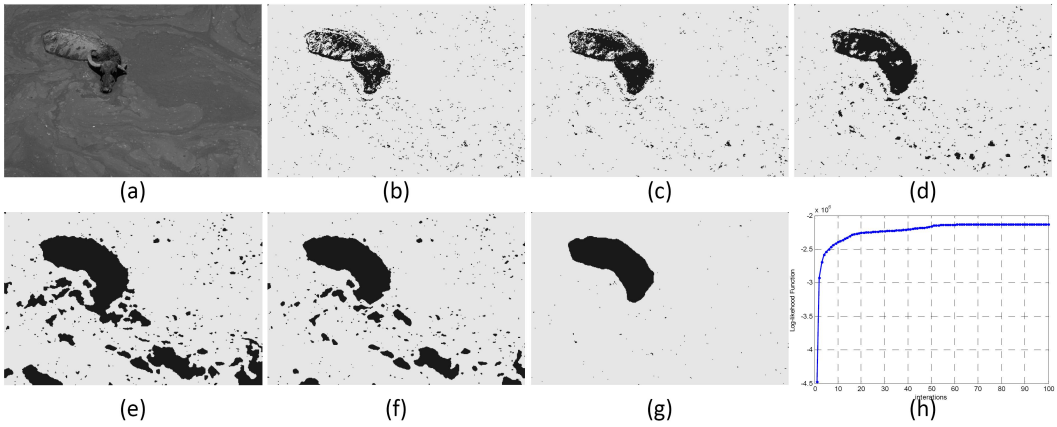

Figure 5.6: Grayscale natural image segmentation (80099), (a): original image, (b): Standard GMM, (c): SVFMM, (d): CA–SVFMM, (e): SIMF, (f) MEANF, (g): Proposed method, (h): Maximization progress of the log-likelihood of the proposed method of this experiment.

Another real world grayscale image is used to test the efficiency and effectiveness of our proposed algorithm. We observe that the effect of noise in the bottom left-hand side of this image is high. The objective is to segment the original image, as shown in Figure 5.7(a), into two labels. As can be seen from the results in Figure 5.7, Standard GMM, SVFMM, CA-SVFMM, SIMF, MEANF could not successfully segment this image, and there is a high degree of misclassification. Compared with these methods, the proposed method, as shown in Figure 5.7(g), successfully segments the image.

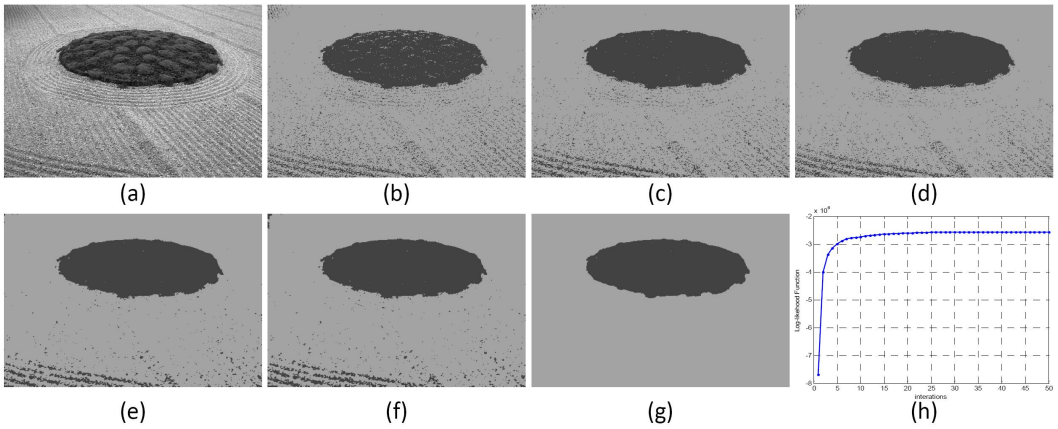

Figure 5.7: Grayscale natural image segmentation (86016), (a): original image, (b): Standard GMM, (c): SVFMM, (d): CA–SVFMM, (e): SIMF, (f) MEANF, (g): Proposed method, (h): Maximization progress of the log-likelihood of the proposed method.

A grayscale real-world image, shown in Figure 5.8(a), was segmented it into three labels: "road", "grass", and "tree". A low segmentation accuracy is obtained with standard GMM, as shown in Figure 5.8(b). This result is used as the initialization step for all the remaining methods. SVFMM and SIMF methods in Figure 5.8(c) and Figure 5.8(d) can produce a better segmentation, even though the effect of noise remains quite high. Compared with these methods, we find that the proposed method, shown in Figure 5.8(f) successfully segment all objects, and the effect of noise on the final segmented image is quite low.

5.3.3 Segmentation of Colored Images

In Figure 5.9, we show the segmentation results of a real-world color image. The orginal image, as shown in Figure 5.9(a), was used for segmentation into

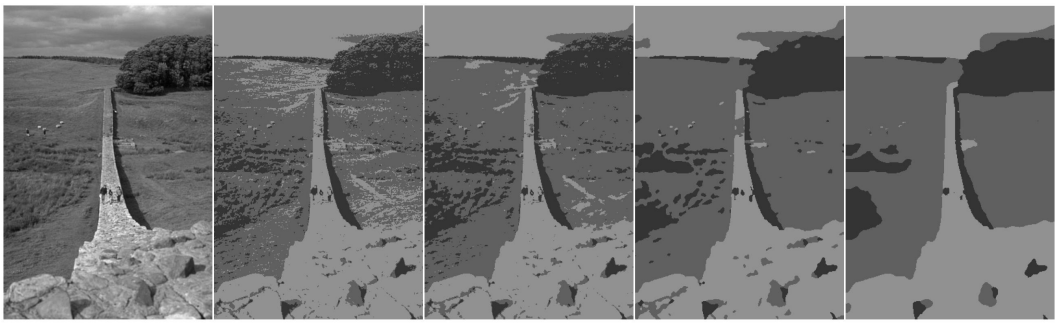

Figure 5.8: Grayscale natural image segmentation (374067), (a): original image, (b): Standard GMM, (c): SVFMM, (d): SIMF, (e): Proposed method.

three labels: "snow", "sky", and "others". The initialization for all algorithms was carried out by using the standard GMM as shown in Figure 5.9(b). Extraction accuracies of the SVFMM, CA-SVFMM, SIMF and MEANF methods are shown in Figure 5.9(c) to Figure 5.9(f). As shown in Figure 5.9, the segmentation accuracy for SIMF and MEANF methods is quite poor. CA-SVFMM can produce a better segmentation; however, the edge between the "snow" and the "sky" is lost. A closer inspection of the "sky" area indicates that a small portion of pixels have been misclassified. The proposed method in Figure 5.9(g), can better classify with more detail along the sharp edge between "snow" and the "sky", as compared with the CA-SVFMM method. Figure 5.9(h) shows the maximization progress of the log-likelihood of the proposed method in this experiment.

In order to further test the accuracy of the proposed algorithm, another real word color image is used. The image shown in Figure 5.10(b) is obtained by corrupting the original image in Figure 5.10(a) with Gaussian noise (0 mean, 0.0015 variance). The objective is to segment the image into three la-

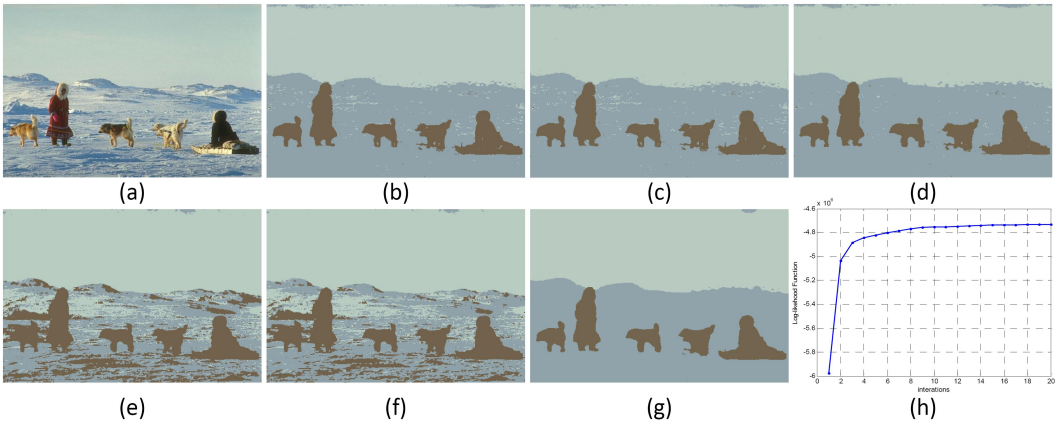

Figure 5.9: Color image segmentation (310007), (a): original image, (b): Standard GMM, (c): SVFMM, (d): CA–SVFMM, (e): SIMF, (f) MEANF, (g): Proposed method, (h): Maximization progress of the log-likelihood of the proposed method.

bels. Figure 5.10(c) presents the segmentation results obtained by employing the standard GMM method. The segmented image is used to initialize the remaining methods. A visual inspection of the results indicate that SVFMM in Figure 5.10(d), CA-SVFMM in Figure 5.10(e), ICM in Figure 5.10(f), and MEANF in Figure 5.10(g) are unable to segment the image successfully. The effect of noise on the final segmented image is highly noticeable. These methods lose much of the image sharpness and salient details in and around the "hands" and "legs" region of the girl in the image. Compared with these algorithms, the proposed method leads to a smoother segmentation. The effect of noise on the final segmented image, obtained by employing the proposed method is far less.

A set of real world color images are used to evaluate the performance of the proposed method against SVFMM, CA-SVFMM, ICM, SIMF, MEANF

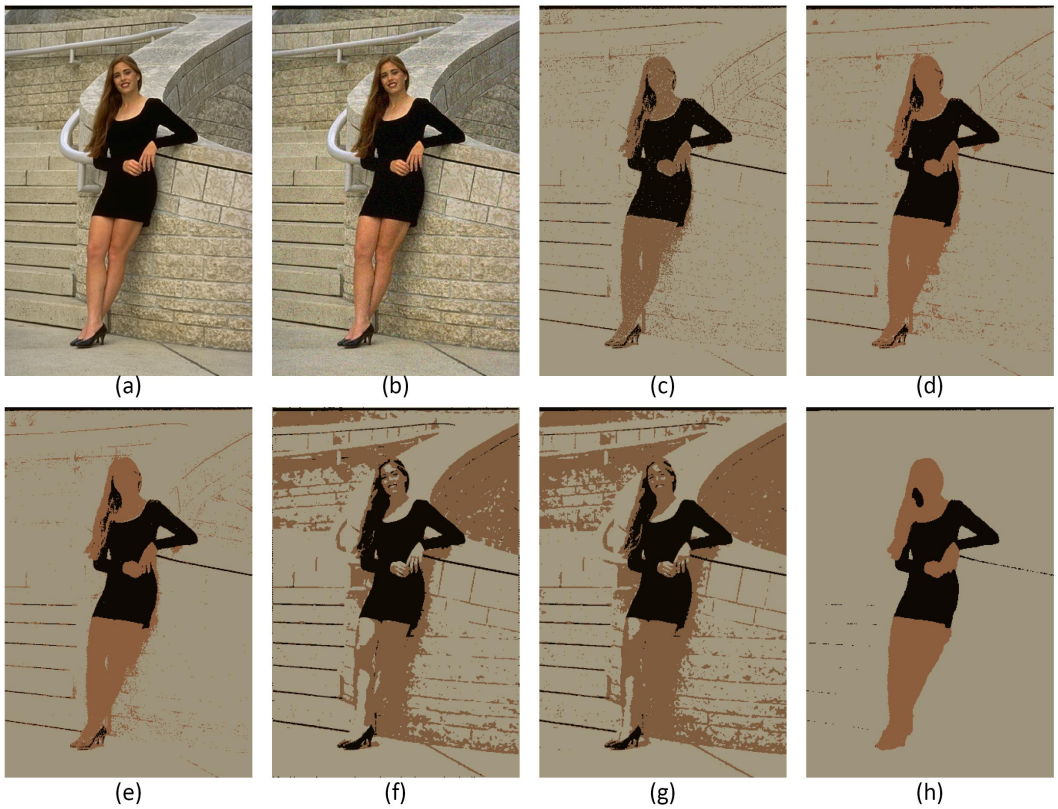

Figure 5.10: Color image segmentation (388016), (a): original image, (b): Corrupted original image with Gaussian noise (0 mean, 0.0015 variance), (c): Standard GMM, (d): SVFMM, (e): CA–SVFMM, (f): ICM, (g) MEANF, (h): Proposed method.

methods. Table 5.4 contains the cumulative results obtained for all methods, for the given set of real world images. As can be easily seen, on average, the proposed method outperforms other methods with a higher PR. Figure 5.6 shows some of the other real-world images used for segmentation by employing SVFMM, MEANF and the proposed method, respectively. The first row shows the original images, followed by the corresponding segmented images in the second, third, and the last row. Figure 5.11 clearly indicates that our proposed method achieves a better segmentation accuracy.

Table 5.4: Comparison of image segmentation results on Berkeley's color image segmentation dataset: Probabilistic Rand (PR) Index.

Image	labels	SVFMM	CA-SVFMM	ICM	SIMF	MEANF	Proposed method
86000	4	0.811	0.814	0.794	0.773	0.785	0.825
101085	2	0.598	0.600	0.572	0.566	0.569	0.603
100080	5	0.802	0.801	0.778	0.768	0.767	0.811
105053	2	0.778	0.782	0.756	0.650	0.689	0.817
241004	5	0.839	0.841	0.838	0.841	0.842	0.841
175043	2	0.800	0.801	0.798	0.713	0.722	0.802
374067	4	0.815	0.814	0.801	0.781	0.783	0.818
24063	3	0.840	0.839	0.838	0.839	0.839	0.841
106025	4	0.832	0.833	0.824	0.796	0.803	0.844
147091	3	0.813	0.815	0.770	0.773	0.775	0.824
277095	3	0.811	0.812	0.773	0.637	0.690	0.814
113009	3	0.671	0.670	0.645	0.614	0.622	0.676
296007	3	0.839	0.840	0.841	0.832	0.841	0.841
189080	4	0.861	0.862	0.814	0.805	0.807	0.867
38092	4	0.871	0.873	0.820	0.820	0.817	0.879
41004	3	0.908	0.909	0.891	0.848	0.872	0.917
Mean	-	0.806	0.807	0.785	0.754	0.764	0.814

5.4 Conclusions

We have presented a new mixture model for image segmentation that incorporates the spatial relationships based on MRF. Compared with other MRF based mixture models, our proposed method directly applies the EM algorithm to optimize the parameters, making it simple, fast, and easy to implement. The proposed method has been tested with many synthetic and real world grayscale and colored images, thereby demonstrating the excellent performance in noisy conditions, as compared to other mixture model based approaches.

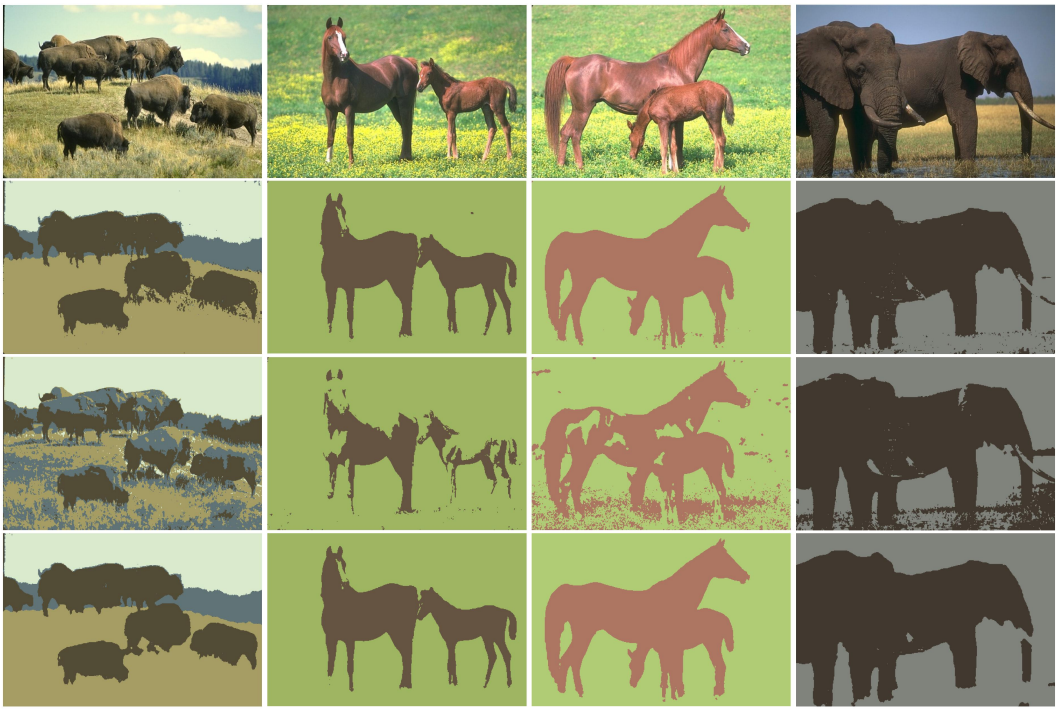

Figure 5.11: (first row): original image, (second row): SVFMM, (third row): MEANF, (last row): Proposed method.

Chapter 6

Conclusions

6.1 Summary of Main Contributions

In this thesis, new fully unsupervised image segmentation algorithms based on mixture model with spatial neighborhood relationships have been described. The proposed methods have been tested with both synthetic and real images, and have been shown to be robust with respect to noise, efficient with respect to the number of parameters used, and sufficiently accurate with respect to the classification results.

A review of the various model-based techniques used to achieve segmentation is described. In Chapter 2, the first group of model-based techniques is described, beginning with the using of standard GMM to solve the unsupervised segmentation problem. The advantages and disadvantages of the standard GMM are then discussed. In order to estimate the parameters of the model, various techniques on maximizing their likelihood are described,

beginning with the EM algorithm, then continuing with the gradient-based optimization techniques.

The second group of model-based techniques for unsupervised segmentation is described in Chapter 3. These methods are based on the MRF to take into account the spatial correlation between the neighboring pixels and to reduce the sensitivity of the segmentation result with respect to noise. In this chapter, two types of models based on MRF are presented. In the first type, mixture models with MRF have been employed for pixel labels. In order to take into account the spatial correlation, an MRF model in the second type is used to model the joint distribution of the priors of each pixel label, instead of the joint distribution of the pixel labels as in first type. The main differences between the first and the second group of model-based techniques are also discussed in this chapter.

In Chapter 4, we propose an extension of the standard GMM for image segmentation, which utilizes a novel approach to incorporate the spatial relationships between neighboring pixels into the standard GMM. The proposed model is easy to implement and compared with the existing MRF models, requires a fewer parameters. We also propose a new method to estimate the model parameters in order to minimize the higher bound on the data negative log-likelihood, based on the gradient method. Experimental results obtained on noisy synthetic and real world grayscale images demonstrate the robustness, accuracy and effectiveness of the proposed model in image segmentation, as compared to other methods based on standard GMM and MRF models.

In Chapter 5, another way to incorporate spatial information between the neighboring pixels into the GMM based on MRF is proposed. In comparison to other mixture models that are complex and computationally expensive, the proposed method is robust and fast to implement. In mixture models based on MRF, the M-step of the EM algorithm cannot be directly applied to the prior distribution π_{ij} for maximization of the log-likelihood with respect to the corresponding parameters. Compared with these models, our proposed method directly applies the EM algorithm to optimize the parameters, which makes it much simpler. Finally, our approach is used to segment many synthetic and real-world grayscale and colored images with excellent results.

6.2 Future Directions

Several extensions are possible for future work. In this thesis, the number of labels (K) is currently set by the user based on prior knowledge. This is an open question and remains the subject of current research. One possible improvement on the current models is to investigate the ways to automatically optimize this parameter.

Another limitation is that a constant β is used throughout the image and for every label. A constant value of β reduces the impact of noise in homogeneous tissues but negatively affects segmentation along the border of two tissues. In the highly noisy image, it can erase the detail of the boundary of two tissues. One possible solution to overcome this problem is to use a different value of β throughout the image.

Another possible extension of this work is to study the use of non-linear smoothing filter to improve the quality of image segmentation. In the context of this thesis, one of the simplest linear filters is used. Our spatial process is based on the mean smoothing filter, where we replace each pixel value in an image with the average value of its neighbors.

Recently, to improve the robustness of the algorithm to outliers, finite Student's-t mixture model (SMM) has been proposed. The main advantage of the Student's-t distribution is that it is heavily tailed than Gaussian, and hence finite mixture model of the longertailed multivariate Student's-t distribution provides a much more robust approach to the GMM. It has proven to be quite effective for image segmentation. However, in the existing SMM, the prior distribution does not depend on the pixel index and has the same value for all pixels. Moreover, it does not take into account the spatial constraints in an image. Additionally, in order to estimate the model parameters by adopting the EM algorithm, the Student's-t distribution in the previous SMM is represented as an infinite mixture of scaled Gaussians, which corresponds to an increase in the algorithms' complexity. A combination between finite Student's-t mixture model with our methods to develop a new model could be considered in future work.

Segmentation of medical images such as computed tomography and magnetic resonance imaging is an important diagnostic imaging technique. Accurate segmentation of these images offers an opportunity to provide more information about the objects and their boundaries for clinical investigation.

However, fully automatic segmentation of medical images into a number of non-overlapping regions i.e., gray matter, white matter, cerebrospinal fluid, and background is difficult because they are frequently corrupted by high levels of noise, and poor contrast along boundaries. The proposed models could be applied for segmenting medical images in the future.

References

[1] Jian X., Thanh N. M., and Wu Q. M. J., "FPGA Implementation of Blob Recognition," *IEEE Canadian Conference on Computer and Robot Vision*, Newfoundland, 2011.

[2] Ahuja S., Thanh N. M., and Wu Q. M. J., "Real–Time 3–D Location, Orientation and Tracking of Tools in a Work–Cell," *Auto 21*, May 2009.

[3] Thanh N. M., Ahuja S., Wu Q. M. J., "Real–time 3D Location, Orientation and Tracking of Tools in a Work-Cell using a Single Camera," Computer Vision and Sensing Systems Laboratory, University of Windsor, 2009,(Internal Paper).

[4] Schroeter P., Vesin J. M., Langenberger T., and Meuli R., "Robust parameter estimation of intensity distributions for brain magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 172–186, 1998.

[5] Ashburner J., and Friston K. J., "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.

[6] Ahmed M. N., Yamany S. M., Mohamed N., Farag A. A., and Moriarty T., "A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.

[7] Thanh N. M., and Wu Q. M. J., "Robust Student's–t Mixture Model with Spatial Constraints and Its Application in Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, Accepted, 2011.

[8] Thanh N. M., and Wu Q. M. J., "A Combination of Positive and Negative Fuzzy Rules for Image Classification Problem," *IEEE International Conference on Machine Learning and Applications*, pp. 741–746, 2008.

[9] Thanh N. M., Sarwer, M. G., and Wu Q. M. J., "A New Probability Neural Network for Image Classification Problem," *The 10th IASTED International Conference, Signal and Image Processing*, pp. 1–6, 2008.

[10] Sarwer M. G, Thanh N. M., and Wu Q. M. J., "Fast Motion Estimation of H.264/AVC by Adaptive Early Termination," *The 10th IASTED International Conference, Signal and Image Processing*, pp. 140–145, 2008.

[11] Thanh N. M., and Wu Q. M. J., "A Fuzzy Logic Model based Markov Random Field for Medical Image Segmentation," Submitted 2011.

[12] Thanh N. M., Ahuja S., and Wu Q. M. J., "A Real–Time Ellipse Detection Based on Edge Grouping," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3280–3286, 2009.

[13] Antoni B. C., and Nuno V., "Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 909–926, 2008.

[14] Antoni B. C., Zhang–Sheng J. L., and Nuno V., "Privacy preserving crowd monitoring: Counting people without people models or tracking," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7, 2008.

[15] Fionn M., Adrian E. R., and Starck J. L., "Bayesian inference for multiband image segmentation via model-based cluster trees," *Image and Vision Computing*, vol. 23, no. 6, pp. 587–596, 2005.

[16] Krak N. C. and Boellaard R., Hoekstra O. S., Twisk J. W., Hoekstra C. J., and Lammertsma A. A., "Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 32, no. 10, pp. 294–301, 2005.

[17] Erdi Y. E., Mawlawi O., Larson S. M., Imbriaco M., Yeung H., Finn R., and Humm J. L., "Segmentation of lung lesion volume by adaptative positron emission tomography image thresholding," *Cancer*, vol. 80, no. 12 Suppl, pp. 2505–2509, 1997.

[18] Jarritt H., Carson K., Hounsel A. R., and Visvikis D., "The role of PET/CT scanning in radiotherapy planning," *British Journal of Radiology*, vol. 79, pp. S27–S35, 2006.

[19] Nestle U., Kremp S., Schaefer–Schuler A., Sebastian–Welch C., Hellwig D., Rube C., and Kirsch C. M., "Comparison of Different Methods for Delineation of 18F–FDG PETPositive Tissue for Target Volume Definition in Radiotherapy of Patients with NonSmall Cell Lung Cancer," *Journal of Nuclear Medicine*, vol. 46, no. 8, pp. 1342–1348, 2005.

[20] Chow C. K., and Kaneko T., "Automatic boundary detection of the left–ventricle from cineangiograms," *Computers and Biomedical Research*, vol. 5, no. 4, pp. 388–410, 1972.

[21] Nakagawa Y., and Rosenfeld A., "Some experiments on variable thresholding," *Pattern Recognition*, vol. 5, no. 4, pp. 191–204, 1979.

[22] Yanowitz S. D., and Bruckstein A. M., "A new method for image segmentation," *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 1, pp. 82–95, 1989.

[23] Pal N., and Pal S., "A review of image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, pp. 1277–1294, 1993.

[24] Bors A. G., and Pitas I., "Optical Flow Estimation and Moving Object Segmentation Based on Median Radial Basis Function Network," *IEEE Transactions on Image Processing*, vol. 7, no. 5, pp. 693–702, 1998.

[25] Manjunath B. S., Simchony T., and Chellappa R., "Stochastic and deterministic network for texture segmentation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1039–1049, 1990.

[26] Ghosh A., Pal N. R., and Pal S. K., "Self–organization for object extraction using multilayer neural networks and fuzziness measure," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 1, pp. 54–68, 1993.

[27] Amartur S. C., Piraino D., and Takefuji Y., "Optimization neural networks for the segmentation of magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 11, pp. 215–220, 1992.

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

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

[30] Comaniciu D., and Meer P., "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[31] Carreira–Perpinan M. A., "Fast nonparametric clustering with Gaussian blurring mean-shift," *International Conference on Machine Learning*, pp. 153–160, 2006.

[32] Linda G. S., and George C. S., "*Computer Vision*", New Jersey, Prentice–Hall, pp. 279325, 2001.

[33] Tabb M., and Ahuja N., "Unsupervised multiscale image segmentation by integrated edge and region detection," *IEEE Transactions on Image Processing*, vol. 6, no. 5, pp. 642–655, 1997.

[34] Chen C. H., Pau L. F., and Wang P. S. P., "*The Handbook of Pattern Recognition and Computer Vision*", 2nd Edition, World Scientific Publishing Co., 1998.

[35] Cai W., Chen S., and Zhang D., "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.

[36] Thanh N. M., and Wu Q. M. J., "A Fuzzy C-Means based Spatial Pixel and Membership Relationships for image segmentation," *IEEE Canadian Conference on Computer and Robot Vision*, Newfoundland, 2011.

[37] Jain A. K., and Dubes R. C., "*Algorithms for Clustering Data*", Englewood Cliffs, NJ: PrenticeHall, 1988.

[38] Bezdek J. C., "*Pattern Recognition with Fuzzy Objective Function Algorithms*", New York: Plenum, 1981.

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

[40] Feng Y.Q., and Chen W.F., "Brain MR image segmentation using fuzzy clustering with spatial constraints based on Markov random field theory," *Lecture Notes in Computer Science*, vol. 3150, pp. 188–195, 2004.

[41] Pham D. L., and Prince J. L., "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recognition Letters*, vol. 20, no. 1, pp. 57–68, 1999.

[42] Krinidis S., and Chatzis V., "A Robust Fuzzy Local Information C–means Clustering Algorithm," *IEEE Transactions on Image Processing*, no. 19, vol. 5, pp. 1328–1337, 2004.

[43] Liu J., and Yang Y. H., "Multiresolution color image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 7, pp. 689–700, 1994.

[44] Stanford D. C., "*Fast automatic unsupervised image segmentation and curve detection in spatial point pattern*", Ph.D. dissertation, Department of Statistics, University of Washington, Seattle, 1999.

[45] Langan D. A., Modestino J. W., and Jun Z., "Cluster validation for unsupervised stochastic model-based image segmentation," *IEEE Transactions on Image Processing*, vol. 7, no. 2, pp. 180–195, 1998.

[46] Titterington D. M., Smith A. F. M., and Makov U. E., "*Statistical Analysis of Finite Mixture Distributions*", Hoboken, NJ: Wiley, 1985.

[47] Jain A. K., Duin R. P. W., and Mao J. C., "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[48] Redner R. A. and Walker H. F., "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195239, 1984.

[49] Diplaros A., Gevers T., and Vlassis N., "Skin detection using the EM algorithm with spatial constraints," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 3071–3075, 2004.

[50] Haim P., Joseph F., and Ian J., "A study of Gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recognition*, vol. 39, no. 4, pp. 695–706, 2006.

[51] Carson C., Belongie S., Greenspan H., and Malik J., "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, 2002.

[52] McLachlan G. J., and Krishnan T., "*The EM Algorithm and Extensions*", Wiley Series in Probability and Statistics, New York, Wiley, 1997.

[53] Hebert T. J., and Lu K., "Expectation maximization algorithms, null spaces, and MAP image restoration," *IEEE Transactions on Image Processing*, vol. 4, pp. 1084–1095, 1995.

[54] Hebert T., and Leahy R., "A generalized EM algorithm for 3-d bayesian reconstruction from poisson data using Gibbs priors," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 194–202, 1989.

[55] Fessler J. A., and Hero A. O. III., "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

[56] Figueiredo M. A. T., and Jain A. K., "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.

[57] Dempster P., Laird N. M., and Rubin D. B., "Maximum likelihood from incomplete data via EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[58] Sanjay G. S., and Hebert T. J., "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.

[59] Choi H. S., Haynor D. R., and Kim Y., "Partial volume tissue classification of multichannel magnetic resonance images–A mixel model," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 395–407, 1991.

[60] Chellappa R., and Jain A., "*Markov Random Fields: Theory and Application*" Boston, MA: Academic, 1993.

[61] Ibanez M. V., and Simo A., "Parameter estimation in Markov random field image modeling with imperfect observations: A comparative study," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2377–2389, 2003.

[62] Santago P., and Gage H. D., "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 566–574, Sep. 1993.

[63] Forbes F., and Peyrard N., "Hidden Markov random field model selection criteria based on mean field-like approximations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1089–1101, 2003.

[64] Celeux G., Forbes F., and Peyrard N., "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131–144, 2003.

[65] Chalmond B., "An iterative Gibbsian technique for reconstruction of mary images," *Pattern Recognition*, vol. 22, no. 6, pp. 747–761, 1989.

[66] Zhang J., "The mean field theory in EM procedures for Markov random fields," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2570–2583, 1992.

[67] Qian W., and Titterington D. M., "Estimation of parameters in hidden Markov models," *Philosophical transactions of the Royal Society*, vol. 337, no. 1647, pp. 407–428, 1991.

[68] Laferte J. M., Perez P., and Heitz F., "Discrete Markov image modeling and inference on the quadtree," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 390–404, 2000.

[69] Robinson M., Azimi-Sadjadi M., and Salazar J., "A temporally adaptive classifier for multispectral imagery," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 159–165, 2004.

[70] Fjortoft R., Delignon Y., Pieczynski W., Sigelle M., and Tupin F., "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 675–686, 2003.

[71] Marroquin J. L., Santana E. A., and Botello S., "Hidden Markov measure field models for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1380–1387, 2003.

[72] Blekas K., Likas A., Galatsanos N. P., and Lagaris I. E., "A spatially constrained mixture model for image segmentation," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 494–498, 2005.

[73] Nikou C., Galatsanos N., and Likas A., "A class–adaptive spatially variant mixture model for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1121–1130, 2007.

[74] Nikou C., Likas A., and Galatsanos N. P., "A Bayesian Framework for Image Segmentation with Spatially Varying Mixtures," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2278–2289, 2010.

[75] McLachlan G., and Peel D., "*Finite Mixture Models*", New York, Wiley, 2000.

[76] Nasios N., and Bors A.G., "Variational learning for Gaussian mixture models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 36, no. 4, pp. 849–862, 2006.

[77] Roweis S. T., Saul L. K., and Hinton G. E., "Global coordination of local linear models," *Advances in Neural Information Processing Systems 14*, Dietterich T. G., Becker S., and Ghahramani Z., Eds. Cambridge, MA: MIT Press, pp. 889–896, 2002.

[78] Verbeek J. J., Vlassis N., and Krose B., "Self-organizing mixture models," *Neurocomputing*, vol. 63, pp. 99123, 2005.

[79] Bishop C. M., "*Neural Networks for Pattern Recognition*", Oxford University Press, Walton Street, Oxford, 1995.

[80] Bishop C. M., "*Pattern Recognition and Machine Learning*", Springer, 2006.

[81] Rudin W., "*Real and Complex Analysis*", Third Edition, McGraw-Hill, pp. 62–65, 1987.

[82] Thanh M. N., and Wu Q. M. J., "Maximum likelihood neural network based on the correlation among neighboring pixels for noisy image segmentation," *IEEE International Conference on Image Processing*, pp. 3020–3023, 2008.

[83] Thanh M. N., Wu Q. M. J., and Ahuja S., "An Extension of the Standard Mixture Model for Image Segmentation," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1326–1338, 2010.

[84] Thanh N. M., and Wu Q. M. J., "Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation," Submitted 2011.

[85] Jang J. R., Sun C. T., and Mizutani E., "*Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*", Prentice–Hall, pp. 133–134, 1997.

[86] Seber G. A. F., "*Multivariate Observations*", Wiley, New York, pp. 482–488, 1984.

[87] Dang M., and Govaert G., "Spatial fuzzy clustering using EM and Markov random fields," *International Journal of System Research and Information Science*, vol. 8, no. 4, pp. 183–202, 1998.

[88] Besag J., "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[89] Chatzis S. P., and Varvarigou T. A., "A Fuzzy Clustering Approach Toward Hidden Markov Random Field Models for Enhanced Spatially Constrained Image Segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1351–1361, 2008.

[90] Sun J., and Gu D., "Bayesian image segmentation based on an inhomogeneous hidden Markov random field," *IEEE International Conference on Pattern Recognition*, vol. 1, pp. 596–599, 2004.

[91] Zhang Y., Brady M., and Smith S., "Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation–Maximization Algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[92] Geman S., and Geman D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI–6, no. 6, pp. 721–741, 1984.

[93] Unnikrishnan R., Pantofaru C., and Hebert M., "A measure for objective evaluation of image segmentation algorithms," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 3, pp. 34–41, 2005.

[94] Dice L. R., "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[95] Clifford P., "Markov random fields in statistics," *Disorder in Physical Systems, A Volume in Honour of John M. Hammersley on the Occasion of his 70th Birthday*, G. Grimmett and D. Welsh, Oxford, UK, Clarendon, 1990.

[96] Besag J., "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, vol. 36, no. 2, pp. 192–326, 1974.

[97] Besag J., "Statistical analysis of non-lattice data," *Journal of the Royal Statistical Society, Series D*, vol. 24, no. 3, pp. 179–195, 1975.

[98] Martin D., Fowlkes C., Tal D., and Malik J., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *International Conference on Computer Vision*, vol. 2, pp. 416–423, 2001.

[99] Li S. Z., "*Markov Random Field Modeling in Image Analysis*", SpringerVerlag, 2009.

[100] Diplaros A., Vlassis N., and Gevers T., "A spatially constrained generative model and an EM algorithm for image segmentation," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 798–808, 2007.

[101] Santago P., and Gage H.D., "Statistical models of partial volume effect," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1531–1540, 1995.

[102] Green P. J., "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm," *IEEE Transactions on Medical Imaging*, vol. 9, no. 1, pp. 84–93, 1990.

[103] Sclove S. L., "Application of the conditional population–mixture model to image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI–5, no. 4, pp. 428–433, 1983.

[104] Liang Z., MacFall J. R., and Harrington D. P., "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Transactions on Medical Imaging*, vol. 13, no. 3, pp. 441–449, 1994.

[105] Liang Z., Jaszczak R. J., and Coleman R. E., "On reconstruction and segmentation of piecewise continuous images," *Information Processing in Medical Imaging*, vol. 12, pp. 94–104, 1991.

[106] Sfikas G., Nikou C., and Galatsanos N., "Robust Image Segmentation with Mixtures of Student's t-Distributions," *IEEE International Conference on Image Processing*, vol. 1, pp. 273–276, 2007.

Appendix A

We apply the gradient descent technique to update the parameters $\Xi = \{\Theta, \Pi\} = (\mu_j, \sigma_j, c_j, b_j, \alpha)$, j=(1,2,...,K). The error function in Eq.(4.10) can be re-written as:

$$E(\Theta^{(t)}, \Pi^{(t)} | \Theta^{(t+1)}, \Pi^{(t+1)}) = -\sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)} \left\{\log \pi_{ij} + \log \Phi(\mathrm{x}_i|\Theta_j)\right\} \qquad \text{(A.1)}$$

where

$$\pi_{ij} = \left[\sum_{m\in\mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)\right]^{\alpha} \Bigg/ \sum_{k=1}^{K}\left[\sum_{m\in\mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_k)^2}{2b_k^2}\right)\right]^{\alpha}$$

and

$$\Phi(\mathrm{x}_i|\Theta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\mathrm{x}_i - \mu_j)^2}{2\sigma_j^2}\right)$$

as shown in Eq.(4.3) and Eq.(2.23), respectively. The derivative of E with respect to μ_j is given by:

$$\frac{\partial E}{\partial \mu_j} = -\sum_{i=1}^{N} z_{ij}^{(t)} \frac{\mathrm{x}_i - \mu_j}{\sigma_j^2} \qquad \text{(A.2)}$$

and similarly, the derivative of E with respect to σ_j, is given by:

$$\frac{\partial E}{\partial \sigma_j} = -\sum_{i=1}^{N} z_{ij}^{(t)} \left(-\frac{1}{\sigma_j} + \frac{(\mathrm{x}_i - \mu_j)^2}{\sigma_j^3}\right) \qquad \text{(A.3)}$$

The derivative of E with respect to c_j can be expressed as:

$$\frac{\partial E}{\partial c_j} = -\sum_{i=1}^{N} z_{ij}^{(t)} \left\{ \frac{\alpha \sum\limits_{m \in \mathcal{N}i} \frac{(\mathrm{x}_m - c_j)}{b_j^2} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)}{\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)} \right\} + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik}^{(t)} \left\{ \frac{\alpha \left[\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)\right]^{\alpha-1} \sum\limits_{m \in \mathcal{N}i} \frac{(\mathrm{x}_m - c_j)}{b_j^2} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)}{\sum\limits_{l=1}^{K} \left[\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_l)^2}{2b_l^2}\right)\right]^{\alpha}} \right\} \tag{A.4}$$

Now, considering the derivative of the term E with respect to b_j, we have:

$$\frac{\partial E}{\partial b_j} = -\sum_{i=1}^{N} z_{ij}^{(t)} \left\{ \frac{\alpha \sum\limits_{m \in \mathcal{N}i} \frac{(\mathrm{x}_m - c_j)^2}{b_j^3} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)}{\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)} \right\} + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik}^{(t)} \left\{ \frac{\alpha \left[\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)\right]^{\alpha-1} \sum\limits_{m \in \mathcal{N}i} \frac{(\mathrm{x}_m - c_j)^2}{b_j^3} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)}{\sum\limits_{l=1}^{K} \left[\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_l)^2}{2b_l^2}\right)\right]^{\alpha}} \right\} \tag{A.5}$$

The derivative of E with respect to α is given by:

$$\frac{\partial E}{\partial \alpha} = -\sum_{i=1}^{N} \sum_{j}^{K} z_{ij}^{(t)} \log\left(\sum_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_j)^2}{2b_j^2}\right)\right) + \sum_{i=1}^{N} \sum_{j}^{K} z_{ij}^{(t)} \left\{ \frac{\sum\limits_{k=1}^{K} \left(\left[\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_k)^2}{2b_k^2}\right)\right]^{\alpha} \log\left(\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_k)^2}{2b_k^2}\right)\right)\right)}{\sum\limits_{l=1}^{K} \left[\sum\limits_{m \in \mathcal{N}i} \exp\left(-\frac{(\mathrm{x}_m - c_l)^2}{2b_l^2}\right)\right]^{\alpha}} \right\} \tag{A.6}$$

Publications

[1] 1. Thanh N. M., and Wu Q. M. J., "*Fuzzy System with Positive and Negative Rules*", Machine Learning, ISBN: 978-953-307-033-9, INTECH, 2010.

[2] Thanh N. M., and Wu Q. M. J., "Robust Student's-t Mixture Model with Spatial Constraints and Its Application in Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, Accepted, 2011.

[3] Thanh N. M., and Wu Q. M. J., "Gaussian Mixture Model based Spatial Neighborhood Relationships for Pixel Labeling Problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, Accepted, 2011.

[4] Thanh N. M., and Wu Q. M. J., Ahuja S., "An Extension of the Standard Mixture Model for Image Segmentation," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1326–1338, 2010.

[5] Thanh N. M., and Wu Q. M. J., "Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation," Submitted 2011.

[6] Thanh N. M., and Wu Q. M. J., "A Fuzzy Logic Model based Markov Random Field for Medical Image Segmentation," Submitted 2011.

[7] Thanh N. M., and Wu Q. M. J., "Dirichlet Bayesian Mixture Model: Application to Image Segmentation," Submitted 2010.

[8] Thanh N. M., and Wu, Q. M. J., "An Extension of Standard Fuzzy C-Means Clustering for Image Segmentation," Submitted 2010.

[9] Thanh N. M., and Wu Q. M. J., "A Fuzzy C-Means based Spatial Pixel and Membership Relationships for image segmentation", *IEEE Canadian Conference on Computer and Robot Vision*, Newfoundland, 2011.

[10] Jian X., Thanh N. M., and Wu Q. M. J., "FPGA Implementation of Blob Recognition," *IEEE Canadian Conference on Computer and Robot Vision*, 2011.

[11] Thanh N. M., Ahuja S., and Wu Q. M. J., "A Real-Time Ellipse Detection Based on Edge Grouping," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3280–3286, 2009.

[12] Ahuja S., Thanh N. M., and Wu Q. M. J., "Real-Time 3-D Location, Orientation and Tracking of Tools in a Work-Cell," *Auto 21*, May 2009.

[13] Thanh N. M., and Wu Q. M. J., "Maximum Likelihood Neural Network based on the Correlation among Neighboring Pixels for Noisy Image Segmentation," *IEEE International Conference on Image Processing*, pp. 3020–3023, 2008.

[14] Thanh N. M., Ahuja S., and Wu Q. M. J., "Real-time 3D Location, Orientation and Tracking of Tools in a Work-Cell using a Single Camera," Computer Vision and Sensing Systems Laboratory, University of Windsor, 2009,(Internal Paper).

[15] Thanh N. M., Sarwer, M. G., and Wu Q. M. J., "A New Probability Neural Network for Image Classification Problem," *The 10th IASTED International Conference, Signal and Image Processing*, pp. 1–6, 2008.

[16] Sarwer M. G, Thanh N. M., and Wu Q. M. J., "Fast Motion Estimation of H.264/AVC by Adaptive Early Termination," *The 10th IASTED International Conference, Signal and Image Processing*, pp. 140–145, 2008.

[17] Thanh N. M., and Wu Q. M. J., "A Combination of Positive and Negative Fuzzy Rules for Image Classification Problem," *IEEE International Conference on Machine Learning and Applications*, pp. 741–746, 2008.

Vita Auctoris

Thanh Minh Nguyen was born in 1981 in Vietnam. He received the B.E. degree in electrical and electronic engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in April 2004, and the MA.Sc. degree from the Department of Electrical Engineering, Dayeh University, Changhua, Taiwan, in June 2006. He received a Doctor of Philosophy degree from the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada, in August 2011. His current research interests include computer vision and pattern recognition.