IMAGE LABELING BY ENERGY MINIMIZATION WITH APPEARANCE AND SHAPE PRIORS

By

Asem Mohamed Ahmed Ali

A Dissertation
Submitted to the Faculty of the Graduate School of the University of Louisville in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering
University of Louisville
Louisville, Kentucky

May 2008
IMAGE LABELING BY ENERGY MINIMIZATION WITH APPEARANCE AND SHAPE PRIORS

By

Asem Mohamed Ahmed Ali

A Dissertation Approved on

by the Following Reading and Examination Committee:

________________________
Aly A. Farag, Ph.D., Dissertation Director

________________________
Thomas L. Starr, Ph.D.

________________________
Tamer Inanc, Ph.D.

________________________
Mohamed N. Ahmed, Ph.D.

________________________
Prasanna Sahoo, Ph.D.

________________________
Georgy L. Gemil’farb, Ph.D.
DEDICATION

To the greatest women in my life, my mother, my wife, and my daughter.
& to the soul of my father.
ACKNOWLEDGMENTS

First of all, my deepest thanks are due to Almighty God, the merciful, the compassionate for uncountable gifts given to me.

I would like to express my deepest gratitude to my advisor, Prof. Aly A. Farag, for giving me the opportunity to be a member in his research group, for his continuous encouragement, and for his support over the course of this work. He provided a very rich working environment with many opportunities to develop new ideas, work in promising applications, get experience in diverse areas, and meet well-known people in the field.

I would like to thank Prof. Thomas L. Starr for giving me the opportunity to be a member in his research group in the face recognition project, and for useful discussions.

I would like to thank Prof. Georgy Gimel’farb for useful discussions and assistance. He has never hesitated to share his experience in Markov Gibbs random field, image processing and computer vision fields. He has been a good teacher.

I would like to thank Dr. Tamer Inanc, Dr. Mohamed Ahmed, Prof. Prasanna Sahoo, and Dr. Moumen Ahmed for agreeing to be on my dissertation committee, for the useful consultation and the fruitful discussions.

I would like to thank Dr. Ayman El-Baz for useful discussions and his assistance in publishing in respected conferences. He has never hesitated to share his experience in Markov Gibbs random field, image processing. He has been a good teacher and a great friend.

I would like to thank all the members of Computer Vision and Image Processing Laboratory at University of Louisville, both past and present. Special thanks to Mr. Mike Miller for his continuous dedication to help and for his sup-
port in hard times. Also, I would like to thank Mr. Chuck Sites for his valuable technical help during the work in the lab projects.

Very special thanks to my family for their encouragement and support, without which this dissertation and research would not have been possible. My deepest gratitude to my mother, sisters, my brother, and lovable daughter Menah. Finally, words cannot describe how I am indebted to my mother and my wife for their pain, suffering and sacrifices made during the journey of this study.
ABSTRACT

IMAGE LABELING BY ENERGY MINIMIZATION WITH APPEARANCE AND SHAPE PRIORS

Asem Mohamed Ahmed Ali

April, 14, 2008

This work addresses modeling and analysis of images, in particular, labeling problems applied to image segmentation and restoration. The objective of this work is to develop accurate mathematical models combining image appearance (i.e., pixel intensities and spatial interaction between the pixels) and shape information in order to describe objects-of-interest in the images. The intensity model estimates the marginal density for each class in the image under consideration. A new unsupervised technique based on maximizing a derived joint likelihood function is proposed to model these marginal densities by Gaussian distributions. The estimation of the new technique is refined by adding Gaussian components with sign alternate using the modified expectation maximization algorithm [1]. Spatial interaction that describes the relation between pixels in each class is modeled using a Markov-Gibbs Random Field (MGRF) with Potts prior. The Gibbs potential is chosen to be asymmetric, which provides more chances to guarantee that the Gibbs energy function is submodular, so it can be minimized using a standard graph cuts approach in polynomial time. Unlike conventional approaches, the parameter of the proposed model is analytically estimated. The estimates are derived in line with the maximum likelihood approach by Gimel’farb [2]. Statistical results highlight the robustness of the proposed analytical estimation approach.
over conventional methods. Finally, the shape variations between an object and its candidates are modeled using a new probabilistic model based on a Poisson distribution.

The proposed models can be used to boost the performance of the known pixel labeling techniques. In this connection, one of the frameworks proposed in this dissertation is an unsupervised maximum-a-posteriori (MAP) framework for labeling N-Dimensional (N-D) grayscale images. In this framework, the input image and its labeling are modeled by a conventional joint Markov-Gibbs random field (MGRF) of independent N-D signals and locally interdependent pixel labels. To produce a good initial labeling, or pre-labeled image, each empirical marginal distribution of signals is closely approximated by the proposed intensity model. Then, the standard graph cuts approach based on large iterative α-expansion moves in the label space is used to refine the initial labeling under the MGRF model with analytically estimated potentials. Experimental results of synthetic and real gray scale multimodal images clarify that without optimizing any tuning parameters, the proposed approach is fast, robust to noise, and gives accurate results compared to the state-of-the-art algorithms.

Due to the efficient and successful pairwise MRFs solvers in computer vision, pairwise MRF models are popular. However, pairwise MRFs can not model the rich statistics that can be modeled with high order MRFs. Using the higher order cliques could improve the image model. However, optimization algorithms of these models have a too high time complexity to be practicable. This dissertation proposes an efficient transformation that reduces higher order energies with binary variables to quadratic ones. Therefore, the well established approaches that have been successfully used to solve the pairwise energies can be used in solving such higher order ones. The use of the proposed approach is demonstrated on the segmentation problem of color images, and it shows encouraging results. The proposed framework can be used to solve many other computer vision problems.

In order to account for image non-homogeneities outside the domain of
uniform spatial interaction assumed in the MGRF model, a new shape prior is proposed. The prior is learned from a set of training shapes by estimating variations of the shapes. This process uses a new probabilistic distance model such that the marginal distributions of an object and its background are approximated each with a linear combination of the Poisson distribution and sign-alternate Gaussians. First, an initial image is aligned with the training set using this distance model. Then, a new energy function is built by combining the above object and background appearance models with the probabilistic shape model. The optimal labeling is obtained using the min-cut techniques to approximate the global minimum of the energy function. Experiments show that the use of the shape prior improves considerably the accuracy of the graph cuts based image segmentation.
TABLE OF CONTENTS

DEDICATION ....................................................... iii
ACKNOWLEDGMENTS .............................................. iv
ABSTRACT .......................................................... vi
TABLE OF CONTENTS ............................................. ix
LIST OF TABLES .................................................... xiii
LIST OF FIGURES ................................................... xiv
List of Algorithms ................................................ xx

CHAPTER

I. INTRODUCTION ................................................. 1

A. Image Modeling ................................................. 1
   1. Intensity Model ............................................ 3
   2. Spatial Interaction Model ................................. 5
      a. Gaussian random fields ............................. 5
      b. Fractal Model ....................................... 6
      c. Fourier transform ................................. 6
      d. Markov Random Field ......................... 6
   3. Shape Model ............................................. 8

B. Image Labeling ................................................ 9

C. Why This Work Is Needed .................................. 10

D. Dissertation Contributions .................................. 10

E. Document Layout ............................................ 11

II. MARKOV-GIBBS RANDOM FIELD AND LABELING PROBLEM .. 12

A. Markov-Gibbs Random Field (MGRF) ........................ 12
   1. Gibbs Random Fields ................................... 14
   2. Markov Random Fields ................................. 16

B. MGRF Models .................................................. 17
   1. Auto-Models .......................................... 17

C. MGRF-Based Image Labeling .................................. 18
   1. Maximum-A-Posteriori Estimation .................... 18

D. Energy Minimization Techniques ............................ 20
1. Simulated Annealing (SA) ........................................ 20
2. Iterated Conditional Modes (ICM) ............................ 21
3. Max-Product Belief Propagation (BP) ....................... 21
4. Tree-Reweighted Message Passing (TRW) .................. 22

E. Graph Cuts ......................................................... 22
   1. Graphs ....................................................... 23
   2. Min-Cut/Max-Flow problems ............................... 23
   3. Expansion Moves Algorithm ............................... 25

F. Extended Roof Duality ......................................... 25
   1. Roof Duality ............................................... 26
   2. Probing Method ............................................ 27

III. MGRF PARAMETERS ESTIMATION ............................. 29
   A. Related Works .............................................. 30
      1. Coding Estimation ...................................... 30
      2. Least Square Error method (LSQR) .................... 31
      3. Parameter Estimation Using Co-occurrence Probability ... 31
      4. Analytical method for Potts model .................... 32
      5. Others .................................................... 32
   B. The Proposed Approach For Parameter Estimation ....... 33
      1. Pairwise Clique Potential Estimation ................. 34
      2. Triple Clique Potential Estimation .................... 37
   C. Experiments .................................................. 37
   D. Conclusions ................................................ 43

IV. A NOVEL UNSUPERVISED GRAPH CUTS APPROACH FOR N-D
    MULTIMODAL IMAGE LABELING ............................... 46
   A. Introduction ................................................. 46
   B. The Conditional Image Model ............................. 49
      1. Dominant Modes Estimation ............................ 51
   C. Graph Cuts-based Optimal Labeling ........................ 59
   D. Experiments and Discussion ................................ 61
      1. Ground Truth Experiments .............................. 63
      2. Real Images and Applications .......................... 65
         a. Lung Segmentation .................................... 70
   E. Conclusions ................................................ 75

V. OPTIMIZING BINARY MRFs WITH HIGHER ORDER CLIQUES ... 78
# LIST OF TABLES

| TABLE |
|---------------------------------------------------------|----------------------------------|
| 1. Accuracy of the proposed parameter estimation method for binary images of size $128 \times 128$ | 38 |
| 2. Accuracy of the proposed parameter estimation method for images of size $64 \times 64$ | 39 |
| 3. Accuracy of the proposed parameter estimation method for images of size $128 \times 128$ | 40 |
| 4. Accuracy of the proposed parameter estimation method for images of size $256 \times 256$ | 41 |
| 5. Accuracy of the proposed parameter estimation method for images of size $512 \times 512$ | 42 |
| 6. Graph Edge Weights | 61 |
| 7. 1D Statistical results | 63 |
| 8. Accuracy and time performance of the proposed approach segmentation on 7 data sets in comparison to ICM and IT. Average volume $256\times256\times77$ | 75 |
| 9. Graph Edges Weights | 107 |
| 10. Accuracy of the proposed approach segmentation on 33 slices in comparison to Graph Cut without shape and THreshold technique | 112 |
| 11. Stereo setup parameters | 120 |
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Different image types (a) a single 2D video frame of a real 3D scene, (b) a remotely sensed image of the Earth’s surface, (c) a 2D slice of a 3D computed tomography image, (d) a magnetic resonance image, (e) an ultrasound image, and (f) an X-ray image.</td>
<td>2</td>
</tr>
<tr>
<td>2. A binary image of a Dalmatian dog in a background of leaves. Observers combine the intensity and interaction information of the input image with the notion of what a Dalmatian looks like to correctly recognize the dog. Courtesy of Cremers [3].</td>
<td>8</td>
</tr>
<tr>
<td>3. Two examples of the first order neighbors for ( p ) and ( q ).</td>
<td>13</td>
</tr>
<tr>
<td>4. The neighborhood systems up to five order for a pixel ( p ).</td>
<td>14</td>
</tr>
<tr>
<td>5. The clique types of the second order neighborhood and their different potential parameters.</td>
<td>15</td>
</tr>
<tr>
<td>6. An example of undirected Graph: Image’s pixels (a-i) are the graph’s nodes. ( n )-links is constructed for 4-neighborhood system. ( t )-link connect pixels with terminals.</td>
<td>24</td>
</tr>
<tr>
<td>7. Examples of cuts on a graph. (a), (b), and (c) are valid cuts. (d) is invalid cuts; it do not separate the terminal there exist a path {s, a, d, e, h, t}. (e) is invalid cuts; it has a subset {(a, d), (b, e), (c, f)} gives a valid cut.</td>
<td>24</td>
</tr>
<tr>
<td>8. The probing method: (a) The output of roof duality with unlabeled sites. (b) and (c) the outputs of roof duality after fixing site ( p ) with 0 and 1, respectively. It can be drawn that ( f_q ) is always one, so its optimal label is 1. ( f_r ) follows ( f_p ). Therefore, sites ( q ) and ( r ) can be eliminated by letting ( f_q^* = 1 ) and ( f_r^* = f_p^* ).</td>
<td>27</td>
</tr>
<tr>
<td>9. Besag’s scheme for coding sites (a) first order model and (b) second order model.</td>
<td>30</td>
</tr>
<tr>
<td>10. Different types of potential function.</td>
<td>33</td>
</tr>
</tbody>
</table>
11. Samples of synthesized binary images of size $128 \times 128$  
12. Samples of synthesized images of size $64 \times 64$  
13. Samples of synthesized images of size $128 \times 128$  
14. Samples of synthesized images of size $256 \times 256$  
15. Samples of synthesized images of size $512 \times 512$  
16. Results of the proposed method for estimating anisotropic Potts model parameters for images of size $128 \times 128$: (a) $[25.6(0.05) 0.003(0.08) 0.001(0.08) 0.002(0.08)]$, (b) $[0.003(0.09) 25.6(0.05) 0.004(0.09) 0.001(0.09)]$, (c) $[0.001(0.08) 0.001(0.09) 0.005(0.09) 25.9(0.05)]$, and (d) $[0.01(0.09) 0.01(0.09) 25.9(0.05) 0.01(0.09)]$  
17. Samples of synthesized images with 32 colors and high order cliques (a) Sample of realizations generated with $\gamma = 5$ and (b) Sample of realizations generated with $\gamma = 10$  
19. The effect of the user input on the final labeled image. courtesy of Boykov and Funka-Lea [5]  
20. pAIC result for a bimodal synthetic image (a)Empirical and estimated densities, and the 2 Gaussian components. (b) The log likelihood (maximum at 2).  
21. pAIC result for a 3-modal synthetic image (a)Empirical and estimated densities, and the 3 Gaussian components. (b) The log likelihood (maximum at 3).  
22. pAIC result for a 4-modal synthetic image (a)Empirical and estimated densities, and the 4 Gaussian components. (b) The log likelihood (maximum at 4).  
23. pAIC result for a 5-modal synthetic image (a)Empirical and estimated densities, and the 5 Gaussian components. (b) The log likelihood (maximum at 5).
24. Non Gaussian 3-class result: (a) and (b) show the output of the pAIC-EM algorithm, (c) shows the normalized absolute error between the empirical and estimated densities, (d) shows the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm, (e) shows the empirical and estimated densities, and (d) shows the marginal densities with the best thresholds. 

25. Result for CT Lung slice: (a) The CT slice, (b) and (c) the output of the pAIC-EM algorithm, (d) the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm, (e) the empirical and estimated densities, and (f) the marginal densities with the best threshold. 

26. Result for 3-class MRA slice: (a) The MRA slice, (b) and (c) the output of the pAIC-EM algorithm, (d) the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm, (e) the empirical and estimated densities, and (f) the marginal densities with the best thresholds. 

27. Proposed algorithm flowchart. 

28. Effect of choosing $\gamma$ (a) Original image (b) Noisy output $\gamma = 0.05$ (c) Over smoothed output $\gamma = 5$. 

29. The changes of the relative error with $\gamma$, proposed adaptive analytical approach values shown in red asterisk. 

30. 1D image labeling: (a) The original signal is distorted by WGN (b) Distorted signal (c) MLE outputs (d) Proposed algorithm output. 

31. Segmentation results (a) The gray levels image with SNR=1 dB, (b) Proposed Approach Segmentation. (c) The gray levels image with SNR=5 dB, (d) Proposed Approach Segmentation. (Error shown in red) 

32. Segmentation results: for the 3-modal image (a) EDISON output, (b) NCUTS output. For the 5-modal image (c) EDISON output, (d) NCUTS output. (Error shown in red) 

33. The changes of the misclassification error with the SNR for 3-modal and 5-modal images. 

34. The changes of the misclassification error with the SNR for 3-modal and 5-modal images. 

xvi
35. 4-modal image segmentation results (b) $\varepsilon = 0.16\%, \tau = 5$ sec. (c) $\varepsilon = 0.71\%, \tau = 13$ sec. (errors shown in red) .......................... 69

36. Example of a graph that used in volume labeling. Note: Terminals should be connected to all voxels but for illustration purposes, this was not done. .................................................. 69

37. Slices from the synthetic volume. .......................................................... 69

38. Synthetic volume segmentation results (a) 3D segmentation (b) 2D segmentation (Error shown in green). ................................. 70

39. Real image segmentation results of (a) Proposed Algorithm (b) EDISON ($h_s=15, h_c=9, M=5000.$) (c) NCUT ($nsg=2$, and the default parameters). (Courtesy of Shi and Malik [6]) ............................ 71

40. Starfish segmentation results of (a) Proposed Algorithm, (b) EDISON ($h_s=2, h_c=10, M=500.$), (c) NCUT ($nsg=2$, of $=0.04$), (d) Proposed Algorithm (e) EDISON ($h_s=7, h_c=6.6, M=5000.$)(f) NCUT ($nsg=2$, of $=0.04$, th $=0.03$, eem $=5$, et $=0.1$) ............................ 71

41. Kidney segmentation results of (a) Proposed Algorithm $\varepsilon=1.7\%$ (b) EDISON ($h_s=5, h_c=6.1, M=9000.$) $\varepsilon=1.1\%$ (c) NCUT ($nsg=2$, of $=0.01$, th $=0.04$, et $=0.01$) $\varepsilon=1.1\%$ (d) Proposed Algorithm $\varepsilon=0.1\%$ (e) EDISON ($h_s=7, h_c=6.6, M=9000.$) $\varepsilon=0.01\%$ (f) NCUT ($nsg=2$, and the default parameters) $\varepsilon=0.6\%$ .................................................. 71

42. Ling Phantom Segmentation: (a) Phantom gray scale image, (b) the proposed algorithm output, (c) ICM output, (d) IT output. (The misclassified pixels are shown in red color) .............................................. 73

43. Segmentation Results of: proposed algorithm (1st column) , ICM (2nd column), and IT (last column). (The misclassified pixels are shown in red color) .............................................. 74

44. Proposed algorithm’s results. Left: segmented lung volumes (Error are shown in green). Right: samples from volume’s segmented slices (Error are shown in red). Segmentation errors (a) 2.08\%, (b) 2.21\%, (c) 2.17\%, and (d) 1.95\% .................................................. 76

45. Examples of segmented lung slices that have nodules (bounded by yellow circle). Left: IT and middle: ICM approaches misclassified these parts as chest tissues (error is shown in red). However, right: The proposed algorithm correctly classified them as a lung. .............................................. 77
46. Part of an image lattice for $2^{nd}$ order neighborhood system and cliques of size three ................................................. 91

47. Starfish segmentation results. (a) the pairwise cliques results, and (b): the higher order cliques results. ........................................... 93

48. Starfish with occlusions segmentation results. (a,c): the pairwise cliques results, and (b,d): the higher order cliques results. ........... 94

49. Starfish with occlusions more segmentation results. (a,c): the pairwise cliques results, and (b,d): the higher order cliques results. ... 95

50. More segmentation results. (a,c): the pairwise cliques results (Numbers in images refer to the regions with inhomogeneities), and (b,d): the higher order cliques results. ................................................. 96

51. More segmentation results. (a,c,e): the pairwise cliques results (Numbers in images refer to the regions with inhomogeneities), and (b,d,f): the higher order cliques results. ................................................. 97

52. More segmentation results for partially occluded objects. (a,c): the pairwise cliques results (Numbers in images refer to the regions represent artificial occlusion), and (b,d): the higher order cliques results. ................................................. 98

53. Samples of kidney training data images: (a) original , (b) Segmented, (c) Aligned (d) Binary ......................................................... 103

54. (a) The labelled image, (b) The iso-contours ........................................... 104

55. (a,b) Empirical densities and the estimated Poisson distributions, (c,d) Components of distance probabilistic models, (e,f) Final estimated densities ......................................................... 106

56. Gray level probabilistic model for the given image (a) Original image (b) aligned image (c) Initial density estimation (d) LCG components (e) Final density estimation (f) Marginal densities. Segmented Kidney (g) Results of gray level threshold 102.6% (h) Results of Graph cuts without shape constraints 41.9% (i) Proposed approach results 2.5% ......................................................... 108

57. More segmentation results (a)Original images (b) Results of Graph cuts without shape constraints (c) Proposed approach results ....... 110

58. More segmentation results (a)Original images (b) Results of Graph cuts without shape constraints (c) Proposed approach results ....... 111
59. Kidney Phantom (a) The phantom (b) Results of Graph cuts without shape constraints (c) Proposed approach results 113

60. Image matching results. Left relative positions before matching. Right matching results. 116

61. General stereo pair setup. The relation between the depth and the disparity 117

62. Rectified stereo pair setup. The depth is inversely proportional to the disparity 118

63. Example of the depth map: (a) one of the image pair and (b) the corresponding depth map. 118

64. The system setup 120

65. Reconstruction results. (a) The stereo pair, (b) left the depth map and right the reconstructed shape. 121

66. More reconstruction results. Left: One of the stereo pair. Middle: Frontal view from the reconstructed shape. Right: Side view from the reconstructed face 122
# List of Algorithms

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Iterated Conditional Modes (ICM) [7]</td>
<td>21</td>
</tr>
<tr>
<td>2  $\alpha$-Expansion Move Algorithm [8]</td>
<td>25</td>
</tr>
<tr>
<td>3  Extended Roof Duality Algorithm (Probing Method) [9]</td>
<td>28</td>
</tr>
<tr>
<td>4  Gibbs Sampler Algorithm [10]</td>
<td>38</td>
</tr>
<tr>
<td>5  pAIC-EM Algorithm</td>
<td>53</td>
</tr>
<tr>
<td>6  Generating Coefficients of Size $k$ Clique’s Energy Algorithm</td>
<td>83</td>
</tr>
<tr>
<td>7  Transform to Quadratic</td>
<td>88</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

An image is a graphic representation of a scene in two or three dimensional spaces. An image is stored as a raster data set of integer values that represent the intensity of reflected light, heat, or other range of values on the electromagnetic spectrum. Common examples include photographs which are a digitization of two dimensional projections of three dimensional scenes, remotely sensed images (e.g., satellite data), and scanned data (see Fig. 1 (a,b)). Other examples include medical images, where image intensities represent radiation absorption in X-ray imaging and Computed Tomography (CT), acoustic pressure in ultrasound, or radio frequency (RF) signal amplitude in Magnetic Resonance Imaging (MRI) (see Fig. 1(c-f)). Images may be acquired in a continuous domain or in a discrete space. For two dimensional discrete images, the location of each measurement is a “image element” called pixel (in three dimensional space, it is called a voxel), and each object or class in the image is represented by a group of pixels.

A. Image Modeling

The goal of image modeling is to quantitatively specify visual characteristics of the image in few parameters so as to understand natural constraints and general assumptions about the physical world and the imaging process [11]. Stochastic approaches, particularly random field models, have proved useful in modeling real images that have large varieties from one to another. These models have been used in image processing algorithms such as segmentation, enhancement and restoration [12]. Random field models can represent prior information of an image so that the powerful Bayesian decision theory can be applied to solve these prob-
FIGURE 1 – Different image types (a) a single 2D video frame of a real 3D scene, (b) a remotely sensed image of the Earth’s surface, (c) a 2D slice of a 3D computed tomography image, (d) a magnetic resonance image, (e) an ultrasound image, and (f) an X-ray image.
lems. Objects-of-interest in the images are characterized by geometric shapes and
visual appearance, although it is very difficult to formally define these notions. In
this dissertation, the visual appearance is characterized by marginal probability
distribution of pixel or voxel intensities and by spatial interaction between pixels
or voxels in each object. The shape is characterized by typical object boundaries on
the images. One of the main objectives of this dissertation is to develop more ac-
curate mathematical models of these characteristics than the known models. This
section presents an overview for existing image modelling approaches.

1. Intensity Model

This model estimates the marginal density for each class in the given image
from the mixed normalized histogram of the occurrences of the gray levels within
that image. Density estimation has thus been heavily studied, under two primary
umbrellas: parametric and nonparametric methods. Nonparametric methods take
a strong stance of letting the data represent themselves. Nonparametric methods
(e.g., Parzen window [13]) achieve good estimation for any input distribution as
more data are observed. However, these methods have many parameters that need
to be tuned [14]. One of the core methods that nonparametric density estimation
approaches based on is the \(k\)-nearest neighbors (\(k\)-NN) method. These approaches
calculate the probability of a sample by combining the memorized responses for
the \(k\) nearest neighbors of this sample in the training data. In these estimators
(e.g., the Parzen density estimator [13]), it is noticed that the amount of computa-
tion is directly related to the number of training samples. In order to reduce the
computation, Fukunaga and Hayes [15] extracted a representative subset from the
training data. This subset was chosen such that the Parzen density estimation gen-
erated with this reduced subset is very close to the one that generated with the
full data set in the sense of an entropy measure of similarity between the two esti-
mates. Silverman [16] proposed a kernel density estimator using the Fast Fourier
Transform (FFT). He estimated density using univariate Parzen window on regular grids. This method exploits the properties of the FFT, where the FFT of the density estimate is the product of the FFTs of the kernel function and the data. However, this algorithm can not be used in the general cases of density estimates. In order to reduce the number of kernel evaluations, Jeon and Landgrebe [17] proposed a simple branch-and-bound procedure that is applied to the Parzen density estimation. Girolami and He [18] proposed a Parzen window-based density estimator which employs condensed data samples. The advantage of nonparametric methods is the flexibility: they can fit almost any data well. No prior knowledge is required. However, they apparently often have a high computational cost. Also, there is no opportunity to incorporate prior knowledge.

On the other hand, parametric methods are useful when the underlying distribution is known in advance or is simple enough to be modeled by a simple distribution function or a mixture of such functions. The parametric model is very compact (low memory and CPU usage) where only few parameters need to fit. Parameters of a mixture are typically estimated using the Expectation-Maximization (EM) algorithms that converge to the maximum likelihood estimates of the mixture weights (prior probabilities of the mixture components) and parameters of each component [19]. Since Laird et al. [20] extended the EM algorithm to be used in estimating parameters from an incomplete data set, EM became a popular approach in density estimation and many versions of EM are introduced [19]. Recently, Farag et al. [1, 21] used a Linear Combination of Gaussians (LCG) with positive and negative components to estimate the marginal density of each class in a given image. They developed a modified EM algorithm to estimate the parameters of this LCG model.

The limitation of all the aforementioned approaches is that all of them depend on using samples of training data to estimate the marginal density for each class in the given image. In this dissertation, the major objective is to present an unsupervised approach to estimate the marginal density for each class from the
mixed normalized histogram of occurrences of the gray levels. Similar to the work of Farag et al. [1,21], the marginal intensity distribution is modeled by a linear combination of Gaussians (LCG) with sign-alternate components. Parameters of the mixtures are estimated by using the modified EM algorithm. However, in this algorithm the number of classes and initial parameters of their dominant modes are set manually. In this dissertation, both the number of the classes and the initial mixture parameters are estimated from the given empirical distribution using a new technique described in section IV.B.1.

2. Spatial Interaction Model

Spatial interaction helps in eliminating possible ambiguities, correcting errors, and recovering missing information in the image labeling problem. Spatial interaction models describe the relation between the pixels in an image mathematically. To do this, the image is realized as a stochastic process on a random field. This random field is a joint distribution imposed on a set of random variables representing pixel intensities that imposes statistical dependence in a spatially meaningful way. Random field models provide a good tool for blending information about local spatial interaction into a global framework. The literature is rich with models that are used in describing the spatial interaction of image pixels. Each model has its own representation for the relationship between local sites in the random field. This section tries to give a brief review for the popular models. Many random field models have been discussed in Dubes and Jain’s work [11]. Indeed, the authors of [11] provided a taxonomy of these various models.

a. Gaussian random fields  Gaussian random fields are special models that take advantage of the mathematical properties of the gaussian distribution. The Gaussian model requires all interactions between pixels to be Gaussian distributed, and it only allows multiple pairwise interactions to occur. The most popular Gaussian models in image analysis are Simultaneous Auto-Regressive (SAR)

b. Fractal Model Fractals are useful in modeling images of natural surfaces (e.g., clouds, leaves, rivers, ... etc.) that have a statistical quality of roughness and self-similarity at different scales. For a review about fractal models see [22]. Pentland [23] used the fractal models for image segmentation. Also, Garding [24] used these models in textures segmentation.

c. Fourier transform This model is used in textured images classification. Fourier transform mimics the human visual system by extracting the different frequency components and analyzing the image in the frequency domain. To segment a variety of natural and synthetic textures, Coggins and Jain [25] used a set of frequency and orientation selective filters in a multiband filtering approach. Smith [26] used a set of band pass filters followed by zero crossing detection to successfully generate a tree classifier of textures.

d. Markov Random Field More commonly used models in image analysis are Markov Random Field (MRF) models. These models capture the local spatial textural information in an image by assuming that the pixel intensity depends on the intensities of the neighboring pixels. MRF models are among the most successful models that are used to represent visual contextual information in labeling problems [27]. Moreover, MRF models that have exponential priors belong to the class of Gibbs models. While a MRF is defined in terms of local properties, a Gibbs Random Field (GRF) describes the global properties of an image in terms of joint distribution of intensity for all pixels [28]. This class of MRF that has exponential priors, known as the Markov-Gibbs Random Field (MGRF), has been extensively used in image modeling [11, 28]. Gimel’farb [29] proposed a MGRF model that takes into account multiple pairwise pixel interactions. This model is used for images (textures) that are spatially uniform. Such an image is classified by its gray level difference histogram. The author proposed an algorithm based on the maximum likelihood approach to estimate the model parameters. Compared to other models, the author claims that the parameters for his model are larger in number.
but simpler to estimate. Zhu et al, [30] proposed the Filters, Random Fields, and Maximum Entropy (FRAME) model. This model integrates filtering theory and MRF modeling using the maximum entropy principle. To model a set of texture images, they extracted texture features by applying a set of filters to the observed images. Then, the marginal distributions of the images are estimated from the histograms of the filtered images. After that, they fit a distribution for this texture from the marginal distributions using a maximum entropy-based method. The authors claim that this model is more descriptive than conventional MRF models. However, it is computationally very expensive. Although MGRF is a very good tool to model an image, its parameters estimation is still a challenge. Several studies [2, 28, 31–33] have been proposed in the literature to estimate these parameters. Some of these methods are discussed in chapter III.

Farag et al., [1] proposed an analytical approach to estimate the parameter of a specific MGRF model (the homogenous isotropic Potts model governing symmetric pairwise co-occurrences of labels). In this dissertation a similar analytical approach to estimate the parameter of another specific MGRF model (the homogenous isotropic Potts model governing asymmetric pairwise co-occurrences of labels), is proposed. This is discussed in detail in chapter III.

A MGRF model is specified by the set of clique potentials. Most of the aforementioned approaches focused on unary and pairwise cliques. However, this representation can not model the rich statistics of natural scenes [34]. Such rich statistics can be modelled using higher order clique potentials. However, due to computational expense of the optimization algorithms of higher order MGRF models, their use has been quite limited. This dissertation proposes a new efficient algorithm that will transform a higher order binary energy to a quadratic one, and so it can be used in practice. This work is presented in chapter V.
3. Shape Model

In many cases, an image has misleading information (intensity and interaction models). However, the human brain tends to still capture the visual characteristics of the given image. An example of this case is shown in Fig. 2. This example is for a Dalmatian dog in an environment of fallen leaves and grass and due to coarse-graining and binarization, the dog cannot be distinguished from the background based on the texture only. However, human observers combine the intensity and the interaction information of the input image with the notion of what a Dalmatian looks like to correctly recognize the dog. Thus, a shape model provides useful information that balances the missing low level information in cases such as poor image resolution, diffused boundaries, noise, or occlusion.

The literature contains many shape modeling approaches such as the one proposed by Leventon et al. [35]. This approach combines the shape and deformable model by attracting the level set function to the likely shapes from a training set specified by Principal Component Analysis (PCA). To make the shape guide
the segmentation process, Chen et al. [36] defined an energy functional which basically minimizes an Euclidean distance between a given point and its shape prior. Huang et al. [37] combine registration with segmentation in an energy minimization problem. The evolving curve is registered iteratively with a shape model using the level sets. They minimized a certain function in order to estimate the transformation parameters. In [38], shapes are represented with a linear combination of 2D distance maps where the weight estimates maximize the distance between the mean gray values inside and outside the shape. In Paragios’s work [39], a shape prior and its variance obtained from training data are used to define a Gaussian distribution, which is then used in the external energy component of a level sets framework. One of the main limitations in these approaches is that they did not model the shape variations and so they cannot handle shapes with large deformation. In this dissertation, a new shape model is proposed to overcome this limitation, as shown in chapter VI.

B. Image Labeling

The image labeling problem is specified in terms of a set of sites (e.g., image pixels, segments, ... etc.) and a set of labels (e.g., pixel color, texture type, ... etc.). The objective of labeling algorithms is to assign the true label for each site. This problem can be formulated in a Bayesian framework using Markov random fields [27]. In this framework, the task is to find the Maximum-A-Posteriori (MAP) estimate of the underlying quantity. This dissertation focuses on the image pixels as sites, and gray levels as labels. This problem will be discussed in detail in chapter II. The MAP-based methods choose the estimated labeled image that maximizes the posterior probability of the labeled image given the observed image. Many optimization techniques proposed in the literature use stochastic models to solve the labeling problem. A review for some of these optimization approaches is introduced in Sec.II.D.
C. Why This Work Is Needed

The objective of this work is to use these models (intensity, spatial interaction, and shape) to interface with and boost the performance of existing image labeling problem. The image labeling can be used as a formulation for diverse computer vision and image processing applications, such as image segmentation, image restoration, image matching, and stereo. Image segmentation is an important preliminary step in many real world problems such as computer added diagnosis, object recognition, shape analysis. Thus, one of the application sections of this dissertation is dedicated to image segmentation.

D. Dissertation Contributions

This work addresses the image modeling and image analysis, especially the labeling problem for gray scale and color images. The objective of this work is to find accurate mathematical models (intensity, spatial interaction, and shape) that describe all possible information in the image. The main contributions in this dissertation can be categorized into different types:

- Intensity Model:

  1. The number of classes in the given multimodal image is determined by using a new technique based on maximizing a new joint likelihood function.

- Spatial interaction Model:

  1. A new analytical approach to estimate the parameter of a specific MGRF model (the homogenous isotropic Potts model governing asymmetric pairwise co-occurrences of labels) is presented.

  2. A new efficient algorithm that will transform a higher order energy to a quadratic one is proposed, and so it can be used in practice.

- Shape Model:
1. The shape variations are estimated using a new probabilistic model.

- **Algorithms:**

1. Unlike previous graph cuts based segmentation techniques, no user interaction is needed; instead, a new unsupervised MAP-based labeling framework of N-D multimodal gray scale images is proposed.

**E. Document Layout**

This document is presented in seven chapters. The following remarks summarize the scope of each chapter.

**Chapter II** discusses the problem formulation of the image labeling in terms of the model that is used to represent the image, the technique that is used to find the Maximum-A-Posteriori (MAP) estimate.

**Chapter III** proposes an analytical method to estimate the parameter of homogeneous isotropic Potts model for an asymmetric Gibbs potential function.

**Chapter IV** presents a novel unsupervised graph cuts approach for N-D multimodal image labeling (image segmentation and image restoration).

**Chapter V** proposes an efficient transformation that reduces a higher order energy with binary variables to a quadratic one. The use of the proposed method is demonstrated on the segmentation problem of color images, and it shows encouraging results.

**Chapter VI** presents a novel shape representation and application for image segmentation.

**Chapter VII** introduces experiments for human faces reconstruction based on a stereo matching technique as an application for image labeling.

**Chapter VIII** summarizes the main component of the proposed work and presents a plan for future work.
CHAPTER II
MARKOV-GIBBS RANDOM FIELD AND LABELING PROBLEM

The labeling problem gives common notation for many diverse vision and image processing problems such as stereo, image restoration, image matching, and image segmentation. The image labeling problem is specified in terms of a set of sites (e.g., image pixels, edges, segments, ...etc.) and a set of labels ( e.g., pixel color, texture type, ...etc.). The objective of labeling algorithms is to assign the true label for each site. This problem can be formulated in a Bayesian framework using Markov random fields [27]. In this framework the task is to find the Maximum-A-Posteriori (MAP) estimate of the underlying quantity. This chapter is dedicated to the discussion of problem formulation and solving tools. Sec. II.A introduces the Markov-Gibbs Random Field (MGRF) and its properties. Sec. II.B presents common MGRF models that have been used in image modelling. The MGRF-based formulation of image labeling is given in Sec. II.C. Different techniques that have been used to solve labeling problem are explored in sections II.D, II.E, and II.F.

A. Markov-Gibbs Random Field (MGRF)

A random field is defined as a triplet consisting of: a sample space, a class of Borel sets on the sample space, and a probability measure $P$ whose domain is the class of Borel sets [11]. A random field model is a specification of $P$ for a particular class of random variables, such as intensities at an image pixels. For an observed image, a stochastic model can be constructed as follows. Let $P = \{1, 2, \ldots, n\}$ be a set of $n$ sites that represents the set of image pixels. Let $G = \{0, \ldots, Q - 1\}$ and $L = \{0, \ldots, K - 1\}$ denote the set of gray levels and region labels, respectively,
where, $Q$ is the number of gray levels, and $K$ is the number of labels.

**Definition 1.** A digital image is defined by a function $I : \mathcal{P} \to \mathcal{G}$ that maps the sites onto the set of signal values.

**Definition 2.** A labeled image is defined by a function $f : \mathcal{P} \to \mathcal{L}$ that maps the sites onto the set of labels.

Let $F = \{F_1, F_2, \ldots, F_n\}$ be a set of random variables defined on $\mathcal{P}$. Hence, $f = \{f_1, f_2, \ldots, f_n\}$ is defined as a configuration of the field $F$. Denote by $\mathcal{F}$ the set of all labelings $\mathcal{L}^n$.

Since the image has a natural structure that is 2D array, this helps to define a geometric neighborhood system $\mathcal{N}$ consisting of a set of all neighboring pairs $\{p, q\}$ where $p, q \in \mathcal{P}$. The most popular neighborhood systems in image analysis is the first order neighbors where the four nearest neighbors sharing a side with the given pixel. Fig. 3 shows an example of this neighborhood system where the neighborhood of $p$ is $N_p = \{a, b, c, d\}$, and the neighborhood of $q$ is $N_q = \{x, y\}$.

The symmetric neighborhood $N_p$ satisfies the following properties:

1. $p \notin N_p$

2. If $p \in N_q$ then $q \in N_p$.

Fig. 4 illustrates the neighborhood systems up to five order for a pixel $p$. 

![FIGURE 3 – Two examples of the first order neighbors for $p$ and $q$.](image-url)
1. Gibbs Random Fields

In 1901, Gibbs used Boltzmann’s distribution of energy states in molecules to express the probability of a whole system with many degrees of freedom being in a state with a certain energy [27]. A Gibbs random field (GRF) describes the properties of an image in terms of the joint distribution of pixels labels. A discrete Gibbs random field provides a global model for an image by specifying a probability mass function in the following form:

$$P(f) = \frac{1}{Z} \exp(-U(f)/T),$$

(1)

where $Z$ is a normalizing constant called the partition function, $T$ is a control parameter called temperature. The term $U(f)$ denotes the Gibbs energy [28], and is given by:

$$U(f) = \sum_{c \in C} V_c(f),$$

(2)

where $V_c$ is known as the potential function, or the clique function and $C$ is the set of all cliques. A clique is defined as [27]:

**Definition 3.** A clique is a set of sites (e.g., pixels in an image) in which all pairs of sites are mutual neighbors.

Fig. 5 illustrates the clique types of the second order neighborhood system.
FIGURE 5–The clique types of the second order neighborhood and their different potential parameters
2. Markov Random Fields

Markov Random Fields (MRF) were introduced to image analysis by Hassner and Sklansky [40]. A Gibbs random field describes the global properties of an image in terms of the joint distributions of pixels labels, whereas, a Markov random field is defined in terms of local properties. Markov random fields provide a convenient prior for modeling spatial interactions between image pixels.

**Definition 4.** The random field $F$, with respect to a neighborhood system $\mathcal{N}$, is a discrete Markov random field if its probability mass function $P(F = f)$ satisfies the following properties:

1. $P(F = f) > 0$ for all $f \in \mathcal{F}$, \hspace{1cm} (Positivity)
2. $P(F_p = f_p|F_p-\{p\} = f_{p-\{p\}}) = P(F_p = f_p|F_{N_p} = f_{N_p})$, \hspace{1cm} (Markov Property)
3. $P(F_p = f_p|F_{N_p} = f_{N_p})$ is the same for all sites $p$, \hspace{1cm} (Homogeneity)

where $\mathcal{P} - \{p\}$ denotes set difference and $f_{N_p}$ denotes all labels of pixels in $N_p$.

MRF probability mass function $P(F = f)$ will be abbreviated as $P(f)$. The Markov property states that a pixel label depends directly only on its neighbors. This property establishes a local model.

**Theorem 1.** **Hammersley-Clifford** [28]: Under the positivity condition, the probability distribution $P(f)$ for an MRF can be represented as the GPD of Eq. (1) with potential function supported by cliques $\mathcal{C}$ of the neighborhood graph describing the neighborhood system $\mathcal{N}$.

This theorem provides a convenient way to specify MRF, where a unique GRF exists for every MRF and vice versa as long as the GRF is defined in terms of cliques on a neighborhood system.
B. MGRF Models

Geman and Geman [27] introduced the MGRF model to engineers as a powerful tool for image modeling. MGRF models have been successfully used in texture and general image analysis and synthesis (For details see [2] and references therein). The literature is rich with MGRF models, each of which tries to select the potential functions that are suitable for a specific system behavior. Here, a simple review of the most popular and related discrete models is given.

1. Auto-Models

The Gibbs energy can be defined by specifying interactions between sites in the image. In most of the image processing and computer vision literature, the Gibbs energy has been defined in terms of the “single-site” potentials and the “two-site” potentials. This is called the pairwise interaction models. As Picard described in [41], the single-site potentials, also called the “external field”, allows one to impose structure on a pattern from an outside source. The two-site potential, also called a “bonding parameter”, influences the “attraction” or “repulsion” between neighboring pairs of pixels in the image. The different models corresponding to this form of the energy are typically called “auto-models”. Besag [28] formulated the energy function of these models as follows:

\[
U(f) = \sum_{p \in P} \left( V_p(f_p) + \sum_{q \in N_p} V_{pq}(f_p, f_q) \right),
\]

where \( V_p(\cdot) \) is the potential function for single-pixel cliques, and \( V_{pq}(\cdot, \cdot) \) is the potential function for all cliques of size 2, with \( V_{pq}(f_p, f_q) = V_{pq}(f_q, f_p) \) and \( V_{pq}(f_p, f_p) = 0 \). In the homogeneous models (site independent) \( V_p(\cdot) \) is represented by \( V(\cdot) \) and \( V_{pq}(\cdot, \cdot) \) is represented by \( V_q(\cdot, \cdot) \) (i.e., \( V_q(\cdot, \cdot) \) depends on the orientation of the neighbor as show in Fig. 5). In the homogeneous isotropic models \( V_{pq}(\cdot, \cdot) \) is represented by \( V(\cdot, \cdot) \).

An example of a Gibbs model having an energy function of this form is the
homogeneous auto-binomial model used by Cross and Jain [42] where

\[
V(f_p) = \gamma_0 f_p - \ln\left(\frac{K!}{f_p!(K-f_p)!}\right),
\]

\[
V_q(f_p, f_q) = \gamma_q f_p f_q,
\]

where \(\gamma_0\) controls the influence of the external field, and \(\gamma_q\) influences the interaction between neighboring pairs. \(\gamma_q\) is called the pairwise potential (e.g., \(\gamma_1, \gamma_2, \gamma_3, \gamma_4\)) that depends on the orientation of site \(q\) relative to its neighbor \(p\) as show in Fig. 5.

In the isotropic models, \(\gamma_q = \gamma\). Derin-Elliott’ model [31] can also be expressed in this framework as follows

\[
V(f_p) = \gamma_0 f_p,
\]

\[
V_q(f_p, f_q) = \begin{cases} 
\gamma_q & \text{if } f_p = f_q, \\
-\gamma_q & \text{otherwise}.
\end{cases}
\]

(5)

One of the most popular models in computer vision is the homogeneous Potts model [27]. The Potts model is similar to the Derin-Elliott model, but \(\gamma_0 = 0\). A similar type of the latter model is used in this dissertation.

C. MGRF-Based Image Labeling

In image labeling problems, one tries to recover a number of hidden variables (e.g., image labels) based on observable variables (e.g., image gray levels). In this problem, MGRF models fit within the Bayesian framework of Maximum-A-Posteriori (MAP) estimation, where the objective is to estimate the labeling that solves the maximization problem.

1. Maximum-A-Posteriori Estimation

Since the field \(F\) is not observable, its realization \(f\) (the desired map) is estimated based on the observation \(I\) (the input image). The common way to estimate an MRF is MAP estimation [27]. Following the conventional approaches the input
image and the desired map (labeled image) are described by a joint MGRF model of independent image signals and interdependent region labels. A two-level probability model of the input image and its desired map is given by a joint distribution
\[ P(I, f) = P(f)P(I|f) \]
where \( P(I|f) \) is a conditional distribution of the original image given the map, and \( P(f) \) is an unconditional probability distribution of the map. Note that when the given data is too noisy, the dependence of the data \( I \) on the desired values \( f \) is week i.e., \( P(I|f) \approx P(I) \) [43]. The Bayesian maximum-a-posteriori estimate of the map \( f \), given the image \( I \) is expressed as:
\[ f^* = \arg \max_{f \in F} P(I|f)P(f). \] (6)

In order to assure that the posterior distribution \( P(I|f) \) is a MRF model, this requires conditional independence of the observed random variables \( I = \{I_1, I_2, \cdots , I_n\} \). One way to get this is to assume the noise at each pixel is independent. Therefore
\[ P(I|f) = \prod_{p \in P} P(I_p \mid f_p). \] (7)

By replacing \( P(I|f) \) and \( P(f) \) in Eq. (6) using their expressions from equations (1), (2), and (7), and after simple algebraic manipulations, the following expression is obtained
\[ f^* = \arg \max_{f \in F} \exp \left( \sum_{p \in P} \log(P(I_p \mid f_p)) - \sum_{c \in C} \frac{V_c(f)}{T} \right). \] (8)

Since the temperature \( T \) is constant for the given image, it can be removed from the expression and its value implicitly estimated with the potential parameter. In order to have a complete description for the MGRF model, one should specify the clique potential function. By choosing the MGRF model to be the pairwise homogeneous Potts model described in Sec.II.B, Eq. (8) can be rewritten as follows:
\[ f^* = \arg \max_{f \in F} \exp \left( \sum_{p \in P} \log(P(I_p \mid f_p)) - \sum_{(p,q) \in N} V(p; f_q) \right). \] (9)

Unfortunately, this problem has no analytical solution. However, maximizing the
likelihood in Eq. (9) is equivalent to minimizing the following energy function:

\[ E(f) = \sum_{\{p,q\} \in N} V(f_p, f_q) + \sum_{p \in P} D(f_p), \]  

(10)

where \( D(f_p) = -\log(P(I_p | f_p)) \) which usually called the data penalty term.

D. Energy Minimization Techniques

To solve the MAP estimation problem Eq. (10) many approaches were proposed. Classical iterative search algorithms can be either stochastic (e.g., simulated annealing [27]) or deterministic (e.g., iterated conditional modes [7]). Recently, new graph-based algorithms such as Graph Cuts [44, 45], Tree-ReWeighted message passing (TRW) [46], Belief Propagation (BP) [47], and roof duality [48] have proven to be very powerful in minimizing such energy Eq. (10). Many studies (e.g., [9, 49]) in the literature have been done to investigate the performance of these algorithms in solving different computer vision problems. These studies show that modern energy minimization methods are much superior than classical methods. In this section and the following two sections some of these algorithms are summarized.

1. Simulated Annealing (SA)

The objective of this algorithm is to find MAP estimates of all labels simultaneously. Simulated annealing algorithm is based on the Metropolis approach [50] and it has been popularized by Geman and Geman [27], who used SA to solve the image labeling problem. The idea is to sample from Gibbs distribution with energy \( \frac{U(f)}{T} \), where the temperature parameter \( T \) is slowly decreased to 0. With certain temperature schedules, annealing can be guaranteed to find the global solution in the limit [27]. However, the schedules that lead to this global need potentially long runtime [8], and so sub-optimal schedules are used in practice. In this case, the algorithm is not expected to find the global solution.
2. Iterated Conditional Modes (ICM)

The ICM algorithm was proposed by Besag [7] to compute the MAP estimate in a computationally simple manner that is faster than the simulated annealing. However, it is a local energy optimization technique. The algorithm is very sensitive to the initial labeling. Choosing the prior MRF model is a critical step in this algorithm. An outline of the ICM is described in Algorithm 1.

Algorithm 1 Iterated Conditional Modes (ICM) [7]

1: Choose a MRF model for $P(f)$.
2: Select labeling $\hat{f}$ that maximizes $P(I|f)$
3: while $i < N_{iter}$ do
4: for all $p \in P$ do
5: Update $\hat{f}_p$ by the value of $f_p$ that maximizes $P(I_P|f_P)P(f_p|\hat{f}_{N_p})$
6: end for
7: increase $i$.
8: end while

3. Max-Product Belief Propagation (BP)

The BP algorithm approximately minimizes energies such as Eq. (10). It gives an exact minimization if the graph of the energy is a tree. The key idea of the BP can be described as follows. It passes messages around the graph defined by the four-connected image grid. Defined by $m^i_{pq}$ is the message that a node $p$ sends to a neighboring node $q$ at an iteration $i$. Each message is a vector of dimension $|\mathcal{L}|$. All messages are initialized to zero, and at each iteration they are updated as follows:

$$m^i_{pq}(f_q) = \min_{f_p} (V(f_p, f_q) + D(f_p) + \sum_{r \in N_p - \{q\}} m^{i-1}_{rp}(f_p)).$$

(11)

The algorithm keeps passing messages for edges until all messages become valid (i.e., the convergence). The message is said to be valid if the updating process Eq.
(11) does not change it (or it is changed by a constant independent of \( f_q \)). After
that a belief vector is computed for each node as

\[
b(f_q) = D(f_q) + \sum_{r \in N_q} m_{rq}^{N_{iter}}(f_q),
\]

(12)

where \( N_{iter} \) is the number of iterations. Finally, the optimal label at each node is
selected such that it minimizes each belief individually. The BP can solve a more
general class of functions than graph cuts, but it has some drawbacks. It diverges
in the case of graphs that have loops, such cases exist in many computer vision
problems. Also, it gives solutions with higher energy than graph cuts [49].

4. Tree-Reweighted Message Passing (TRW)

The TRW is a message passing algorithm similar to the BP algorithm. How-
ever, the message update rule is different as follows:

\[
m_{pq}^i(f_q) = \min_{f_p} \{ V(f_p, f_q) + A_{pq}(D(f_p) + \sum_{r \in N_p} m_{rp}^{i-1}(f_p)) - m_{qp}^{i-1}(f_p) \}.
\]

(13)

The coefficients \( A_{pq} \) are estimated as shown in [49] as follows. The image grid is
subdivided into a set of trees such that each edge is in at least one tree. \( A_{pq} \) is
the probability that a tree, which is chosen randomly under certain distribution,
contains the edge \((p, q)\) given that it contains \( p \). Note that if \( A_{pq} = 1 \) Eq. (13)
would be identical to Eq. (11). One of the advantages of the TRW algorithm is
that it computes a lower bound on the energy. Although the original TRW does
not guarantee the increasing of the lower bound with time, Sequential TRW (TRW-
S) proposed in [46] guarantees that the lower bound estimate is not to decrease
(convergence properties). TRW-S is guaranteed to give the same performance of
the roof duality, but it is much slower [9].

E. Graph Cuts

The work in [49] illustrates that the expansion moves (a graph cuts algo-
rithm) outperforms the other competitive methods in all tested problems in terms
of accuracy and time efficiency. So this technique is used as a minimization tool in this dissertation. Note that different graph-based energy minimization methods may use different graph constructions. Also there are different rules for converting graph cuts into image labeling. For more details see [8,51]. In this section the reconstruction of the graph and the rules that are used to minimize an equation such as Eq. (10) is reviewed.

1. Graphs

The weighted undirected graph $G = (V,E)$ is a set of vertices $V$, and a set of edges $E$ connecting the vertices. Each edge is assigned a nonnegative weight. The set of vertices $V$ corresponds to the set of image pixels $P$, and some additional special nodes called terminals. These terminals correspond to the set of labels that can be assigned to an image pixels. This work deals only with graphs that have two terminals. These terminals are usually called the source $s$ and the sink $t$. An example of this graph is shown in Fig. 6. The set of edges $E$ consists of two subsets. The first subset, the $n$-links, contains edges that connect the neighboring pixels in the image. The second subset, the $t$-links, contains edges that connect the pixels with the terminals. Each edge is assigned a cost. The cost of a $t$-link connecting a node and a terminal corresponds to the penalty of assigning the corresponding label to the pixel. This cost corresponds to the second term in Eq. (10). The cost of a $n$-link between two pixels is the penalty of disconnecting them. This cost corresponds to the first term in Eq. (10).

2. Min-Cut/Max-Flow problems

An $s/t$ cut on a graph $G$ is a set of edges $E_c \subset E$ such that terminals are separated in the induced graph $G(G_c) = (V,E - E_c)$. The cut divides the set of image pixels into two disjoint subsets. No proper subset of $E_c$ separates the terminals in $G(G_c)$. Examples of valid and invalid cuts are shown in Fig. 7. The sum of weights
FIGURE 6 – An example of undirected Graph: Image’s pixels (a-i) are the graph’s nodes. n-links is constructed for 4-neighborhood system. t-link connect pixels with terminals.

FIGURE 7 – Examples of cuts on a graph. (a), (b), and (c) are valid cuts. (d) is invalid cuts; it do not separate the terminal there exist a path \( \{s, a, d, e, h, t\} \). (e) is invalid cuts; it has a subset \( \{(a, d), (b, e), (c, f)\} \) gives a valid cut.

of edges, which belong to a cut, is the cut cost \( |E_c| \). The Min-Cut problem is to find a cut that has the minimum cost among all cuts. Min-Cut/Max-Flow algorithms in combinatorial optimization show that a globally minimum \( s/t \) cut can be computed efficiently in a low-order polynomial time by computing the maximum flow from \( s \) to \( t \) [52]. Boykov and Kolmogorov [53] described a modified max-flow algorithm that significantly outperforms the original max-flow techniques. In this dissertation, this algorithm is used to find the minimum cut among all the cuts in the graph. Since the cut divides the set of image pixels into two disjoint subsets each set has one terminal, each pixel is assigned a unique label. Therefore, if the edge weights are properly set based on the energy function parameters, a minimum cost cut will correspond to a labeling with minimum value of this energy [53].

24
3. Expansion Moves Algorithm

The expansion moves algorithm was proposed by Boykov et al. [44] to minimize an energy function such as Eq. (10) with non binary variables by repeatedly minimizing an energy function with binary variables using the Max-Flow/Min-Cut method. It is an effective algorithm for minimizing discontinuity-preserving energy functions. This algorithm can be applied to pair-wise interactions that are submodular on the space of labels (e.g., Potts function) [45]. The potential function \( V(\ldots) \) is submodular if:

\[
V(l_1, l_2) + V(l_2, l_3) \geq V(l_1, l_3) + V(l_2, l_2),
\]

holds for all labels \( l_1, l_2, \) and \( l_3 \in \mathcal{L} \). A labeling \( \hat{f} \) is defined to be an \( \alpha \)-expansion move from a labeling \( f \) if every pixel either keeps its old label, \( \hat{f}_p = f_p \), or switch to a particular label \( \alpha \), \( \hat{f}_p = \alpha \). Then the algorithm cycles through the labels \( \alpha \) in some order and finds the lowest energy \( \alpha \)-expansion move from the current labeling. The algorithm terminates when there are no moves for any label with lower energy. The expansion moves algorithm gives a local minimum lies within a multiplicative factor of the global minimum . This factor depends on the potential function, e.g. for the Potts model the factor is two [44]. The outline of this algorithm is shown in Algorithm 2.

**Algorithm 2** \( \alpha \)-Expansion Move Algorithm [8]

1: Start with any arbitrary labeling \( f \)
2: Set \( success = 0 \)
3: For each label \( \alpha \in \mathcal{L} \) (any order)
   
   find \( \hat{f} = \arg \min E(f) \) among \( f \) within one \( \alpha \)-expansion1 of \( \hat{f} \)
   
   if \( E(\hat{f}) < E(\hat{f}) \) set \( \hat{f} = \hat{f} \) and \( success = 1 \)
4: If \( success = 1 \) goto 2
5: Return \( \hat{f} \)
F. Extended Roof Duality

As described in the previous section, for the case of binary pairwise MRF (i.e., \( \mathcal{L} = \{0, 1\} \)), a global minimum can be computed in polynomial time as a minimum \( s/t \) cut if every pairwise term satisfies

\[
V(0, 1) + V(1, 0) \geq V(0, 0) + V(1, 1).
\]

However, in many vision applications this submodularity condition is not satisfied. Roof duality [48] and its extended version, extended roof duality [9], can be used to minimize non-submodular functions. Roof duality can be considered as a generalization of the standard graph cut algorithm. For the submodular functions, the two algorithms give the same answer in almost the same time. For non-submodular functions, roof duality produces part of an optimal solution. The extended roof duality algorithm outperforms other algorithms in solving many problems that have been demonstrated in [9]. Thus in this dissertation, extended roof duality algorithm is used to minimize functions outside the scope of expansion moves algorithm.

1. Roof Duality

The main idea of this approach is to solve a particular linear programming relaxation of the energy Eq. (10), where the binary constraints \( f_p \in \{0, 1\} \) are replaced with \( f_p \in \{0, 1, \frac{1}{2}\} \) for every site \( p \in \mathcal{P} \). Usually, the partial labeling is defined with \( f_p \in \{0, 1, \emptyset\} \) where \( \emptyset \) means that node is unlabeled. Similar to the submodular case in the standard graph cut approach, the problem is reduced to the computation of a minimum \( s/t \) cut in a certain graph. However, the size of the graph is doubled in the non-submodular case. In addition to the spacial nodes (the source \( s \) and the sink \( t \) which correspond to labels 0 and 1), for each site \( p \in \mathcal{P} \), two nodes \( p, \bar{p} \) are added to \( \mathcal{V} \), (they correspond to the variable \( f_p \) and its complement \( \bar{f}_p = 1 - f_p \)). For each non zero term in energy Eq. (10), two directed edges
2. Probing Method

When the number of non-submodular terms is small, the roof duality works well. However, in more difficult cases it may leave many nodes unlabeled. Many extensions are proposed to enhance this technique. One of these extensions is the “probing” method introduced in [9], which can be described as follows. Let \( f \) be the output of the roof duality algorithm with node \( p \) unlabeled. By fixing \( p \) to 0 and then to 1 and run the roof duality algorithm in each case, two partial labelings \( f^0 \) and \( f^1 \) are generated. Define the set \( \mathcal{U} \) as follows:

\[
\mathcal{U} = \left[ \text{dom}(f^0) \cap \text{dom}(f^1) \right] - \left[ \text{dom}(f) \cup \{p\} \right],
\]

where \( \text{dom}(f) \) is the domain of \( f \) (the set of labeled nodes). For a global minima \( f^* \) and using the roof duality property [54], the following can be drawn

\[
f^*_p = l \Rightarrow f^*_q = f^l_q \quad \forall l \in \{0, 1\}, q \in \mathcal{U}.
\]

Thus, nodes in \( \mathcal{U} \) can be removed (by fixing or contracting) from the energy without affecting the global minimum. An illustration example is shown in Fig. 8.
ing a node to 0 and to 1 may label different sets of nodes (i.e., \( \text{dom}(f^0) \neq \text{dom}(f^1) \)).

In this case to exploit this information, a pairwise term \( V(f_p, f_q) \) is added to the energy where \( V(l, 1 - f_q^1) = C_n \) (\( C_n \) sufficiently large non-negative constant.) and all other terms are zeros.

The outline of the probing method is summarized in Algorithm 3.

**Algorithm 3** Extended Roof Duality Algorithm (Probing Method) [9]

1: Run the roof duality algorithm for the given energy.
2: Select unlabeled node \( p \), and fix it to 0 and to 1. Then run the roof duality algorithm to get \( f^0 \) and \( f^1 \). Then compute \( \mathcal{U} \).
3: Remove nodes in \( \mathcal{U} \) by fixing or contracting.
4: Add a directed constraints for all edges \( (p, q) \in \mathcal{E} \) with \( q \in \text{dom}(f^0) - \text{dom}(f^1) \) or \( q \in \text{dom}(f^1) - \text{dom}(f^0) \).
5: If the energy has changed run the roof duality again and update the unlabeled nodes.
CHAPTER III
MGRF PARAMETERS ESTIMATION

Fitting an MRF model to an image requires estimating its parameters $\gamma_q$ from a sample of the image. The literature is rich with works that propose different MGRF models, as described in Sec. II.B, which are suitable for a specific system behavior. Usually, these works identify their models’ parameters using an optimization technique. This technique tries to maximize either the likelihood or the entropy of the proposed probability distributions. This chapter proposes an analytical method to estimate the homogeneous isotropic Potts model for an asymmetric Gibbs potential function.

- **Maximum Likelihood Estimation (MLE)** is the most popular estimator used in estimating unknown parameters of a distribution (e.g., [29]). Define by $\Theta$ the vector of potential parameters (e.g., for a homogeneous anisotropic pairwise Potts model $\Theta = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]$ and for a homogeneous anisotropic Potts model with triple cliques $\Theta = [\gamma_5, \gamma_6, \gamma_7, \gamma_8]$). The Gibbs probability distribution can be represented as a function in $\Theta$ as follows:

$$P(f) = \frac{1}{Z} \exp \left( U(f, \Theta) \right),$$

and the log-likelihood function is defined by

$$L(f|\Theta) = \frac{1}{|P|} \log P(f).$$

Thus, the maximum log-likelihood estimator can be defined by

$$\Theta^* = \arg \max_{\Theta} \left( U(f, \Theta) - \log(Z(\Theta)) \right).$$

Equation (19) can be solved by the differentiation of the log-likelihood. However, the second term $\log(Z(\Theta))$ is intractable. Thus, numerical techniques are usually used to find a solution for this problem.

29
A. Related Works

In this section some popular methods used to estimate the parameters for MGRF models are discussed.

1. Coding Estimation

The coding method was proposed by Besag [28]. In this method, the image grid is partitioned into coding patterns. The codings are chosen such that a pixel and its neighbors cannot be members of the same coding pattern. This implies that the distribution of the pixel values within one coding pattern are independent on the pixel values of the other coding patterns. In order to get an efficient estimator, the number of coding patterns should be as low as possible. Thus the efficient coding of a first-order MGRF consists of two patterns (checkerboard) shown in Fig. 9 (a), and of a second-order MRF consists of four patterns as shown in Fig. 9 (b).

All pixels coded $j$ are used for $j$th set of parameter estimates, $j = 1, 2, 3, 4$. Using this coding and MRF properties, the colors of sites in each coding are conditionally independent

$$P(f_p, f_q | f_{N_p}, f_{N_q}) = P(f_p | f_{N_p})P(f_q | f_{N_p}).$$
The coding method estimates the vector parameters \( \Theta \) by finding the vector \( \Theta_j \) which maximizes the log-likelihood in coding \( j \)

\[
L_j(\Theta) = \sum_{p \in P_j} \log \left( \frac{\exp(-U(f, \Theta))}{\sum_{l \in L} \exp(-U(l, \Theta))} \right),
\]

where \( P_j \) is the set of pixels that have the code \( j \). After optimizing \( L_j(\Theta) \), the estimated vector for the second order model is defined as follows

\[
\Theta = \frac{1}{4} \sum_{j=1}^{4} \Theta_j.
\]

2. Least Square Error method (LSQR)

This method was proposed by Derin and Elliot [31], the corresponding model is described in Sec. II.B.1. They established different \( 3 \times 3 \) label blocks of pixels. For a pixel \( p \) with a label \( f_p \) and the 8 neighborhood \( N_p \), the block is \( (f_p, f_{N_p}) \). Each different \( 3 \times 3 \) block of labels establishes a block type. Define \( l_1, l_2 \) as labels of a particular pixel \( p \) with a neighborhood \( N_p \). One can formulate the following equation

\[
\sum_{q \in N_p} (V(l_1, f_q) - V(l_2, f_q)) = \log \frac{P(l_2|f_{N_p}) + \epsilon}{P(l_1|f_{N_p}) + \epsilon},
\]

where \( \epsilon \) is a small number (e.g., \( \frac{1}{512} \)) to avoid zero probability. The ratio \( \frac{P(l_2|f_{N_p})}{P(l_1|f_{N_p})} \) is estimated by counting the number of blocks of type \((l_2, f_{N_p})\) and dividing by the number of blocks of type \((l_1, f_{N_p})\). A second-order binary MGRF has 256 such equations. In order to estimate the model parameters using least square methods, one needs to solve this overdetermined system of linear equations.

3. Parameter Estimation Using Co-occurrence Probability

Cremers and Grady [33] computed the Gibbs energy \( U(f) \) from the histograms of joint co-occurrence of label pairs (or triplets). They assumed that the
co-occurrence probability for any two variables (or three variables) does not depend on other variables. Under this assumption they simplified the Gibbs energy in pairwise case to the form

\[ U(f) = -\frac{1}{\Gamma} \sum_{\{p \neq q\} \in P} P(f_p, f_q), \]  

(23)

where the constant \( \Gamma = \binom{n}{2} \) denotes the number of ways to generate such pairings divided by the number of times each pair appears in the overall product. Then potential parameters \( \gamma_{l_1l_2} \) are related to the probability of co-occurrence of labels \( l_1 \) and \( l_2 \) as follows:

\[ \gamma_{l_1l_2} = -\log P(f_p = l_1 \cap f_q = l_2). \]  

(24)

### 4. Analytical method for Potts model

Farag et al., [1] proposed an analytical approach to estimate the parameter of a homogenous isotropic MGRF Potts model. They defined the potential function of Potts model governing symmetric pairwise co-occurrences of the region labels as \( V = \{V(l_1, l_2) = \gamma \text{ if } l_1 = l_2 \text{ and } V(l_1, l_2) = -\gamma \text{ if } l_1 \neq l_2: l_1, l_2 \in \mathcal{L}\} \). To identify the homogeneous isotropic Potts model that describes the label image \( f \), they need to estimate only the potential value \( \gamma \). This parameter was obtained analytically using the Maximum Likelihood Estimator (MLE) for a generic MGRF [2]. Hence, the potential interaction is given by the following equation:

\[ \gamma = \frac{K^2}{2(K-1)} \left( \tilde{\mathbf{s}}_{\text{eq}}(f) - \frac{1}{K} \right), \]  

(25)

where \( \tilde{\mathbf{s}}_{\text{eq}}(.) \) denotes the relative frequency of the equal labels in the pixel pairs.

### 5. Others

Many different approaches were proposed to estimate the MGRF parameter. To estimate the unconditional probability distribution \( P(f) \), Olga [8] discussed different types of potential function \( V(., .) \), see Fig. 10. In all these forms the po-
potential parameter was set by hand. Boykov and Funka-Lea [5], estimated potential parameters of the Potts model using simple function that is inversely proportional to the gray level difference between the two pixels and to their distance as follows:

$$\gamma_{p,q} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma_c^2}\right) \cdot \frac{1}{\text{dist}(p, q)},$$

(26)

where $\sigma_c$ is estimated as camera noise. Many other works [4, 55, 56] used the same criteria. Usually, potential parameter of Potts model is chosen based on local intensity gradient, Laplacian zero-crossing, gradient direction, geometric, or other criteria. However, these models depend on parameters which must be set by hand.

B. The Proposed Approach For Parameter Estimation

Unlike common computer vision studies, this work adopts the pairwise and triple homogenous isotropic MGRF model to be the image model with Potts prior. Similar to Farag et al. [1], the parameter of this model is analytically estimated. However, this work focuses on asymmetric pairwise co-occurrences of the region labels. The asymmetric Potts model is chosen to provide more chances to guarantee that the Gibbs energy function is submodular, so it can be minimized using a standard graph cuts approach in polynomial time. In this case, the Gibbs potential governing asymmetric pairwise co-occurrences of the region labels can be described as follows:

$$V(f_p, f_q) = \begin{cases} 
0 & \text{if } f_p = f_q, \\
\gamma & \text{otherwise}.
\end{cases}$$

(27)
Then the MGRF model of region maps is specified by the following Gibbs probability distribution:

\[
P(f) = \frac{1}{Z} \exp \left( - \sum_{\{p,q\} \in \mathcal{N}} V(f_p, f_q) \right);
\]

\[
= \frac{1}{Z} \exp \left( - \gamma |T| \tilde{\mathcal{F}}_{\text{neq}}(f) \right).
\]  

(28)

Here, \( T = \{ \{p, q\} : p, q \in \mathcal{P}; \{p, q\} \in \mathcal{N} \} \) is the family of the neighboring pixel pairs supporting the Gibbs potentials, \(|T|\) is the cardinality of that family, and \( \tilde{\mathcal{F}}_{\text{neq}}(f) \) denotes the relative frequency of the not equal labels in the pixel pairs of that family:

\[
\tilde{\mathcal{F}}_{\text{neq}}(f) = \frac{1}{|T|} \sum_{\{p, q\} \in T} \delta(f_p \neq f_q),
\]  

(29)

where, the indicator function, \( \delta(A) \) equals 1 when the condition \( A \) is true, and zero otherwise. To completely identify the Potts model that describe the label image \( f \), the potential value \( \gamma \) have to be estimate.

1. Pairwise Clique Potential Estimation

To estimate the model parameter \( \gamma \) that specifies the Gibbs potential, the MGRF model is identified using a reasonably close first approximation of the maximum likelihood estimation of \( \gamma \). It is derived in accordance with [2] from the log-likelihood

\[
L(f|\gamma) = \frac{1}{|\mathcal{P}|} \log P(f).
\]  

(30)

Using Eq. (28), the partition function \( Z \) can be written as follows:

\[
Z = \sum_{\hat{f} \in \mathcal{F}} \exp \left( - \gamma |T| \tilde{\mathcal{F}}_{\text{neq}}(\hat{f}) \right).
\]  

(31)

Then the log-likelihood of Eq. (30) can be rewritten as follows:

\[
L(f|\gamma) = -\gamma \rho \tilde{\mathcal{F}}_{\text{neq}}(f) - \frac{1}{|\mathcal{P}|} \log \sum_{\hat{f} \in \mathcal{F}} \exp \left( - \gamma |T| \tilde{\mathcal{F}}_{\text{neq}}(\hat{f}) \right),
\]  

(32)
where \( \rho = \frac{|T|}{|P|} \). The approximation is obtained by truncating the Taylor’s series expansion of \( L(f|\gamma) \) to the first three terms in the close vicinity of the zero potential, \( \gamma = 0 \):

\[
L(f|\gamma) \approx L(f|0) + \gamma \frac{dL(f|\gamma)}{d\gamma} \bigg|_{\gamma=0} + \frac{1}{2} \gamma^2 \frac{d^2L(f|\gamma)}{d\gamma^2} \bigg|_{\gamma=0} .
\]

(33)

The first derivative of the log-likelihood Eq. (32) is given by

\[
\frac{dL(f|\gamma)}{d\gamma} = -\rho \delta_{\text{neq}}(f) + \rho \sum_{\hat{f} \in \mathcal{F}} \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right)
\]

\[
\sum_{\hat{f} \in \mathcal{F}} \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right)
\]

\[
= -\rho \delta_{\text{neq}}(f) + \rho \mathbb{E}\{\delta_{\text{neq}}(\hat{f})|\gamma\},
\]

(34)

where \( \mathbb{E}\{.\} \) denotes math expectation. By replacing \( \delta_{\text{neq}}(.) \) with \( 1 - \delta_{\text{eq}}(.) \), the first derivative becomes:

\[
\frac{dL(f|\gamma)}{d\gamma} = -\rho (1 - \delta_{\text{eq}}(f)) + \rho \mathbb{E}\{(1 - \delta_{\text{eq}}(\hat{f})|\gamma\} = \rho (\delta_{\text{eq}}(f) - \mathbb{E}\{\delta_{\text{eq}}(\hat{f})|\gamma\}).
\]

(35)

If \( \gamma = 0 \), this MGRF becomes the Independent Random Field (IRF) of of equiprobable \( K \) labels. Every label has the same probability \( 1/K \), and the expectation can be computed as follows:

\[
\mathbb{E}\{\delta_{\text{eq}}(\hat{f})|0\} = \frac{1}{|T|} \sum_{\{p,q\} \in \mathcal{T}} \mathbb{E}\{\delta(f_p = \hat{f}_q)\} = \frac{1}{|T|} |T| \mathbb{E}\{\delta(f_p = \hat{f}_q)\} = \frac{1}{K}.
\]

(36)

Thus in the vicinity of the origin \( \gamma = 0 \), the first derivative of the log-likelihood is equal to

\[
\frac{dL(f|\gamma)}{d\gamma} \bigg|_{\gamma=0} = \rho (\delta_{\text{eq}}(f) - \frac{1}{K}).
\]

(37)

The second derivative of the log-likelihood is given by

\[
\frac{d^2L(f|\gamma)}{d\gamma^2} = -\rho^2 |P| \sum_{\hat{f} \in \mathcal{F}} \delta_{\text{neq}}^2(\hat{f}) \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right) \sum_{\hat{f} \in \mathcal{F}} \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right)
\]

\[
\left( \sum_{\hat{f} \in \mathcal{F}} \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right) \right)^2
\]

\[
- \rho^2 |P| \left( \sum_{\hat{f} \in \mathcal{F}} \delta_{\text{neq}}(\hat{f}) \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right) \right)^2
\]

\[
\left( \sum_{\hat{f} \in \mathcal{F}} \exp \left( -\gamma |T| \delta_{\text{neq}}(\hat{f}) \right) \right).
\]

(38)
In a similar way, in the vicinity of the origin $\gamma = 0$, the second derivative of the log-likelihood is equal to

$$
\frac{d^2L(f|\gamma)}{d\gamma^2}|_{\gamma=0} = -\rho^2|P| \left( \mathbb{E}\{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})^2|\gamma\} - \mathbb{E}^2\{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})|\gamma\} \right)
$$

(38)

$$
= -\rho^2|P| \text{ var } \{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})|\gamma\} = -\rho^2|P| \text{ var } \{(1 - \tilde{\mathcal{F}}_{\text{eq}}(\hat{f}))|\gamma\}
$$

$$
= -\rho^2|P| \text{ var } \{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})|\gamma\}.
$$

For the IRF the frequency variance can be estimated as follows:

$$
\text{ var } \{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})|0\} = \mathbb{E}\{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})^2|0\} - \mathbb{E}^2\{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})|0\}
$$

$$
= \mathbb{E}\left\{ \frac{1}{|T|} \sum_{\{p,q\} \in T} \delta(f_p = f_q) \right\} - \frac{1}{K^2}
$$

$$
= \frac{1}{|T|^2} \mathbb{E}\left\{ \sum_{\{p,q\} \in T} \delta(f_p = f_q) \right\} + \sum_{\{p,q\} \in T} \delta(f_p = f_q) \sum_{\{i\neq p,j\neq q\} \in T} \delta(f_i = f_j) - \frac{1}{K^2}
$$

$$
= \frac{1}{|T|^2} \left( |T| \frac{1}{K} + |T|(|T| - 1) \frac{1}{K^2} \right) - \frac{1}{K^2}
$$

$$
= \frac{1}{\rho|P|} \frac{K-1}{K^2}.
$$

(39)

Thus in the vicinity of the origin, the second derivative of the log-likelihood is equal to

$$
\frac{d^2L(f|\gamma)}{d\gamma^2}|_{\gamma=0} = -\rho^2|P| \text{ var } \{\tilde{\mathcal{F}}_{\text{eq}}(\hat{f})|0\} = -\rho \frac{K-1}{K^2}.
$$

(40)

Finally, the approximated log-likelihood Eq. (33) becomes

$$
L(f|\gamma) \approx -|P| \log K + \rho\gamma (\tilde{\mathcal{F}}_{\text{eq}}(f) - \frac{1}{K}) - \frac{1}{2} \gamma^2 \rho \frac{K-1}{K^2}.
$$

(41)

For the approximate log-likelihood of Eq. (41), let $\frac{dL(f|\gamma)}{d\gamma} = 0$. This results in the following approximate MLE of $\gamma$:

$$
\gamma^* = \frac{K^2}{K-1} (\tilde{\mathcal{F}}_{\text{eq}}(f) - \frac{1}{K})
$$

$$
= \frac{K^2}{K-1} (1 - \tilde{\mathcal{F}}_{\text{eq}}(f) - \frac{1}{K})
$$

$$
= \frac{K^2}{K-1} (\frac{K-1}{K} - \tilde{\mathcal{F}}_{\text{eq}}(f)).
$$

(42)
2. Triple Clique Potential Estimation

The Gibbs potential governing asymmetric triple co-occurrences of the region labels can be described as follows:

\[ V(f_p, f_q, f_r) = \gamma (1 - \delta(f_p = f_q = f_r)). \quad (43) \]

Following the same method used in pairwise potentials, one can prove that the potentials of the third order cliques have the same analytical form of Eq. (42) but with the frequency

\[ \tilde{\varphi}_\text{neq}(f) = \frac{1}{|T|} \sum_{(p,q,r) \in T} (1 - \delta(f_p = f_q = f_r)), \quad (44) \]

where \(T = \{ \{p,q,r\} : p, q, r \in \mathcal{P}; \{p,q,r\} \in \mathcal{N} \}\) is the family of the neighboring pixel triples supporting the Gibbs potentials.

C. Experiments

The robustness of the proposed method for estimating Gibbs potentials of the Potts model is tested by applying it on simulated texture images with known potential values. The simulated texture images is generated using Gibbs sampler approach [10] which is explained in Algorithm 4. The idea of the synthesis process is to find the configuration \(f\) in \(\mathcal{F}\) which maximizes the probability \(P(f)\). The advantage of Algorithm 4 is that it eliminates the need for computing the partition function.

To assess the robustness of the proposed approach, many experiments are conducted. In the first experiment, four binary different realizations of homogeneous isotropic Potts model are generated. Samples of these realizations for images of size 128 × 128 are shown in Fig. 11. To get accurate statistics, 100 realizations are generated from each type. The proposed method is used to estimate the model parameter \(\gamma\) for these data sets. The means and the variances (written between parentheses) of the 100 realizations for each type are shown in Table (1).
Algorithm 4 Gibbs Sampler Algorithm [10]

1: Start with any random labeling $f$

2: for all $p \in \mathcal{P}$ do

3: Choose $l \in \mathcal{L}$ at random and let $\hat{f}_p = l$, and $\hat{f}_q = f_p$ for all $q \neq p$

4: Let $P = \min\{1, P(F = \hat{f})/P(F = f)\}$.

5: Replace $f$ by $\hat{f}$ with probability $P$.

6: end for

7: Repeat (2) $N_{\text{iter}}$ times

---

FIGURE 11 – Samples of synthesized binary images of size $128 \times 128$

---

TABLE 1

ACCURACY OF THE PROPOSED PARAMETER ESTIMATION METHOD FOR BINARY IMAGES OF SIZE $128 \times 128$

<table>
<thead>
<tr>
<th>Actual parameter $\gamma$</th>
<th>0.1</th>
<th>0.75</th>
<th>1.0</th>
<th>1.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed $\gamma^*$</td>
<td>0.12 (0.009)</td>
<td>0.77 (0.014)</td>
<td>1.04 (0.013)</td>
<td>1.78 (0.013)</td>
</tr>
<tr>
<td>CM</td>
<td>0.94 (0.003)</td>
<td>1.02 (0.0)</td>
<td>1.09 (0.011)</td>
<td>1.79 (0.066)</td>
</tr>
<tr>
<td>LSQR</td>
<td>0.11 (0.016)</td>
<td>0.64 (0.041)</td>
<td>0.85 (0.054)</td>
<td>1.79 (0.091)</td>
</tr>
</tbody>
</table>
In the second experiment, four different realizations of Potts model are generated with 32 colors (i.e., $K = 32$). Samples of these realizations for images of size $64 \times 64$ are shown in Fig. 12. Also, 100 realizations are generated from each type. The proposed method is used to estimate the model parameter $\gamma$ for these data sets. The means and the variances of the 100 realizations for each type are shown in Table (2). Figures 13, 14, and 15 show more samples of the four realizations for different image sizes. The estimation results are shown in Tables (3), (4), and (5), respectively. For a comparison purpose, the estimations of the Coding Method (CM) [28] and Least Square Error method (LSQR) [31] are also illustrated. CM allows for an easy formulation of the estimator for auto-binomial model. However, it is generally considered difficult to use reliably [31]. Picard in [57] confirmed that CM’s performance varies widely for different data, which means CM works well for some images (e.g., Figures 11-15 c and d), but poorly for others (e.g., Figures 11-15 b). Also, Picard in [57] mentioned that CM’s estimations sometimes need
FIGURE 13 – Samples of synthesized images of size $128 \times 128$

(a) $\gamma = 0.5$  (b) $\gamma = 5$
(c) $\gamma = 10$  (d) $\gamma = 25$

TABLE 3
ACCURACY OF THE PROPOSED PARAMETER ESTIMATION METHOD FOR IMAGES OF SIZE $128 \times 128$

<table>
<thead>
<tr>
<th>Actual parameter $\gamma$</th>
<th>0.5</th>
<th>5</th>
<th>10</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed $\gamma^*$</td>
<td>0.51 (0.029)</td>
<td>5.4 (0.05)</td>
<td>10.1 (0.06)</td>
<td>25.7 (0.11)</td>
</tr>
<tr>
<td>CM</td>
<td>0.43 (0.017)</td>
<td>12.3 (0.07)</td>
<td>13.1 (0.07)</td>
<td>22.7 (0.83)</td>
</tr>
</tbody>
</table>
FIGURE 14 – Samples of synthesized images of size 256 × 256

TABLE 4
ACCURACY OF THE PROPOSED PARAMETER ESTIMATION METHOD FOR IMAGES OF SIZE 256 × 256

<table>
<thead>
<tr>
<th>Actual parameter γ</th>
<th>0.5</th>
<th>5.0</th>
<th>10.0</th>
<th>25.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed γ*</td>
<td>0.51 (0.016)</td>
<td>5.5 (0.02)</td>
<td>10.2 (0.03)</td>
<td>25.7 (0.05)</td>
</tr>
<tr>
<td>CM</td>
<td>0.43 (0.01)</td>
<td>12.3 (0.03)</td>
<td>13.1 (0.04)</td>
<td>22.34 (0.36)</td>
</tr>
</tbody>
</table>

...to be adjusted upward. In this experiment, CM estimations of Potts model parameters with 32 colors are adjusted by 10. Another disadvantage of CM is that the used optimization technique requires an initial guess for the solution and convergence conditions and may run into local optima without reaching the proper solution [11] as shown in cases of binary realizations for γ = 0.1 and 0.75, and for 32 colors realization for γ = 5. Derin and Elliott in their implementation LSQR, claimed accurate estimation by using the most frequently occurring blocks types, however, there is no consistent way to establish these types of blocks [11]. Also, as described in Sec.III.A.2, for binary realizations (e.g., Figures 11) to estimate the
FIGURE 15 – Samples of synthesized images of size $512 \times 512$

TABLE 5
ACCURACY OF THE PROPOSED PARAMETER ESTIMATION METHOD FOR IMAGES OF SIZE $512 \times 512$

<table>
<thead>
<tr>
<th>Actual parameter $\gamma$</th>
<th>0.5</th>
<th>5</th>
<th>10</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed $\gamma^*$</td>
<td>0.51 (0.008)</td>
<td>5.5 (0.01)</td>
<td>10.3 (0.01)</td>
<td>25.9 (0.03)</td>
</tr>
<tr>
<td>CM</td>
<td>0.43 (0.004)</td>
<td>12.3 (0.01)</td>
<td>13.03 (0.02)</td>
<td>22.2 (0.17)</td>
</tr>
</tbody>
</table>
model parameters using LSQR, one needs to solve an overdetermined system of linear equations with up to $2^8 = 256$ equations. This is not practical in the case of realizations with 32 colors, where the overdetermined system of equations has up to $32^8$ equations.

The proposed algorithm is tested on anisotropic Potts models. Four different Potts models with parameters $[\gamma_1 = 25, \gamma_2 = \gamma_3 = \gamma_4 = 0]$, $[\gamma_2 = 25, \gamma_1 = \gamma_3 = \gamma_4 = 0]$, $[\gamma_3 = 25, \gamma_1 = \gamma_2 = \gamma_4 = 0]$, and $[\gamma_4 = 25, \gamma_1 = \gamma_2 = \gamma_3 = 0]$ and with 32 colors are generated. Samples of these realizations for images of size $128 \times 128$ are shown in Fig. 16. Also, 100 realizations are generated from each type, and the proposed method is used to estimate the model parameters for these data sets. The means and the variances of the 100 realizations for each type are shown in the figure.

In the last experiment, two different realizations of Potts model with triple cliques are synthesized, samples are shown in Fig. 17. The means and the variances of the estimated parameters of 100 samples from each type are also shown in Fig. 17.

D. Conclusions

This chapter proposed an analytical method to estimate the homogeneous isotropic Potts model with asymmetric Gibbs potential function. The experiments showed that the proposed analytical estimates of the MGRF parameters outperformed the classical methods (e.g., CM and LSQR). Also, the proposed approach was tested in an anisotropic model and performed well. The statistical results highlighted the robustness of the proposed analytical estimation approach over the conventional methods. This accurate identification of the MGRF model will demonstrate promising results in segmentation problem as will be discussed in detail in the following chapters.
FIGURE 16 – Results of the proposed method for estimating anisotropic Potts model parameters for images of size $128 \times 128$: (a) $[25.6(0.05) \ 0.003(0.08) \ 0.001(0.08) \ 0.002(0.08)]$, (b) $[0.003(0.09) \ 25.6(0.05) \ 0.004(0.09) \ 0.001(0.09)]$, (c) $[0.001(0.08) \ 0.001(0.09) \ 0.005(0.09) \ 25.9(0.05)]$, and (d) $[0.01(0.09) \ 0.01(0.09) \ 25.9(0.05) \ 0.01(0.09)]$. 
FIGURE 17 – Samples of synthesized images with 32 colors and high order cliques (a) Sample of realizations generated with $\gamma = 5$ and (b) Sample of realizations generated with $\gamma = 10$
CHAPTER IV
A NOVEL UNSUPERVISED GRAPH CUTS APPROACH FOR N-D MULTIMODAL IMAGE LABELING

This chapter proposes a new unsupervised MAP-based labeling (image segmentation and image restoration) framework of N-D multimodal gray scale images. As described in Sec. II.C.1, the input image and its desired map (labeled image) are described by a joint Markov-Gibbs random field model of independent image signals and interdependent region labels. However, the main focus in the proposed approach is on more accurate model identification for the MGRF model and the gray levels distribution model. The parameter of the MGRF model is analytically estimated as described in the previous chapter III.B. In this chapter, Sec.IV.B introduces an accurate model of the gray levels distribution where the gray levels distribution of the given image is approximated by a Linear Combination of Gaussians (LCG). In order to make the approach unsupervised, Sec. IV.B.1 proposes a new technique based on maximizing a new joint likelihood function to estimate the number of classes in the given image. An initial labeling (pre-labeled image) is generated using the LCG-model. Then the $\alpha$-expansion move Algorithm 2 iteratively refines the initial labeled image by using the MGRF with analytically estimated potential. Experimental results show that the developed technique gives promising accurate results compared to other known algorithms.

A. Introduction

Image labeling, segmentation and restoration, is one of the most important low-level computer vision tasks. This chapter addresses the problem of accurate unsupervised labeling of multimodal gray scale images, where each region of in-
terest relates to a single dominant mode (or peak) of the empirical marginal probability distribution of gray levels. The goal of the proposed algorithm is to extract the major regions (e.g. classes, patches, objects) of the given multimodal image while ignoring the small intra-region variations, which is known as image labeling.

Recently, energy-based algorithms appeared as robust image labeling approaches. Similarly, the proposed approach uses graph cuts technique to minimize the energy function that is discussed in Sec. II.C.1:

$$E(f) = \sum_{\{p,q\} \in N} V(f_p, f_q) + \sum_{p \in P} D(f_p).$$  \hspace{1cm} (45)

The literature is rich with image labeling techniques. However, only some whose basics depend on the energy optimization are discuss here. Greig et al. [58] discovered the power of graph cuts algorithms from combinatorial optimization, and showed that graph cuts can be used for binary image restoration. The problem was formulated as MAP estimation of a MRF. Shi and Malik [6] proposed the normalized cut criteria, an unbiased measure of both the total dissimilarity between the different image regions as well as the total similarity within the image regions, for graph partitioning. To compute the minimum cut, which corresponds to optimum segmentation, they solved an eigenvalue system. Boykov and Jolly [55] proposed a framework that uses s/t graph cuts to get a globally optimal object extraction method for N-dimensional images. They minimized a cost function which combines region and boundary properties of segments as well as topological constraints. That work illustrated the effectiveness of formulating the object segmentation problem via graph cuts. Since Boykov and Jolly introduced their graph cuts segmentation technique in their paper [55], it became one of the leading approaches in interactive N-D image segmentations, and many publications extended this work in different directions. Blake et al. [59] used a mixture of the Markov-Gibbs random field (MGRF) to approximate the regional properties of segments and the spatial interaction between segments. Geo-cuts [60] combines geometric cues with energy function. GrabCut [61] reduces the human interaction
by using the iterative graph cut approach. Obj-cuts [62] combines the object detection with the segmentation, and incorporates the global shape priors in MRF. To overcome the time complexity and memory overhead of Boykov and Jolly’s approach for high resolution data, Lombaert et al. [63] performed graph cuts on a low-resolution image/volume and propagated the solution to the next higher resolution level by only computing the graph cuts at that level in a narrow band surrounding the projected foreground/background interface. Instead of minimizing the energy function Eq. (45) using Max-flow/Min-cut method, Keuchel [64] solved the multiclass image labeling problem using a semidefinite relaxation technique. This technique makes the energy form less restrictive, and the shape concept is imposed into the energy function. However, this increases the computational time dramatically.

Although interactive segmentation imposes some useful topological constraints, it depends on the user inputs which highly affects the labeling results. Unlike previous graph cuts based segmentation and restoration techniques, in the proposed approach, no user interaction is needed; instead, the image is initially pre-labeled using its gray levels. Indeed, to model the low level information in the given image, the gray levels distribution of this image is precisely approximated with a linear combination of Gaussian distributions with positive and negative components. One of the contributions of this work is that the number of dominant modes in the LCG model (number of classes in the given multimodal image) is determined by using a new technique based on maximizing a new joint likelihood function. To overcome the intra-region variations, the proposed approach does not depend only on the image gray levels but it uses the graph cuts approach to combine the image gray levels information and the spatial relationships between the region labels. As explained in Sec. III.A.5, the potentials of Potts model, which describe the spatial pairwise interaction between two neighboring pixels, are usually estimated using simple functions that are proportional to the gray levels difference between the two pixels and inversely proportional to their distance. Unlike these
conventional techniques, in this dissertation the potentials of Potts model are estimated using a new analytical approach which is presented in Sec. III.B. After the image is initially labeled, the energy function Eq. (45) is formulated using both image appearance models (LCG and MGRF models). This function is minimized using a multi-way graph cuts Algorithm 2, described in Sec. II.E.3, to get the final and optimal segmentation of the input image.

B. The Conditional Image Model

As discussed in Sec. II.C.1 to solve the labeling problem, one needs to estimate the unconditional $P(f)$ and the conditional $P(I|f)$ image models. The former is completely identified by estimating the parameter of MGRF as presented in Sec. III.B. The latter is discussed in this section.

Many works were presented in the computer vision field to identify this model. Some of these related works are reviewed in this section. To restore the original image from a noisy version, Olga [8] estimate the conditional distribution of the noisy image given the map as follows

$$P(I_p | f_p) = A_p \cdot \exp(-D_p(f_p)),$$

where $A_p$ is normalizing constant, and $D_p(f_p) = (I_p - f_p)^2$. In her work, she assumed that the number and the values of the labels are known. To segment an object from its background, in the works by Boykov et al., [4,5,55,56], the user manually selects some seeds, as shown in Fig. 18. They used the intensity of these seeds to estimate the conditional distributions of the object and the background. Blake et al. [59] made the user to draw a fat pen trail enclosing the object boundary. Therefore; the image is classified to object, background, and unknown region. They use this information to estimate the conditional distribution using Gaussian mixture Markov random field model. Although user interaction imposes some useful topological constraints, it depends on the user inputs which highly affects the labeling results as shown in Fig. 19. Unlike these techniques, in this work
the conditional distribution is estimated from the given multimodal image data, intensity distribution.

To accurately estimate this conditional distribution $P(I|f)$, the gray levels marginal density of each class is approximated using a LCG with $C_{p,l}$ positive and $C_{n,l}$ negative components as follows:

$$P(I_p|f_p) = P(g|l) = \sum_{r=1}^{C_{p,l}} w_{p,r,l} \varphi(g|\theta_{p,r,l}) - \sum_{s=1}^{C_{n,l}} w_{n,s,l} \varphi(g|\theta_{n,s,l}), \quad (47)$$

where, $\varphi(g|\theta)$ is a Gaussian density with parameter $\theta$ (mean $\mu$ and variance $\sigma^2$), $w_{p,r,l}$ denotes the $r^{th}$ positive weight in class $l$, $w_{n,s,l}$ denotes the $s^{th}$ negative weight in class $l$. The summation of these weights is one: $\sum_{r=1}^{C_{p,l}} w_{p,r,l} - \sum_{s=1}^{C_{n,l}} w_{n,s,l} = 1$. In order to estimate the parameters of the LCG model, the modified EM algorithm [1] is used to deal with the positive and negative components. In the modified EM algorithm [1], the number of classes $K$ and the initial parameters of its dominant modes are set manually. In this dissertation, these parameters are estimated by a new technique described in the following section.
1. Dominant Modes Estimation

To complete the proposed modeling, one needs to estimate the number of image classes. Assume for any given multimodal image that its number of classes is equal to the number of dominant modes (peaks in the image gray levels frequency distribution), and each dominant mode is roughly approximated with a single Gaussian distribution. In this dissertation, a new technique is developed using Akaike Information Criterion (AIC)-type criterion [65] to estimate the number of classes in the given multimodal image. The main idea behind this technique is that the image is described by a mixture of Gaussian distributions and the number of dominant modes is estimated by finding the minimum number of Gaussian distributions that maximizes the likelihood function of this model. Consider the likelihood function of this model is defined as

$$\ell(\theta, I) = \prod_{p \in P} \sum_{j=1}^{k} \pi_j \varphi(\theta_j, I_p),$$  \hspace{1cm} (48)

where $k$ is number of components, the prior $\pi$’s are constrained by $\sum_j \pi_j = 1$. Let $\Delta_{pj} \in \{0, 1\}$ be a set of indicator variables for mixture components independent of the input $I$. Note $\sum_j \Delta_{pj} = 1$ as well as $\Delta_{pj}$ are independent for distinct pixels and $P(\Delta_{pj} = 1 \mid I) = \frac{\pi_j \varphi(\theta_j, I_p)}{\sum_{j=1}^{k} \pi_j \varphi(\theta_j, I_p)}$. Given the set of indicators $\Delta = \{\Delta_{pj}\}$ and the input $I$ the complete log-likelihood is given by

$$L(\theta, \Delta, I) = \sum_{p,j} \Delta_{pj} \log \varphi(\theta_j, I_p).$$  \hspace{1cm} (49)

Since $\Delta$ is actually unknown the “partial” log-likelihood is suggested to describe the mixture models:

$$L(\theta, I) = \sum_{p,j} \hat{\Delta}_{pj} \log \varphi(\theta_j, I_p),$$  \hspace{1cm} (50)

where $\hat{\Delta}_{pj}$ is the posterior probability of the label $j$ given the input image. And it is defined as $\hat{\Delta}_{pj} = P(\Delta_{pj} = 1 \mid I)$. Given model component penalty $\mathcal{N}$, the “partial”
likelihood function leads to a “partial” AIC (pAIC)

\[ pAIC \propto \sum_{p,j} \hat{\Delta}_{pj} \log \varphi(\theta_j, I_p) - \mathcal{N}(k + 1) \]

\[ = \sum_{p,j} \hat{\Delta}_{pj} (\log \varphi(\theta_j, I_p) - \mathcal{N}(k + 1)/n) \equiv D(k). \tag{51} \]

**Sufficient Condition for Monotonicity of pAIC** Let \( \hat{\pi}_j = \sum_p \hat{\Delta}_{pj}/n \). For given values of the parameter \( \pi \), \( \theta \) and \( \Delta \) one would like to increase RHS of Eq. (51) by assigning \( \min_j \hat{\pi}_j = 0 \) and re-weighting remaining \((k - 1)\) \( \hat{\pi}'s \) so as to satisfy the constrain \( \sum_j \hat{\pi}_j = 1 \). This could be then later used in the iterative steps of the EM-type procedure. Re-label the mixtures so as to have \( \min_j \hat{\pi}_j = \hat{\pi}_1 \). Denote modified \( D(k) \) by \( \hat{D}(k - 1) \), \( A = \min_{p,j \geq 2} \log \varphi(\theta_j, I_p) \), and \( B = \max_p \log \varphi(\theta_1, I_p) \). Note \( \sum_{p,j \geq 2} \hat{\Delta}_{pj} = n(1 - \hat{\pi}_1) \) and denote \( \log(\varphi(\theta_j, I_p)) \) by \( \varphi_{pj} \). Consider

\[ \hat{D}(k - 1) - D(k) \]

\[ = \sum_{p,j \geq 2} \hat{\Delta}_{pj} \left( \frac{\varphi_{pj} - \mathcal{N}k/n}{1 - \hat{\pi}_1} - \varphi_{pj} + \mathcal{N}(k + 1)/n \right) \]

\[ - \sum_p \hat{\Delta}_{p1} (\varphi_{p1} - \mathcal{N}(k + 1)/n) \]

\[ = \sum_{p,j \geq 2} \hat{\Delta}_{pj} \left( \frac{\varphi_{pj} - \mathcal{N}k/n}{1 - \hat{\pi}_1} - \varphi_{pj} + \mathcal{N}(k + 1)/n \right) \]

\[ - \sum_p \hat{\Delta}_{p1} \varphi_{p1} + \mathcal{N}(k + 1)\hat{\pi}_1 \]

\[ \geq (A - \mathcal{N}k/n)(1 - \hat{\pi}_1)n\hat{\pi}_1 - nB\hat{\pi}_1 + \mathcal{N}(k + 1)\hat{\pi}_1 \]

\[ = n\hat{\pi}_1(A - B) + \mathcal{N}. \tag{52} \]

Thus if the condition

\[ \hat{\pi}_1(A - B) + \mathcal{N}/n \geq 0, \tag{53} \]

is satisfied then \( \hat{D}(k - 1) - D(k) \geq 0 \) and the pAIC is increased as a result of the adjustment. The proposed algorithm is summarized in Algorithm 5.

To emphasize the ability of the pAIC algorithm in detecting the number of classes in the multimodal images, the proposed pAIC algorithm is tested using
Algorithm 5 pAIC-EM Algorithm

1: Initialize the estimates of the model parameters $\hat{\pi}, \hat{\theta}$ over-fitting the number of mixtures $k$

2: Perform the expectation step of the EM algorithm

3: For the smallest $\hat{\pi}$ check the condition Eq. (53).

   If it is satisfied, remove the corresponding component and adjust the remaining $\hat{\pi}$’s,

   otherwise do nothing

4: Perform the maximization step of EM

5: Repeat 2-4 until pAIC does not change by more than pre-specified error

Different multimodal images. Figures (20-23) show samples of pAIC results for bi-modal, 3-modal, 4-modal, and 5-modal synthetic images, and illustrate that each log likelihood is maximum at the correct number of classes. Since the synthetic images come from Gaussian mixture distributions, the resultant distributions which are created by approximating only the dominant modes of the probability density function is almost sufficient to give accurate solutions.

However, for real images this is not the case, so a more accurate model is needed. The latter is the LCG model with positive and negative components. In Fig. 24, (a) and (b) show the output of the pAIC-EM algorithm for a synthetic tri-modal image that is generated using a Gaussian mixture with positive and negative components. (c) shows the normalized absolute error between the empirical and estimated densities. (d) shows the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm. (e) shows the empirical and estimated densities. Finally (f) shows the marginal densities with the best thresholds. The proposed algorithm was tested on real images. Fig. 25 shows a typical human chest Computer Tomography (CT) slice (a), its empirical marginal grey levels distribution approximated with the dominant normal mixture (b), and the log likelihood maximum at 2 (c)
FIGURE 20 – pAIC result for a bimodal synthetic image (a) Empirical and estimated densities, and the 2 Gaussian components. (b) The log likelihood (maximum at 2).

FIGURE 21 – pAIC result for a 3-modal synthetic image (a) Empirical and estimated densities, and the 3 Gaussian components. (b) The log likelihood (maximum at 3).
FIGURE 22 – pAIC result for a 4-modal synthetic image (a) Empirical and estimated densities, and the 4 Gaussian components. (b) The log likelihood (maximum at 4).

FIGURE 23 – pAIC result for a 5-modal synthetic image (a) Empirical and estimated densities, and the 5 Gaussian components. (b) The log likelihood (maximum at 5).
FIGURE 24 – Non Gaussian 3-class result: (a) and (b) show the output of the pAIC-EM algorithm, (c) shows the normalized absolute error between the empirical and estimated densities, (d) shows the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm, (e) shows the empirical and estimated densities, and (d) shows the marginal densities with the best thresholds.
FIGURE 25—Result for CT Lung slice: (a) The CT slice, (b) and (c) the output of the pAIC-EM algorithm, (d) the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm, (e) the empirical and estimated densities, and (f) the marginal densities with the best threshold.
FIGURE 26 – Result for 3-class MRA slice: (a) The MRA slice, (b) and (c) the output of the pAIC-EM algorithm, (d) the dominant component generated by pAIC-EM and the refining components, positives and negatives, generated by the modified EM algorithm, (e) the empirical and estimated densities, and (f) the marginal densities with the best thresholds.
(note (b) and (c) are pAIC-EM outputs). The two dominant modes represent the
darker lung area and its brighter background, respectively. Also, Fig. 25 shows
the 12 components of the final LCG (d), the empirical and estimated densities (e),
and the final LCG approximation of each class for the best separation threshold
t = 109 (f) (note (d), (e), and (f) are mEM outputs). Fig. 26 shows a Magnetic Resonance Angiography (MRA) slice (a), its empirical marginal grey levels distribution approximated with the dominant normal mixture (b), and the log likelihood maximum at 3 (c) (note (b) and (c) are pAIC-EM outputs). The three dominant modes represent dark bones and fat, brain tissues, and bright blood vessels, respectively. Also, Fig. 26 shows the 9 components of the final LCG (d), the empirical and estimated densities (e), and the final LCG approximation of each class for the best separation thresholds t_1 = 53, and t_2 = 191 (f) (note (d), (e), and (f) are mEM outputs). For all experiments, the initial parameters are k = 10 Gaussians with \( \hat{\theta}_j \) (\( \mu_j = j \times [Q - 1]/[k] \) and \( \sigma_j^2 = 5 \)), and \( \hat{\pi}_j = 1/k \). The model component penalty \( \mathcal{M} \) can be easily selected to be greater than the increasing in the likelihood that happens by adding a one Gaussian distribution to the model.

C. Graph Cuts-based Optimal Labeling

After the image models were presented, now the goal is to estimate the desired map \( f \) by minimizing the energy function Eq. (45). The flow chart of the complete algorithm is shown in Fig. 27. To minimize this energy, the input image is initially labeled based on its gray levels probabilistic model described in Sec.IV.B. Then the resulting labeled image is used as the best initialization to the \( \alpha \)-expansion move algorithm described in Sec. II.E.3. The \( \alpha \)-expansion move algorithm repeatedly minimizes the energy function Eq. (45), which is defined over a finite set of labels by minimizing another version of this function with binary variables using Max-flow/Min-cut method. In each iteration of \( \alpha \)-expansion move algorithm, the updated labeled image is used to update the MGRF potentials \( \gamma \) as
in Eq. (42). To minimize this binary version of the energy function, a weighted undirected graph is created with vertices corresponding to the set of image pixels/voxels, $P$, and two special terminal vertices $s$ (source, the new label “0”), and $t$ (sink, the current label “1”). The neighborhood system $\mathcal{N}$, is chosen to be the nearest 4-neighborhood in the 2D case (or 6-neighborhood in the 3D case). Each edge in the set of edges connecting the graph vertices is assigned a nonnegative weight as follows. For each $p, q \in P$, and $\{p, q\} \in \mathcal{N}$, the weights are shown in Table (6). Then the optimal labeling is obtained by finding the minimum cost cut on this
graph. The minimum cost cut is computed in polynomial time for two terminal graph cuts with positive edge weights via $s/t$ Min-Cut/Max-Flow algorithm [53].

**D. Experiments and Discussion**

To assess the performance of the proposed approach, it is tested on several N-D multimodal images. First, the advantage of the adaptive analytical approach that is proposed to compute the spatial interaction parameter $\gamma$ is highlighted. As shown in Fig. 28, for a small value of $\gamma$ the resultant labeled image will be noisy (it emphasizes the data, the 2nd term in Eq. (45)). For a large value of $\gamma$ the corresponding labeled image is oversmoothed and some classes disappeared. For this image, Fig. 29 shows the change of the relative error with $\gamma$. Also values of $\gamma$ computed with the proposed adaptive analytical approach are shown. These values correspond to the range of $\gamma$ that gives the minimum error; this emphasizes the correctness of the proposed approach.

**Validation**: proposed approach is compared with both the mean shift algorithm [66] and the normalized cuts algorithm [6]. Note that when using these codes, several trials have been conducted in order to select the tuning parameters that give the best results. These parameters are (EDISON: spatial and color bandwidths $h_s$ and $h_c$; Min. region $M$). (NCUTS: number of segments ‘$nsg$’, offset of

---

**Table 6**

**GRAPH EDGE WEIGHTS.**

<table>
<thead>
<tr>
<th>Edge</th>
<th>Weight</th>
<th>for</th>
</tr>
</thead>
<tbody>
<tr>
<td>${p, q}$</td>
<td>$\gamma$</td>
<td>$f_p \neq f_q$</td>
</tr>
<tr>
<td>${p, q}$</td>
<td>0</td>
<td>$f_p = f_q$</td>
</tr>
<tr>
<td>${s, p}$</td>
<td>$-\ln[P(I_p</td>
<td>“1”)])$</td>
</tr>
<tr>
<td>${p, t}$</td>
<td>$-\ln[P(I_p</td>
<td>“0”)])$</td>
</tr>
</tbody>
</table>
FIGURE 28 – Effect of choosing $\gamma$ (a) Original image (b) Noisy output $\gamma = 0.05$ (c) Over smoothed output $\gamma = 5$

FIGURE 29 – The changes of the relative error with $\gamma$, proposed adaptive analytical approach values shown in red asterisk
TABLE 7
1D STATISTICAL RESULTS.

<table>
<thead>
<tr>
<th>(\sigma_n)</th>
<th>0.5</th>
<th>0.7</th>
<th>1.0</th>
<th>1.2</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu_a)%</td>
<td>0.25</td>
<td>1.0</td>
<td>3.3</td>
<td>6.7</td>
<td>11.05</td>
</tr>
<tr>
<td>(\sigma_a)%</td>
<td>0.2</td>
<td>0.65</td>
<td>1.9</td>
<td>8.0</td>
<td>13.1</td>
</tr>
</tbody>
</table>

the symmetric similarity matrix ‘of’; elongation parameter for edge map ‘eem’; symmetric similarity matrix threshold ‘th’; error tolerance in eigensolver ‘et’).

1. Ground Truth Experiments

Ground truth experiments (on synthetic 1D, 2D and 3D images) are done to statistically measure the proposed algorithm performance. Fig. 30 shows a 1D image (signal) with 2 classes (1,-1) distorted by a White Gaussian Noise (WGN) with standard deviation \(\sigma_n = 1\). To illustrate that the standard MLE, which only uses the proposed LCG model threshold without spatial relations, does not give satisfactory restoration, Fig. 30 presents the MLE result as well as the proposed approach result. To get some statistical analysis, the original signal is distorted with different WGNs (with different \(\sigma_n\)). For each \(\sigma_n\) 50 noisy signals are generated. Average relative errors of proposed approach outputs in comparison to the original signal are computed for this data set. Means \(\mu_a\) and standard deviations \(\sigma_a\) of the relative errors are shown in Table (7). The results in this table illustrate the robustness to noise of the proposed algorithm.

2D Case: To assess the robustness of the proposed approach, it is tested on 600 synthetic 2D multimodal images and compared the results with ground truths. Examples of these images and proposed approach results are shown in Fig. 31. Relative errors \(\varepsilon\) of the results in comparison to ground truths and computation times \(\tau\) are also given. Fig. 32 shows EDISON outputs \((h_s=2,h_c=1.5,M=8000,)\), and NCUTS outputs (default parameters with \(nsg=3\) and 5) for the same images.

To get some statistics, 10 3-modal data sets are generated, each of which
FIGURE 30 – 1D image labeling: (a) The original signal is distorted by WGN (b) Distorted signal (c) MLE outputs (d) Proposed algorithm output.
consists of 30 images, with different Signal to Noise Ratios (SNR). The proposed algorithm is applied on the data sets and computed the average relative error for each data set. Fig. 33 shows the SNR and the corresponding average relative segmentation error. The error at SNR -5 dB is dramatically large and the proposed algorithm missed one of the objects due to the great amount of noise. The same scenario is repeated for some 5-modal images (see Fig. 31(c)), and the corresponding error versus SNR is plotted on Fig. 34. For a comparison purpose, the segmentation error of the ICM Algorithm, EDISON, and NCUTS techniques are also illustrated. It is worth mentioning that since the peaks in the image gray levels histogram disappeared at low SNR, the experiments at low SNR are presented to show the robustness of the framework given the number of classes.

Fig. 35 shows the results of the proposed algorithm and that of EDISON \((h_s=8, h_r=7.0, M=20)\) on a 4-modal image. In this case, one cannot compare with NCUTS since this algorithm was designed so as separate segments take different labels even if they share the same gray levels.

**3D Case:** To test the algorithm on 3D case, a bimodal volume phantom \((256x256x50)\) is generated. The relative error \(\varepsilon=0.53\%\) between the proposed approach segmentation and the ground truth confirms the high accuracy of the proposed segmentation framework in 3D cases. To emphasize the importance of segmenting the 3D object using a 3D graph (e.g., Fig. 36) instead of independently segmenting each 2D slice of the volume (e.g, see Fig. 37), the two corresponding segmentation results are illustrated in Fig. 38 (a) and (b), respectively.

2. **Real Images and Applications**

Figures 39, 40 and 41 show more segmentations of the proposed approach, EDISON, and NCUTS of real images. In each output, the object boundary (shown in red) is drawn, and the computation time is given. As shown in Fig. 40(d, e, and f) the gray levels inhomogeneities of the starfish and its background lead to some
FIGURE 31 – Segmentation results (a) The gray levels image with SNR=1 dB, (b) Proposed Approach Segmentation.(c) The gray levels image with SNR=5 dB, (d) Proposed Approach Segmentation. (Error shown in red)
(a) $\varepsilon = 3.86\%, \tau = 26$ sec.
(b) $\varepsilon = 1.48\%, \tau = 13$ sec.
(c) $\varepsilon = 2.24\%, \tau = 2$ sec.
(d) $\varepsilon = 1.44\%, \tau = 18$ sec.

FIGURE 32 – Segmentation results: for the 3-modal image (a) EDISON output, (b) NCUTS output. For the 5-modal image (c) EDISON output, (d) NCUTS output. (Error shown in red)
FIGURE 33 – The changes of the misclassification error with the SNR for 3-modal and 5-modal images

FIGURE 34 – The changes of the misclassification error with the SNR for 3-modal and 5-modal images
FIGURE 35 – 4-modal image segmentation results (b) $\varepsilon = 0.16\%$, $\tau = 5$ sec. (c) $\varepsilon = 0.71\%$, $\tau = 13$ sec. (errors shown in red)

FIGURE 36 – Example of a graph that used in volume labeling. Note: Terminals should be connected to all voxels but for illustration purposes, this was not done.

FIGURE 37 – Slices from the synthetic volume.
errors. However, proposed algorithm’s result is more accurate than the others for this specific case (Fig. 40(d)). This error can be overcome by using the high order cliques model as shown in the next chapter.

a. **Lung Segmentation**  Medical images are good examples of multimodal images. In such a case, errors are evaluated with respect to ground truths produced by a radiologist. To assess the performance of the proposed framework on practical problems, it is applied on lung segmentation problem [67]. Due to the closeness of the gray levels between the abnormal tissues in the lung and the chest tissues, interactive segmentation for the computed tomography (CT) lung images is improper. In order to measure the accuracy of the proposed approach on medical data, a geometric phantom is created with the same gray levels distribution in regions as in the lung CT images at hand using the inverse mapping approach [1]. The error 0.26% between proposed algorithm results and ground truth confirms the high accuracy of the proposed segmentation framework. For comparison, Fig. 42 shows the binary results obtained with the proposed technique, Iterative Threshold (IT) [68] approach, and ICM [7] Algorithm 1. Also the proposed
FIGURE 39 – Real image segmentation results of (a) Proposed Algorithm (b) EDISON ($h_s=15, h_c=9, M=5000$.) (c) NCUT ($nsg=2$, and the default parameters). (Courtesy of Shi and Malik [6])

FIGURE 40 – Starfish segmentation results of (a) Proposed Algorithm, (b) EDISON ($h_s=2, h_c=10, M=500$.), (c) NCUT ($nsg=2$, of=0.04), (d) Proposed Algorithm (e) EDISON ($h_s=7, h_c=6.6, M=5000$.) (f) NCUT ($nsg=2$, of=0.04, $th=0.03, eem=5, et=0.1$)
FIGURE 41 – Kidney segmentation results of (a) Proposed Algorithm $\varepsilon=1.7\%$ (b) EDISON ($h_s=5$, $h_c=6.1$, $M=9000$) $\varepsilon=1.1\%$ (c) NCUT ($nsg=2$, $of=0.01$, $th=0.04$, $et=0.01$) $\varepsilon=1.1\%$ (d) Proposed Algorithm $\varepsilon=0.1\%$ (e) EDISON ($h_s=7$, $h_c=6.6$, $M=9000$) $\varepsilon=0.01\%$ (f) NCUT ($nsg=2$, and the default parameters) $\varepsilon=0.6\%$
FIGURE 42 – Ling Phantom Segmentation: (a) Phantom gray scale image, (b) the proposed algorithm output, (c) ICM output, (d) IT output. (The misclassified pixels are shown in red color)

ICM, and IT approaches were run on 60 axial human chest slices obtained by spiral-scan low-dose computer tomography (LDCT), the 8-mm-thick LDCT slices were reconstructed every 4 mm with the scanning pitch of 1.5 mm. The results for three of them are shown in Fig. 43. In this experiment, the errors are evaluated with respect to the ground truth produced by an expert (a radiologist). The percentage error of the misclassified pixels is shown for each approach.

The proposed algorithm is extended to segment the whole lung volume simultaneously. This helps to overcome the large gray levels inhomogeneities in lung data. The proposed approach is applied on seven human chest CT scans. Some of these results are presented on Fig. 44. The main problem in the segmentation of ICM and IT is that the misclassified voxels include abnormal lung tissues (lung cancer), bronchi and bronchioles as shown in Fig. 45. These tissues are important if lung segmentation is a pre-step in a detection of lung nodules system.
FIGURE 43 – Segmentation Results of: proposed algorithm (1st column), ICM (2nd column), and IT (last column). (The misclassified pixels are shown in red color)
The motivation behind the proposed segmentation approach is to exclude such errors as far as possible. As expected, all misclassified pixels in the results of the proposed algorithm are located at the boundary. The statistical analysis is shown in Table (8). For comparison, the statistical analysis of ICM technique results, and the IT approach results are also shown. The unpaired t-test is used to show that the differences in the mean errors between the proposed segmentation, and (ICM/and IT) are statistically significant (the two-tailed value $P$ is less than 0.0006).

### E. Conclusions

In this chapter, a novel approach [69, 70] was presented for automatic multimodal gray scale image labeling using the graph cuts algorithm. A joint MGRF model was used to describe the input image and its desired map with more accurate model identification. The number of classes in the given image was determined using a new technique [71] based on maximizing a new joint likelihood function. The image gray levels distribution was precisely approximated by LCG distributions with positive and negative components. Therefore, no user interac-
FIGURE 44 – Proposed algorithm’s results. Left: segmented lung volumes (Error are shown in green). Right: samples from volume’s segmented slices (Error are shown in red). Segmentation errors (a) 2.08%, (b) 2.21%, (c) 2.17%, and (d) 1.95%
FIGURE 45 – Examples of segmented lung slices that have nodules (bounded by yellow circle). Left: IT and middle: ICM approaches misclassified these parts as chest tissues (error is shown in red). However, right: The proposed algorithm correctly classified them as a lung.

tion was needed; the image was initially segmented using this LCG model. Finally, an energy function using the previous models was formulated, and was globally minimized using graph cuts. Experimental results of synthetic and real gray scale multimodal images clarified that without optimizing any tuning parameters, the proposed approach was fast, robust to noise, and gave accurate results compared to the state-of-the-art algorithms (e.g., [6, 66]). Moreover, the proposed approach was easily extended to segment 3D volumes.
CHAPTER V

OPTIMIZING BINARY MRFs WITH HIGHER ORDER CLIQUES

Due to the explosion of efficient and successful pairwise MRFs solvers in computer vision, previous chapters focus on the pairwise MRF model. However, a question is still raised: does any link exist between the pairwise and higher order MRFs such that the like solutions can be applied to the latter models? This chapter explores such a link for binary MRFs that allows one to represent Gibbs energy of signal interaction with a polynomial function. First a new algorithm that converts higher order energy that represents high order MRFs to a polynomial function is presented. Then energy minimization tools for the pairwise MRF models can be easily applied to the higher order counterparts. The proposed framework demonstrates very promising experimental results of image segmentation and can be used to solve other computer vision problems.

A. Introduction

Recently, as explained in Sec. II.D, discrete optimizers (e.g., graph cuts, BP, and TRW) became essential tools in the computer vision field. These tools are used to solve many computer vision problems. Where, the framework of such problems is justified in terms of maximum a posteriori configurations in a MRF, and the MAP-MRF problem is formulated as a minimization of an energy function. However, this chapter focuses only on binary MRFs that play an important role in computer vision since Boykov et al. [44] proposed an approximate graph-cut algorithm for energy minimization with iterative expansion moves. As explained in Algorithm 2 Sec.II.E.3, this algorithm reduces the problem with multivalued variables to a sequence of subproblems with binary variables.
Most of the energy-based computer vision frameworks represent the MRF energy on an image lattice in terms of unary and pairwise clique potentials. However, this representation is insufficient for modeling rich statistics of natural scenes [34]. The latter require higher order clique potentials being capable to describe complex interactions between variables. Adding potentials for the higher order cliques could improve the image model [72, 73]. However, optimization algorithms of these models have too high time complexity to be practicable. For example, a conventional approximate energy minimization framework with belief propagation (BP) is too computationally expensive for MRFs with higher order cliques, and Lan et al. [34] proposed approximations to make BP practical in these cases. However, the results are competitive with only simple local optimization based on gradient descent technique. Recently, Kohli et al. [74] proposed a generalized $P^\alpha$ family of clique potentials for the Potts MRF model and showed that optimal graph-cut moves for the family have polynomial time complexity. However, just as in the standard graph-cut approaches based on the $\alpha$-expansion or $\alpha\beta$-swap, the energy terms for this family have to be submodular.

Instead of developing efficient energy minimization techniques for higher order MRFs, this work chooses an alternative strategy of reusing well established approaches that have been successful for the pairwise models and proposes an efficient transformation of an energy function for a higher order MRF into a quadratic function. First, the potential energy for higher order cliques is converted into a polynomial form, an algebraic proof explaining when this form can be graph representable is introduced, the graph reconstruction for such an energy is explicitly shown. Then the higher-order polynomial is reduced to a specific quadratic one. The latter may have submodular and/or nonsubmodular terms, and a few approaches have been proposed to minimize such functions. For instance, Rother et al. [75] truncate nonsubmodular terms in order to obtain an approximate submodular function to be minimized. This truncation leads to a reasonable solution when the number of the nonsubmodular terms is small. As discussed in Sec.II.F.1,
recently Rother et al. [9] proposed an efficient optimization algorithm for nonsubmodular binary MRFs, called the extended roof duality. However, it is limited to only quadratic energy functions. The proposed work expands notably the class of the nonsubmodular MRFs that can be minimized using this algorithm. In this chapter, extended roof duality is used to minimize the proposed quadratic version of the higher order energy. To illustrate potentialities of the higher order MRFs in modeling complex scenes, the performance of the proposed approach has been assessed experimentally in application to image segmentation. The obtained results confirm that the proposed optimized MRF framework can be efficiently used in practice.

B. Preliminaries

Recall, the goal image labeling $f$ in the MAP approach is a realization of a Markov-Gibbs random field (MGRF) $F$ defined over an arithmetic 2D lattice $\mathcal{P} = \{1, 2, \cdots, n\}$ with a neighborhood system $\mathcal{N}$. Energy functions for an MGRF with only unary and pairwise cliques can be written in the following form:

$$E(f) = \sum_{p \in \mathcal{P}} D(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V(f_p, f_q).$$  \hfill (54)

The unary terms $D(.)$ encode the data penalty function, and the pairwise terms $V(.,.)$ are interaction potentials. For simplicity, in this chapter both unary and pairwise terms will be represented as function $V(.)$. So the energy function has the following form:

$$E(f) = \sum_{p \in \mathcal{P}} V(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V(f_p, f_q).$$  \hfill (55)

The energy minimum $E(f^*) = \min_f E(f)$ corresponds to the MAP labeling $f^*$. For a binary MGRF, the set of labels consists of two values, $\mathcal{L} = \{0, 1\}$, each variable $f_p$ is a binary variable, and the energy function Eq. (55) can be written in a quadratic polynomial form:

$$E(f) = a_0 + \sum_{p \in \mathcal{P}} a_p f_p + \sum_{\{p,q\} \in \mathcal{N}} a_{pq} f_p f_q,$$  \hfill (56)
where, \( a_0, a_p \) and \( a_{pq} \) are real numbers depending on \( V(0), V(1), \ldots, V(1, 1) \) in a straightforward way.

Generally, let \( \mathcal{L}^n = \{(f_1, f_2, \ldots, f_n) | \ f_p \in \mathcal{L} \ \forall \ p = 1, \ldots, n\} \), and let \( E^k(f) = E^k(f_1, f_2, \ldots, f_n) \) be a real valued polynomial function of \( n \) bivalent variables and real coefficients and defining a Gibbs energy with higher order potentials (in contrast to the above quadratic function \( E \)). Such function \( E^k(f) \) is called a pseudo-Boolean function [76] and can be uniquely represented as a multi-linear polynomial [54] as follows:

\[
E^k(f) = \sum_{S \subseteq P} a_S \prod_{p \in S} f_p, \tag{57}
\]

where \( a_S \) are non-zero real numbers, and the product over the empty set is 1 by definition.

C. Polynomial Forms of Clique Potentials

To be transformed into a quadratic energy, the higher order energy function should be represented in a multi-linear polynomial form Eq. (57). This section considers how the clique potentials can be represented in a polynomial form. An unary term has an obvious polynomial form:

\[
V_{f_p} = V(f_p) = (V_1 - V_0)f_p + V_0, \tag{58}
\]

where \( V_1 \) and \( V_0 \), are the potential values for the labels 1 and 0 for the variable \( f_p \in \mathcal{L} \).

1. Cliques Of Size Two

Let \( f_p, f_q \in \mathcal{L} \), and let \( c_0, c_1, c_2 \) and \( c_3 \) be real coefficients. A clique of size two has a potential function \( V(f_p, f_q) \) that can be generally represented as follows:

\[
V_{f_pf_q} = V(f_p, f_q) = (c_0 f_p + c_1)(c_2 f_q + c_3)
\]

\[
= c_0c_2 f_pf_q + c_1c_2 f_q + c_0c_3 f_p + c_1c_3.
\]
So, it is easy to show that

\[ V_{00} = c_1c_3, \quad V_{01} = c_1c_2 + V_{00}, \quad V_{10} = c_0c_3 + V_{00}, \text{ and} \]

\[ c_0c_2 = V_{11} - (c_1c_2 + c_0c_3 + c_1c_3) = V_{11} + V_{00} - V_{01} - V_{10}. \]

This implies

\[
V_{f_pf_q} = (V_{11} + V_{00} - V_{01} - V_{10})f_pf_q + (V_{01} - V_{00})f_q \\
+ (V_{10} - V_{00})f_p + V_{00}. 
\] (59)

Indeed, representing pairwise potential in the polynomial form Eq. (59) has many advantages. It implies an algebraic proof of the Kolmogorov–Zabih’s submodularity condition [45] similar to the combinatorial optimization theorem (its version and proof are given in [77]):

**Theorem 2.** [Freedman–Drineas; 2005] The quadratic pseudo-Boolean function Eq. (56) can be minimized via graph cut techniques if and only if \( a_{pq} \leq 0 \) \( \forall \{p, q\} \in N \).

According to this theorem, quadratic polynomial function Eq. (59) can be minimized via graph cuts if and only if

\[ V_{11} + V_{00} - V_{01} - V_{10} \leq 0. \] (60)

This follows the definition of Kolmogorov and Zabih [45]:

**Definition 5.** A function of one binary variable is always submodular. A function \( V(f_p, f_q) \) from the family \( F^2 \) is submodular if and only if: \( V_{11} + V_{00} \leq V_{01} + V_{10} \).

The representation in Eq. (59) explicitly shows edges in the graph related to the quadratic term. This construction is similar to what has been introduced in [45] but here each term in Eq. (59), excluding the constant, directly represents a part of the graph.

2. Cliques Of Size \( k \)

A clique of size \( k \) has a potential function \( V(f_p, f_q, \cdots, f_k) \), where \( k \leq n \). Just as before, coefficients of the polynomial form can be estimated using Algorithm 6.
It is worth mentioning that Freedman and Drineas [77] proposed a mathematical formula that can be used to compute these coefficients. However, it is complicated, and the proposed algorithm is much easier for implementation.

**Algorithm 6 Generating Coefficients of Size k Clique’s Energy Algorithm**

1. \( Coef(S, Z, W) \)
2. \( \mathcal{H} = \{1, 2, \ldots, k\} \)
3. if \((S = \emptyset)\) then
   4. Return \(V(f_{\mathcal{H}} = 0)\)
5. end if
6. if \(|S| = 1\) then
   7. if \((Z = \emptyset)\) then
      8. Return \(V(f_S = 1, f_{\mathcal{H}-S} = 0) - V(0)\)
   9. else
      10. Return \(V(f_S = 1, f_Z = W, f_{\mathcal{H}-(S+Z)} = 0) - V(f_Z = W, f_{\mathcal{H}-Z} = 0)\)
   11. end if
12. else
   13. Select \(p \in S\). Then: \(S = S - \{p\}, Z = Z + \{p\}, W_1 = W + \{1\}, \) and \(W_0 = W + \{0\}\)
14. \( Coef(S, Z, W_1) - Coef(S, Z, W_0) \)
15. end if

As an example, Algorithm 6 is used to estimate the polynomial coefficients...
of the potential $V_{pqr} = V(f_p, f_q, f_r); f_p, f_q, f_r \in L$, for a clique of size 3. Then

$$V_{pqr} = \text{Coef}(\{p, q, r\}, \phi, \phi)f_p, f_q, f_r$$

$$+ \text{Coef}(\{p, q\}, \phi, \phi)f_p, f_q$$

$$+ \text{Coef}(\{p, r\}, \phi, \phi)f_p, f_r$$

$$+ \text{Coef}(\{q, r\}, \phi, \phi)f_q, f_r$$

$$+ \text{Coef}(\{p\}, \phi, \phi)f_p + \text{Coef}(\{q\}, \phi, \phi)f_q$$

$$+ \text{Coef}(\{r\}, \phi, \phi)f_r + \text{Coef}(\phi, \phi, \phi). \quad (61)$$

Where the coefficients are computed using Algorithm 6 such as:

$$\text{Coef}(\phi, \phi, \phi) = V_{000}.$$  

$$\text{Coef}(\{p\}, \phi, \phi) = V_{100} - V_{000}.$$  

$$\text{Coef}(\{p, q\}, \phi, \phi) \begin{cases} +\text{Coef}(\{q\}, \{p\}, \{1\}) = V_{110} - V_{100} \\ -\text{Coef}(\{q\}, \{p\}, \{0\}) = V_{000} - V_{010} \end{cases}.$$  

$$\text{Coef}(\{p, q, r\}, \phi, \phi) \begin{cases} +\text{Coef}(\{q, r\}, \{p\}, \{1\}) \\ -\text{Coef}(\{q, r\}, \{p\}, \{0\}) \end{cases},$$  

where

$$+\text{Coef}(\{q, r\}, \{p\}, \{1\}) \begin{cases} +\text{Coef}(\{r\}, \{p, q\}, \{1, 1\}) = V_{111} - V_{110} \\ -\text{Coef}(\{r\}, \{p, q\}, \{1, 0\}) = V_{100} - V_{101} \end{cases}.$$  

$$-\text{Coef}(\{q, r\}, \{p\}, \{0\}) \begin{cases} -\text{Coef}(\{r\}, \{p, q\}, \{0, 1\}) = V_{010} - V_{011} \\ +\text{Coef}(\{r\}, \{p, q\}, \{0, 0\}) = V_{001} - V_{000} \end{cases}.$$
Finally, the potential is as follows:

\[
V_{pqr} = ((V_{111} + V_{100} - V_{110} - V_{101}) \\
- (V_{011} + V_{000} - V_{001} - V_{010})) f_p f_q f_r \\
+ (V_{011} + V_{000} - V_{001} - V_{010}) f_q f_r \\
+ (V_{101} + V_{000} - V_{001} - V_{100}) f_p f_r \\
+ (V_{110} + V_{000} - V_{100} - V_{010}) f_p f_q \\
+ (V_{010} - V_{000}) f_q + (V_{100} - V_{000}) f_p \\
+ (V_{001} - V_{000}) f_r + V_{000}.
\]  

(62)

As will be explained in Sec.V.D, the first (cubic) term will be reduced to be a quadratic term with the same coefficient. Thus, and according to Theorem (2), such energy can be minimized via graph cuts if and only if

\[
(V_{111} + V_{100} - V_{110} - V_{101}) - (V_{011} + V_{000} - V_{001} - V_{010}) \leq 0, \tag{63}
\]

\[
V_{011} + V_{000} - V_{001} - V_{010} \leq 0, \tag{64}
\]

\[
V_{101} + V_{000} - V_{001} - V_{100} \leq 0, \text{ and} \tag{65}
\]

\[
V_{110} + V_{000} - V_{100} - V_{010} \leq 0. \tag{66}
\]

These inequalities represent all the projections of the $V_{pqr}$ on 2 variables. This follows the following definition [45]:

**Definition 6.** A function from the family $F^k$ is submodular if and only if: all its projections on 2 variables are submodular.

Moreover, (and similar to the $F^2$ case), Eq. (62) explicitly shows how the cubic term can be represented in the graph construction. This polynomial representation can be easily expanded to the general case of $F^k$ when Algorithm 6 is used to compute the coefficients.
D. Energy Reduction - the Proposed Approach

The previous section showed so far how the Gibbs energy can be represented in the polynomial form Eq. (57) for any clique size. This section discusses minimization of such energies. Quite successful optimization techniques for graphical models have been proposed to minimize quadratic energies with submodular terms (e.g., [44]) and nonsubmodular terms (e.g., [9]). This section proposes how to exploit these algorithms for minimizing higher order energies by transforming the latter to the quadratic ones. This can be done by adding a number of dummy variables, each such variable substituting the product of the two initial variables as shown in the following algorithm. Different theoretical works [54,76] considered the problem of reducing an optimization of the general pseudo-Boolean function \( E^k(f) \) in polynomial time to the optimization of a quadratic pseudo-Boolean function. In contrast to [76], the proposed algorithm guarantees that the goal quadratic pseudo-Boolean function has the same minimum at the same variables as the initial general pseudo-Boolean function. Also, as distinct from what has been proposed in [54], both a detailed proof to verify the proposed algorithm and an efficient implementation of the latter on a related graph are proposed.

The quadratic pseudo-Boolean function \( E(.) \) is obtained from the pseudo-Boolean function \( E^k(.) \) as follows: replace the occurrence of \( f_p f_q \) in Eq. (57) by \( f_{n+1} \) and add the term \( B \cdot (f_p f_q + 3f_{n+1} - 2f_p f_{n+1} - 2f_q f_{n+1}) \) to \( E(.) \). This gives the following function \( E(f_1, f_2, \cdots, f_{n+1}) \)

\[
E(f) = B \cdot (f_p f_q + (3 - 2f_p - 2f_q)f_{n+1}) + \sum_{S^* \subseteq P} a_{S^*} \prod_{p \in S^*} f_p,
\]  

(67)

where

\[
S^* = \begin{cases} 
(S \setminus \{p, q\}) \cup \{n + 1\} & \{p, q\} \subseteq S \\
S & \{p, q\} \not\subseteq S
\end{cases}
\]
To compute the constant \( B \), Eq. (57) is rewritten first as follows:

\[
E^k(f) = a_0 + \sum_{S_1 \subseteq P} a_{S_1} \prod_{p \in S_1} f_p + \sum_{S_2 \subseteq P} a_{S_2}^+ \prod_{p \in S_2} f_p,
\]

where \( a_0 \) is the absolute term, \( a_{S_1}^{-} \)'s are the negative coefficients, and \( a_{S_2}^{+} \)'s are the positive coefficients. Then, let \( \mathfrak{A} = a_0 + \sum_{S_1 \subseteq P} a_{S_1}^{-} \) be the sum of all negative coefficients in Eq. (68) plus the absolute term. Note that

\[
\mathfrak{A} \leq \min_{f \in \mathcal{L}^n} E^k(f).
\]

Also, denote \( u \) a real number being greater than the minimal value of \( E^k(f) \) on \( \mathcal{L}^n \):

\[
u > \min_{f \in \mathcal{L}^n} E^k(f).
\]

Practically, \( u \) can be any number being greater than a particular value of \( E^k(f) \) on \( \mathcal{L}^n \). Finally, the chosen value \( B \) has to satisfy the relationship \( B \geq u - \mathfrak{A} \).

This replacement is repeated until one gets a quadratic pseudo-Boolean function. Algorithm 7 shows these steps in detail.

At each step, \( E(f_1, f_2, \cdots, f_{n+1}) \) must satisfy the following

**Lemma 1.** Let \( \mathcal{M}_{E^k} = \{ y \in \mathcal{L}^n | E^k(y) = \min_{f \in \mathcal{L}^n} E^k(f) \} \) be a set of all \( y \in \mathcal{L}^n \) such that \( E^k(y) \) is the global minimum of the function \( E^k \) on \( \mathcal{L}^n \). Then

1. \( E(f_1, f_2, \cdots, f_{n+1}) = E^k(f_1, f_2, \cdots, f_n) \),

2. \( (y_1, y_2, \cdots, y_{n+1}) \in \mathcal{M}_E \) iff \( (y_1, y_2, \cdots, y_n) \in \mathcal{M}_{E^k} \).

**Proof 1.** For \( x, y, z \in \mathcal{L} \), it is easy to notice that the function \( \Omega(x, y, z) = xy + 3z - 2xz - 2yz \) can be defined as follows:

\[
\Omega(x, y, z) = \begin{cases} 
0 & xy = z \\
3 & x = y = 0, z = 1 \\
1 & \text{otherwise}
\end{cases}
\]
Algorithm 7 Transform to Quadratic

Require: general pseudo-Boolean function $E^k(f)$ Eq. (57).

1: set $\mathcal{A} = a_0 + \sum_{S_1 \subseteq P} a_{S_1}$, set $u > \min_{f \in \mathcal{L}^n} E^k(f)$ (e.g., $u = E^k(0) + 1$), and set $B \geq u - \mathcal{A}$

2: while $(\exists S' \subseteq P$ and $|S'| > 2)$ do

   Select a pair $\{p, q\} \subseteq S'$ and update the coefficients

   $$a_{\{p,q\}} = a_{\{p,q\}} + B$$
   $$a_{\{p,n+1\}} = a_{\{q,n+1\}} = -2B$$
   $$a_{(n+1)} = 3B$$
   $$a_{(S-\{p,q\}) \cup (n+1)} = a_S, \text{ set } a_S = 0 \forall S \supseteq \{p, q\}$$

4: $n = n + 1$, update the function as in Eq. (67)

end while

Ensure: The quadratic pseudo-Boolean function $E(f)$

If $f_pf_q = f_{n+1}$, then

$$E(f) = B \cdot \Omega(f_p, f_q, f_{n+1}) + \sum_{S' \subseteq P} a_{S'} \prod_{p \in S'} f_p$$

$$= 0 + \sum_{S' \subseteq P} a_{S'} \prod_{p \in S'} f_p$$

$$= \sum_{S \subseteq P} a_S \prod_{p \in S} f_p = E^k(f).$$

More specifically, $E(f)$ has the same minimum value as $E^k(f)$ on $\mathcal{L}^n$. On the other hand, assume $y_p, y_q \neq y_{n+1}$, which implies that $\Omega(y_p, y_q, y_{n+1}) \geq 1$. Also, assume $(y_1, y_2, \cdots, y_{n+1}) \in \mathcal{M}_E$. Then

$$E(y) = B \cdot \Omega(y_p, y_q, y_{n+1}) + \sum_{S' \subseteq P} a_{S'} \prod_{p \in S'} y_p.$$
From equations (69,70) one can conclude that $B > 0$. So

$$E(y) \geq B + \sum_{S \subseteq P} a_S \prod_{p \in S^*} y_p \geq B + \mathbb{A}.$$ 

The choice of $B \geq u - \mathbb{A}$, guarantees that

$$E(y) \geq u.$$

This contradicts the assumption $(y_1, y_2, \ldots, y_{n+1}) \in \mathcal{M}_E$. This proves $(y_1, y_2, \ldots, y_{n+1}) \notin \mathcal{M}_E$ if $y_p y_q \neq y_{n+1}$ and the lemma follows.

By repeatedly applying the construction in Lemma (1), one gets the following theorem (different versions of this theorem can be found in [54, 76]):

**Theorem 3.** Given a general pseudo-Boolean function $E^k(f_1, f_2, \ldots, f_n)$, there exists at most a quadratic pseudo-Boolean function $E(f_1, f_2, \ldots, f_{n+m})$ where $m \geq 0$ such that

1. $(y_1, y_2, \ldots, y_{n+m}) \in \mathcal{M}_E \iff (y_1, y_2, \ldots, y_n) \in \mathcal{M}_{E^k}$

2. The size of the quadratic pseudo-Boolean function is polynomially bounded in size of $E^k$, and so the reduction algorithm will terminate at polynomial time.

**Proof 2.** Repeated application of the construction in the proof of Lemma (1) yields to the point 1 of the theorem.

To prove point 2: Define by $M_3$ the number of terms that have $|S| \geq 3$ (i.e., more than 2 variables, higher order terms) in the function $E^k(f_1, f_2, \ldots, f_n)$.³ In the loop of the Algorithm 7, notice the following:

³Note that a function $E^k$ of $n$ binary variables contains at most $2^n$ terms. This can be computed by summing numbers of terms that have 0 up to $n$ variables.

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$$

Also, it is easy to show that the function $E$ contains at most $2^n - \frac{n^2+n+2}{2}$ terms that have $|S| > 2$ (i.e., more than 2 variables).
• The term of size \( n \) (i.e., \(|S| = n\)) needs at most \( n - 2 \) iterations,

• Also, at each iteration in this loop, at least one of the terms which have \(|S| > 2\) will decrease in size.

Hence the algorithm must terminate in at most a number of iterations \( \ll M_3(n - 2) \). The less sign is presented because the average number of iterations for each term will be less than \( n - 2 \). Indeed, larger number of variables contained by each energy term indicates that these terms share several common variables, and so they will be reduced concurrently. As an example: a function with 10 variables contains at most 968 terms with \(|S| > 2\). Using Algorithm 7, it is reduced with a number of iterations equals to 68 \( \ll 968 \times 8 \). This proves the claim about complexity.

1. Efficient Implementation

The number of dummy variables in the generated quadratic pseudo-Boolean function depends on the selection of the pairs \( \{p, q\} \) in the loop of the Algorithm 7. Finding the optimal selection to minimize this number is an NP-hard problem [54]. Also, searching for this pair in other terms will be exhaustive. However, in most computer vision problems, one deals with images on an arithmetic 2D lattice \( \mathcal{P} \) with \( n \) pixels. The order of the Gibbs energy function to be minimized depends on the particular neighborhood system and the maximal clique size. The prior knowledge about the neighborhood system and the clique size can be used to minimize the number of dummy variables and to eliminate the search for the repeated pair in other terms. This process is demonstrated on the second order neighborhood system and the cliques of the size 3 (see Fig. 5), but it can be generalized for the higher orders. Figure 46 suggests that the second order neighborhood system contains four different cliques of the size 3. Thus, the cubic terms that correspond to the cliques of the size 3 can be converted, to quadratic terms as follows:

• At each pixel \((i, j)\) select the cubic term that corresponds to clique \( \gamma_8 \)
FIGURE 46 – Part of an image lattice for 2\textsuperscript{nd} order neighborhood system and cliques of size three

- Reduce this term and the cubic term of the clique $\gamma_6$ at pixel $(i-1, j-1)$ if possible, by eliminating variables $(i-1, j)$, and $(i, j-1)$

- For pixel $(i, j)$ select the cubic term that corresponds to the clique $\gamma_5$

- Reduce this term and the cubic term of the clique $\gamma_7$ at pixel $(i-1, j+1)$ if possible, by eliminating variables $(i-1, j)$, and $(i, j+1)$

After a single scan for the image, all the cubic terms will be converted to the quadratic terms, and every term will be visited only once. To illustrate the enhancement introduced by the proposed implementation, as an example: the linear search in a list runs in $O(n)$: $n$ is the number of elements. Hence, an image of size $R \times C$ has $4(R-1)(C-1)$ triple cliques in second order neighborhood system window. Each triple clique has 4 terms with $|S| > 1$ with total 9 elements as shown in Eq. (62). So applying Algorithm 7 directly without proposed implementation has an overhead $O(36(R-1)(C-1))$

Notice that this scenario is not unique. Many other scenarios can be chosen for image scanning and selection of pairs of higher order cliques to be reduced. However, in the efficient scenario every higher order term must be converted to a quadratic term after being visited only once.
E. Experimental Results

To illustrate the potential of higher order cliques in modelling complex objects and assess the performance of the proposed algorithm, image segmentation into two classes (object and background) is considered. As described in Sec.II.C the MAP estimate of $f$, given the input image, is equivalent to minimize an energy function of the form (57). Where, the set of labels is $\{0 \equiv \text{“BCK”}, 1 \equiv \text{“OBJ”}\}$ and each pixel’s label represents a variable in this energy. So, one has an $n$ binary variables energy function. The unary term in this energy function $V(f_p)$ is chosen to be:

$$V(f_p) = ||I_p - \hat{I}_p||^2,$$  \hspace{1cm} (71)

where, $I_p$ is the features vector at the pixel $p$, e.g. a 4D vector $I_p = (I_{Li}, I_{ai}, I_{bi}, I_{ti})$ [78], where the first three components are the pixel-wise color L*a*b* components and and $I_{ti}$ is a local texture descriptor [79]. Seeds selected from the input image can be used to estimate feature vectors for the object, $\hat{I}_1$, and background, $\hat{I}_0$. Using feature vectors $\hat{I}_1$ and $\hat{I}_0$, an initial binary map can be estimated. The pairwise and third order cliques’ potentials are analytically estimated from the initial map using the proposed methods described in Sec.III.B and Sec.III.B.2, respectively.

In all experiments, the second order neighborhood system is selected sizes from 1 to 3. By defining the cliques’ potentials (unary, pairwise, and third order), one identifies the target segmentation’s energy that needs to be minimized. After that, Algorithm 6 is used to compute the coefficients of the polynomial that represents the segmentation’s energy. Then Algorithm 7 generates a quadratic version of this polynomial. Finally, the extended roof duality optimization algorithm [9], discussed in Sec.II.F solves the quadratic pseudo-Boolean function. In the experiments that follow, images are segmented twice: first, with unary and pairwise cliques, and then with unary and third order cliques in the MGRF model. Of course, cliques of greater sizes can be more efficient for describing complex regions. The third order is used for illustration purposes only.
Fig. 47 shows the segmentation results of a starfish. As shown in the results, unlike pairwise interaction Fig. 47(a), the high order interaction Fig. 47(b) overcomes the intensity inhomogeneities of the starfish and its background. For more challenging situations, some parts are occluded from the starfish Figures 48 and 49. Also, the high order interaction (see b and d) successes to get the correct boundary of the starfish, however, pairwise interaction (see a and c) could not. The average execution time for this experiment: in the higher order case is 6 sec., in correspondence to 2 sec. in the pairwise case.

More segmentation results for different colored objects are shown in Figures 50, 51, and 52. These images are from the Berkeley Segmentation Dataset [80]. As shown in Fig. 50 (a,c), numbers refer to regions that contain inhomogeneities where the pairwise interaction fails. However, as expected, the high order interaction overcomes them (see b,d). More results are illustrated in Fig. 51. In Fig. 52 some artificial occlusions are made, by letting some object regions take the background color. The results illustrate that the high order interaction still can get the correct segmentations.
FIGURE 48—Starfish with occlusions segmentation results. (a,c): the pairwise cliques results, and (b,d): the higher order cliques results.
FIGURE 49 – Starfish with occlusions more segmentation results. (a,c): the pairwise cliques results, and (b,d): the higher order cliques results.
FIGURE 50—More segmentation results. (a,c): the pairwise cliques results (Numbers in images refer to the regions with inhomogeneities), and (b,d): the higher order cliques results.
FIGURE 51 – More segmentation results. (a,c,e): the pairwise cliques results (Numbers in images refer to the regions with inhomogeneities), and (b,d,f): the higher order cliques results.
FIGURE 52 – More segmentation results for partially occluded objects. (a,c): the pairwise cliques results (Numbers in images refer to the regions represent artificial occlusion), and (b,d): the higher order cliques results.
F. Conclusions

This chapter introduced an efficient link between the binary MGRF models with higher order and pairwise cliques. It proposed an algorithm [81] that can transform a general pseudo-Boolean function into a quadratic pseudo-Boolean function and provably guarantees the obtained quadratic function has the same minimum at the same variables as the initial higher order one. The algorithm was efficiently implemented for image-related graphical models. Thus, one can apply the well known pairwise MGRFs solvers to the higher order MGRFs. The MGRF parameters were analytically estimated. Experimental results showed the proposed framework notably improved image segmentation and therefore may be useful for solving many other computer vision problems.
CHAPTER VI
A NOVEL SHAPE REPRESENTATION AND APPLICATION FOR IMAGE SEGMENTATION

This chapter proposes a novel segmentation approach based on the graph cuts technique with shape constraints. The segmentation approach depends on both image appearance and shape information. Shape information is gathered from a set of training shapes. Then the shape variations are estimated using a new distance probabilistic model. This model approximates the marginal densities of the object and its background in the variability region using a Poisson distribution refined by positive and negative Gaussian components. To segment an object in the given image, first it is aligned with the training images so one can use the distance probabilistic model. As discussed in Sec.IV.B, the object gray level is approximated with a linear combination of Gaussian distributions with positive and negative components. The spatial interaction between the neighboring pixels is identified using the new analytical approach introduced in Sec.III.B. Finally, a new energy function is formulated using both image appearance models and shape constraints. This function is globally minimized using $s/t$ graph cuts to get the optimal segmentation. Experiments show that the proposed technique gives promising results compared to others without shape constraints.

A. Introduction

Segmentation is a fundamental problem in image processing. There are many simple techniques, such as region growing or thresholding, for image segmentation. Although these techniques are widely known due to their simplicity and speed, no accurate segmentation can be achieved using these techniques
because these techniques depend only on the marginal probability distributions, and in most cases signal ranges for different objects overlap. To overcome this problem, many methods try to exploit the spatial interaction between segments as well as the regional properties of segments. Also parametric deformable models (e.g. [82]) and geometrical deformable models (level sets e.g. [83]) are powerful methods and have been used widely for the segmentation problems. However, all these methods tend to fail in the case of noise, gray level inhomogeneities, diffused boundaries or occluded shapes, and they don’t take advantage of the a priori models. Therefore segmentation algorithms can not depend only on image information but also have to exploit the prior knowledge of shapes and other properties of the structures to be segmented.

Leventon et al. [35] combine the shape and deformable model by attracting the level set function to the likely shapes from a training set specified by principal component analysis (PCA). Huang et. al. [37], combine registration with segmentation in an energy minimization problem. The evolving curve is registered iteratively with a shape model using the level sets. They minimized a certain function to estimate the transformation parameters. Unfortunately, this approach may stuck in a local minimum and its coefficients still exist to be tuned. In [38], shapes are represented with a linear combination of 2D distance maps where the weight estimates maximize the distance between the mean gray values inside and outside the shape. In [39] a shape prior and its variance obtained from training data are used to define a Gaussian distribution, which is then used in the external energy component of a level sets framework. To make the shape guides the segmentation process, Chen et al. [36] defined an energy functional which basically minimizes an Euclidean distance between a given point and its shape prior.

In this chapter, a new segmentation approach is proposed. This approach uses graph cuts to combine region and boundary properties of segments as well as shape constraints. From a set of aligned images an image consisting of three segments (common object, common background, and shape variability region) is
generated. The shape variations are modelled using a new distance probabilistic model. This distance model approximates the distance marginal densities of the object and its background inside the variability region using a Poisson distribution refined by positive and negative Gaussian components. For each given image, to use the distance probabilistic model, the given image is aligned with the training images. Then its gray level is approximated using an LCG model with positive and negative components. Finally, a new energy function is globally minimized using $s/t$ graph cuts to get the optimal segmentation. This function is formulated such that it combines region and boundary properties, and the shape information.

### B. Proposed Segmentation Framework

In this chapter, the goal is to find the optimal segmentation, best labelling $f$, by minimizing a new energy function which combines region and boundary properties of segments as well as shape constraints. Image appearance models are discussed in sections III.B and IV.B. In this section the shape model is explained.

1. **Shape Model Construction**

   A shape model of an object is created from a training set of images for that object. Fig. 53 illustrates the steps that used to create a human kidney shape model from human kidney Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) slices. Fig. 53(a) shows a sample of the DCE-MRI kidney slices. First, the kidneys are manually segmented (by a radiologist), as shown in Fig. 53(b). Then the segmented kidneys are aligned using 2D rigid registration [84], see Fig. 53(c). The aligned images are converted to binary images, as shown in Fig. 53(d). Finally, a labelled image “shape image” $\mathcal{P}_s = \mathcal{K} \cup \mathcal{B} \cup \mathcal{X}$ is generated as shown in Fig. 54(a). The white color represents $\mathcal{K}$ (kidney), black represents $\mathcal{R}$ (background), and gray is the variability region $\mathcal{X}$. To model the shape variations, variability region $\mathcal{X}$, a distance probabilistic model is used. The distance probabilistic model
FIGURE 53 – Samples of kidney training data images: (a) original, (b) Segmented, (c) Aligned, (d) Binary

describes the object (and background) in the variability region as a function of the normal distance \( d_p \) from a pixel \( p \in \mathcal{X} \) to the kidney/variability contour \( C_{KX} \).

\[
d_p = \min_{c \in C_{KX}} \| p - c \|. \tag{72}
\]

Each set of pixels located at equal distance \( d_p \) from \( C_{KX} \) constitutes an iso-contour \( C_{dp} \) for \( C_{KX} \) as shown in Fig. 54(b) (To clarify the iso-contours, the variability region is enlarged without relative scale to object). The kidney distance histogram is estimated as follows. The histogram entity at distance \( d_p \) is defined as

\[
h_{dp} = \sum_{i=1}^{M_t} \sum_{p \in C_{dp}} \delta(p \in \mathcal{K}_i), \tag{73}
\]
where the indicator function \( \delta(A) \) equals 1 when the condition \( A \) is true, and zero otherwise, \( M_i \) is the number of training images, and \( K_i \) is the kidney region in the \( i^{th} \) training image. The distance \( d_p \) is changed until the whole distance domain available in the variability region is covered. Then the histogram is multiplied by kidney prior value which is defined as follows:

\[
\pi_K = \frac{1}{M_t \cdot |\mathcal{X}|} \sum_{i=1}^{M_t} \sum_{p \in K} \delta(p \in K_i).
\]

(74)

Since each iso-contour \( C_{d_p} \) is a normally propagated wave from \( C_{KX} \). A reasonable assumption is that the probability of an iso-contour \( C_{d_p} \) to be object decays exponentially as \( d_p \) increased. To estimate the marginal density of the kidney, for a discrete index \( d_p \) a Poisson distribution to the object the distance histogram can be fitted. The set of pixels belong to the iso-contour \( C_{d_p} \) will obey a Poisson process. The same scenario is repeated to get the marginal density of the background. The kidney and background distance empirical densities and the estimated Poisson distributions are shown in Fig. 55 (a) and (b), respectively.
2. Distance Probabilistic Model

The distance marginal density of each class \( P(d_p \mid f_p) \) is estimated as follows. Since each class \( f_p \) (object or background) does not follow a perfect Poisson distribution, there will be a deviation between the estimated and the empirical densities. This deviation is modelled by a linear combination of Gaussians with positive and negative components. So the distance marginal density of each class consists of a Poisson distribution and a LCG with \( C_{f_p}^+ \) positive and \( C_{f_p}^- \) negative components as follows:

\[
P(d_p \mid f_p) = \vartheta(d_p \mid \xi_{f_p}) + \sum_{r=1}^{C_{f_p}^+} w_{f_p,r}^+ \phi(d_p \mid \theta_{f_p,r}^+) - \sum_{l=1}^{C_{f_p}^-} w_{f_p,l}^- \phi(d_p \mid \theta_{f_p,l}^-),
\]

where \( \vartheta(d_p \mid \xi_{f_p}) \) is a Poisson density with rate \( \xi \). The Poisson distribution parameter is estimated using the maximum likelihood estimator. As shown in Fig. 55: (c) and (d) illustrate the probabilistic models components for object and background respectively. The empirical and the final estimated densities are shown in Fig. 55 (e) for the kidney and (f) for the background.

3. Graph Cuts-based Optimal Segmentation

Define by \( d \) the set of distances of pixels in the variability region (shape information). Due to the independence of \( I \) and \( d \), a probability model of the shape constraints, input image and its desired map is given by a conditional distribution:

\[
P(f \mid I, d) = P(f)P(I \mid f)P(d \mid f).
\]

Similar to what was explained in Sec.II.C.1, MAP estimate of \( f \) is equivalent to minimizing the following function:

\[
E(f) = \sum_{(p,q) \in \mathcal{N}} V(f_p, f_q) - \sum_{p \in \mathcal{P}} \log(P(I_p \mid f_p)) - \sum_{p \in \mathcal{P}} \log(P(d_p \mid f_p)),
\]

where \( V(f_p, f_q) \) represents the penalty for the discontinuity between pixels \( p \) and \( q \). This model is discussed in Sec.III.B. The second term measures how much
FIGURE 55 – (a,b) Empirical densities and the estimated Poisson distributions, (c,d) Components of distance probabilistic models, (e,f) Final estimated densities
assigning a label $f_p$ to pixel $p$ disagrees with the pixel intensity $I_p$. This model is discussed in Sec.IV.B. The last term measures how much assigning a label $f_p$ to pixel $p$ disagrees with the shape information, this is explained in the previous section.

To segment an object, a graph (e.g. Fig. 6) is constructed and the weight of each edge is defined as shown in Table (9). Then the optimal segmentation boundary between the object and its background is obtained by finding the minimum cost cut on this graph. The minimum cost cut is computed exactly in polynomial time for two terminal graph cuts with positive edges weights via $s/t$ Min-Cut/Max-Flow algorithm [53].

### C. Experiments

The proposed segmentation framework is tested on a data set of DCE-MRI of human kidney. To segment a kidney slice, one can follow the following scenario. The given image is aligned with the aligned training images. The gray level marginal densities of the kidney and its background are approximated using the proposed LCG model with positive and negative components. Fig. 56(a) shows the original image, (b) shows the aligned image, (c) illustrates the empirical densities

<table>
<thead>
<tr>
<th>Edge</th>
<th>Weight</th>
<th>for</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>${p, q}$</td>
<td>$V(f_p, f_q)$</td>
<td>${p, q} \in N$</td>
<td></td>
</tr>
<tr>
<td>${s, p}$</td>
<td>$-\log[P(I_p</td>
<td>“1”) * P(d_p</td>
<td>“1”)]$</td>
</tr>
<tr>
<td></td>
<td>$\infty$</td>
<td>$p \in K$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$p \in R$</td>
<td></td>
</tr>
<tr>
<td>${p, t}$</td>
<td>$-\log[P(I_p</td>
<td>“0”) * P(d_p</td>
<td>“0”)]$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$p \in K$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\infty$</td>
<td>$p \in R$</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 56 – Gray level probabilistic model for the given image (a) Original image (b) aligned image (c) Initial density estimation (d) LCG components (e) Final density estimation (f) Marginal densities. Segmented Kidney (g) Results of gray level threshold 102.6% (h) Results of Graph cuts without shape constraints 41.9% (i) Proposed approach results 2.5%
as well as the initial estimated density using dominant modes in the LCG model, (d) illustrates the LCG components, (e) shows the closeness of the final gray level estimated density and the empirical one. Finally, (f) shows the marginal gray level densities of the object and back ground with the best threshold. To illustrate the closeness of the gray level between the kidney and its background, (g) shows the segmentation using gray level threshold=72. To emphasize the accuracy of the proposed approach, (h) shows the segmentation using the graph cuts technique without using the shape constraints (all the t-links weights will be $-\log P(I_p \mid f_p)$), and (i) shows the results of the proposed approach.

Samples of the segmentation results for different subjects are shown in Figures 57 and 58, (a) illustrates the input images, (b) shows the results of graph cuts technique without shape constraints, and the results of the proposed approach are shown in (c).

**Evaluation:** to evaluate the results the percentage segmentation error from the ground truth (manual segmentation produced by an expert) is calculated as follows:

$$\text{error}\% = \frac{100 \times \text{Number of misclassified pixels}}{\text{Number of Kidney pixels}}$$

For each given image, the binary segmentation is shown as well as the percentage segmentation error. The misclassified pixels are shown in red color.

The statistical analysis of 33 slices, which are different than the training data set is shown in Table (10). The unpaired $t$-test is used to show that the differences in the mean errors between the proposed segmentation, and graph cut without shape prior and the best threshold segmentation are statistically significant (the two-tailed value $P$ is less than 0.0001).

**D. Validation**

Due to the hand shaking errors, it is difficult to get accurate ground truth from manual segmentation. Thus to evaluate the proposed algorithm performance,
FIGURE 57 – More segmentation results (a) Original images (b) Results of Graph cuts without shape constraints (c) Proposed approach results
FIGURE 58 – More segmentation results (a) Original images (b) Results of Graph cuts without shape constraints (c) Proposed approach results
a phantom shown in Fig. 59(a) is created with topology similar to the human kidney. Furthermore, the phantom mimics pyramids that exist in any kidney. The kidney, the pyramids and the background signals for the phantom are generated according to the distributions shown in Fig. 56(f) using the inverse mapping method [1]. Fig. 59(b,c) show that the proposed approach is almost 26 times more accurate than the graph cuts technique without shape constraints.

### E. Conclusions

In this chapter, a new segmentation approach [85] that used graph cuts to combine region and boundary properties of segments as well as shape constraints was proposed. Shape variations were estimated using a new distance probabilistic model. To get the optimal segmentation, a new energy function was formulated using the image appearance models discussed in previous chapters and shape constraints. Then, this function was globally minimized using s/t graph cuts. Experimental results showed that the shape constraints overcame the gray level inhomogeneities problem and precisely guided the graph cuts to accurate segmentations (with mean error 5.4% and standard deviation 1.6%) compared to graph cuts with-
FIGURE 59 – Kidney Phantom (a) The phantom (b) Results of Graph cuts without shape constraints (c) Proposed approach results out shape constraints (mean error 62.7% and standard deviation 27.5%).
CHAPTER VII
STEREO MATCHING-BASED HUMAN FACES RECONSTRUCTION

The image labeling can be used as a formulation for diverse computer vision and image processing applications. In addition to its applicability to image segmentation and image restoration, the image labeling formulation can be utilized to solve one of the most fundamental problems in computer vision, the stereo matching problem.

A. Introduction

Stereo matching is an essential problem in computer vision and it has been studied in a huge number of works (e.g., [8, 44, 49, 51]). Stereo matching is a special case from a general problem called the image matching problem. The latter can be formulated as a labeling problem as follows. Let $I$ and $\tilde{I}$ denote two observed images. Typically, one of these images is chosen to be the reference image $I$ with set of pixel $\mathcal{P}$. The labeling algorithm assigns each pixel $p \in \mathcal{P}$ a label (displacement) $f_p$, such that $I_p$ and $\tilde{I}_{p+f_p}$ are intensities of correspondence pixels. Similar to the formulation in Sec. II.C, image pixels represent the sites. However, instead of the gray levels, displacements $(\partial_x, \partial_y)$ in the image spatial domain are used as the labels. The desired displacement field is the mapping $f : \mathcal{P} \rightarrow \mathcal{L}$, where $\mathcal{L}$ is the set of labels $\{(\partial_x^1, \partial_y^1), \ldots, (\partial_x^K, \partial_y^K)\}$, where $K$ is the number of labels.

Similar to what has been discussed in Sec. II.C, the framework for this problem can be the search for MAP configurations in a MRF model. The MAP problem is formulated as minimizing an interaction energy for the model. Two main assumptions are typically used in this problem: (1) the intensity of each pixel $I_p$ is similar to the intensity of the corresponding pixel in the other image $\tilde{I}_{p+f_p}$ and (2)
the displacement field \( f \) should be smooth. Therefore, the desired displacement field \( f \) is equivalent to minimizing the same energy function in Eq. (10) that incorporated these assumptions. It is rewritten here:

\[
E(f) = \sum_{(p,q) \in N} V(f_p, f_q) + \sum_{p \in P} D(f_p).
\] (79)

A proper method for computing the data penalty term \( D(f_p) \) is introduced in [8, 51]. This method uses the Birchfield and Tomasi approach [86], which handles sampling artifacts with slight variation. So \( D(f_p) \) can be computed as follows [8].

\[
\varrho_1(p, f_p) = \min_{f_p - \frac{1}{2} \leq f_p + \frac{1}{2}} |I_p - \tilde{I}_{p+t}|
\]

\[
\varrho_2(p, f_p) = \min_{p - \frac{1}{2} \leq q \leq p + \frac{1}{2}} |I_q - \tilde{I}_{p+f_p}|
\]

\[
\varrho(p, f_p) = \min(\varrho_1, \varrho_2)
\]

\[
D(f_p) = \varrho(p, f_p)^2.
\] (80)

In the previous formulation of image segmentation, the smoothness term is chosen to be piecewise constant prior. In contrast, in the matching problem the smoothness term is chosen to be piecewise smooth prior to allow smooth variations in the displacement field.

\[
V(f_p, f_q) = \min(|f_p - f_q|, \mathcal{M}),
\] (81)

where \( \mathcal{M} \) is a constant. Note, if \( \mathcal{M} = 1 \) this leads to piecewise constant prior. If \( \mathcal{M} > 1 \) this leads to piecewise smooth prior.

Fig. 60 illustrates simple examples for the image matching problem. In each row in Fig. 60, the left column illustrates the object in the reference image and the relative position of its candidate in the other image. The latter is illustrated by the border of the object. After minimizing Eq. (79) using Algorithm 2, in Sec.II.E.3, and applying the generated displacement fields, the objects are matched as shown in the right column of Fig. 60.
FIGURE 60–Image matching results. Left relative positions before matching. Right matching results.
B. Stereo Matching

In the classical stereo matching problem, the setup consists of two cameras observing a static scene. The objective in this problem is to find the pairs of corresponding points $p$ and $q$ that result from the projection of the same scene point into the two images. As shown in Fig. 61, the distance from the scene point to the cameras is determined by difference in image locations of points $p$ and $q$. This difference is called the disparity. The two cameras are called rectified pair if their positions differ only by a translation in the x-direction. Therefore, the horizontal disparity $p_x - q_x$ of corresponding pair $p$ and $q$ is inversely proportional to the depth of corresponding scene point, as shown in Fig. 62. To reconstruct the 3D shape of an object, one needs to determine the disparities of the correspondences between pixels of the images. Usually, these disparities are represented as gray levels in an image that is called the disparity map or depth map. An example of a depth map is shown in Fig. 63.

Finding the disparity map $f$ for a rectified stereo pair is an image matching problem (image labeling problem). Where, $I$ and $\hat{I}$ represent the left and right images, respectively. The set of label $L$ is the disparity range $\{\partial_1, \cdots, \partial_K\}$. However,
FIGURE 62 – Rectified stereo pair setup. The depth is inversely proportional to the disparity

\[ Z = \frac{B\lambda}{(p_x - q_x)} \]

FIGURE 63 – Example of the depth map: (a) one of the image pair and (b) the corresponding depth map.
the problem formulation discussed in previous section does not encode the constraints of the visual correspondence. The uniqueness is one of these constraints, where each pixel in $I$ corresponds to at most one pixel in $\tilde{I}$. Note, in the previous formulation two pixels in $I$ can be mapped to one pixel in $\tilde{I}$. The occlusion is another constraint, where some pixels do not have correspondences. In contrast, each pixel is assigned a label in previous formulation. To overcome these problems, Kolmogorov in [51] treated the two images symmetrically by computing the disparities for both images in the same time. In this case, $\mathcal{P}$ represents the set of pixels of both images and $f$ is the labeling of both images. To enforce the visibility constraints, the author in [51] modified the data penalty term in the energy function, Eq. (79), such that it is computed only for pixels that have the same disparity in both images. In other words, if pixel $p$ is located in the left image and pixel $q$ is located in the right image, then $D(f_p) = \varrho(p, f_p)^2 \delta(f_p = f_q)$ where $q = p + f_p$ and $p = q + f_q$ (e.g., if $f_p = \partial^1_x$, then $f_q = -\partial^1_x$). After minimizing the energy and finding the labeling $f$, the pixels are considered to be occluded if $q = p + f_p$ and $p \neq q + f_q$. Occluded regions can be filled with the average of their neighbors’ disparities.

1. Human Faces Reconstruction

As an application, the stereo matching approach is used to reconstruct human faces in a 3D face recognition framework. Fig. 64 illustrates the setup that is used to capture images. The setup parameters are shown in Table (11). Fig. 65 shows an example for a reconstructed face. More results are shown in Fig. 66.
FIGURE 64 – The system setup

TABLE 11
STEREO SETUP PARAMETERS

<table>
<thead>
<tr>
<th>Range (m)</th>
<th>Baseline B (m)</th>
<th>Zoom λ (mm)</th>
<th>Focus</th>
<th>Pan/Yaw (degree)</th>
<th>Tilt/Pitch (degree)</th>
<th>Roll (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6</td>
<td>200</td>
<td>Range</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
FIGURE 65 – Reconstruction results. (a) The stereo pair, (b) left the depth map and right the reconstructed shape.
FIGURE 66 – More reconstruction results. Left: One of the stereo pair. Middle: Frontal view from the reconstructed shape. Right: Side view from the reconstructed face
This dissertation addressed the image labeling problem. More specifically, it focused on image modeling, which is a very important component in the image labeling system. The dissertation proposed accurate mathematical models for image appearance and shape models in order to describe objects-of-interest in the images.

- An intensity model, which estimates the marginal density for each class in the given image was modeled using a new unsupervised technique based on maximizing a derived joint likelihood function.

- Spatial interaction that describes the relation between pixels in each class was modeled using a Markov-Gibbs random field with Potts prior where the parameters of the model were analytically estimated. Statistical results of more than two thousand synthetic images confirmed the robustness of the proposed analytical estimation approach over conventional methods.

- A new shape model was proposed. In this model, the shape variations between an object and its candidates were estimated using a new probabilistic model based on a Poisson distribution.

The image appearance models were used in a novel framework for automatic multimodal gray scale image labeling. A joint MGRF model was used to describe the input image and its desired map with more accurate model identification. No user interaction was needed; instead, the image was initially labeled using the proposed intensity model. An energy function using the appearance models was formulated, and was globally minimized using a standard graph cuts
approach. Experimental results showed that without optimizing any tuning parameters, the proposed approach was fast, was robust to noise, and gave accurate results compared to the state-of-the-art algorithms.

To exploit the modeling capability of high order MRFs and the efficiency of pairwise MRF solvers, this dissertation proposed an efficient transform that converts higher order Gibbs energies to a quadratic energies, for binary MRF. This transformation can be applied on many computer vision problems. In this dissertation it was demonstrated on color images segmentation. The experiments showed that the proposed approach performance was encouraging.

Another framework, which exploits the appearance models and the shape model was proposed. To get the optimal segmentation, a new energy function was formulated using these models and was globally minimized using a standard graph cuts approach. Experiments confirmed that the shape constraints overcame the gray levels inhomogeneities problem and precisely guided the graph cuts to accurate segmentations (with mean error $\approx 5.4\%$ and standard deviation $\approx 1.6\%$) compared to graph cuts without shape constraints (mean error $\approx 62.7\%$ and standard deviation $\approx 27.5\%$).

A. Directions for Future Research

There are many possible directions in which the work proposed in this dissertation can be extended or enhanced. These include, but are not limited to, the following:

- The proposed unsupervised framework is limited only for the multimodal gray scale images. Investigating a general framework that is suitable for a more general class of gray scale images and color images as well as texture images is a good extension.

- As in conventional approaches, the proposed work used the standard neighborhood systems (6-neighborhood system in 3D case or 4-neighborhood sys-
tem in 2D case). Studying the effect of selecting important neighbors from a data base of the object on the labeling result is a possible direction for future work.

- New methods for high order cliques Gibbs energies reduction can be investigated, such that the generated quadratic energies is submodular. In this case the optimization problem can be solved in polynomial time.

- The proposed shape model and its segmentation approach depend on aligned data set. So another possible direction is a graph cuts framework that simultaneously does both segmentation and registration processes.

- Another possible direction that could be investigated is the integration of the deformable model (active contour and level set models) into the graph cuts formulation.
REFERENCES


APPENDIX I
NOMENCLATURE

The following convention is used throughout this dissertation.

\( P(\cdot) \)  probability mass function
\( \mathcal{P} \)  set of image pixels
\( n \)  number of image pixels
\( \mathcal{G} \)  set of gray levels
\( \mathcal{L} \)  set of labels
\( Q \)  number of gray levels in the set \( \mathcal{G} \)
\( K \)  number of labels in the set \( \mathcal{L} \)
\( l \)  label in the set \( \mathcal{L} \)
\( \mathbf{I} \)  observed image, mapping \( \mathbf{I} : \mathcal{P} \rightarrow \mathcal{G} \)
\( f, f^*, \hat{f} \)  different labelling, mapping \( f : \mathcal{P} \rightarrow \mathcal{L} \)
\( f_p \)  label of the pixel \( p \in \mathcal{P} \)
\( \mathcal{F} \)  set of all labelings
\( \mathcal{F}_p \)  random variable defined on a location \( p \)
\( \mathcal{F} \)  “random field” set of random variables defined on \( \mathcal{P} \)
\( \mathcal{N} \)  neighborhood system
\( \mathcal{N}_p \)  neighborhood of a pixel \( p \)
\( U(f) \)  Gibbs energy
\( Z \)  normalizing constant in Gibbs distribution
\( T \)  control parameter called \textit{temperature} in Gibbs distribution
\( \mathcal{C} \)  set of all cliques
\( V_c, V_p, V_{pq}, V() \) potential functions

\( \gamma_0 \) controls the influence of the external field

\( \gamma \) influences the interaction between neighboring pairs or tripels

\( \mu, \sigma \) Gaussian distribution parameters \( \theta = (\mu, \sigma) \)

\( \pi \) prior weight or responsability

\( m \) message in BP and TRW-s approaches

\( \Theta \) vectors of model parameters

\( T \) family of the neighboring pixel pairs or tripels supporting the Gibbs potentials

\( N_{iter}, A, B \) constants

\( C_{p,l}, w_{p,r,l} \) parameters of positive and negative components of LCG

\( G \) weighted undirected graph

\( s, t \) terminals of the graph

\( V \) graph vertices

\( E \) graph edges

\( E_c \) set of edges that constitute a cut

\( |E_c| \) cut cost

\( U, S \) sets of pixels

\( S \) set of graph nodes belong to source

\( T \) set of graph nodes belong to sink

\( E \) quadratic energy function

\( E^k \) high order energy function

\( D(.) \) data penalty term

\( L(.) \) Likelihood function

\( \bar{\psi} \) relative frequency of labels in pixel pairs

\( \rho \) ratio \( |T|/|P| \)

\( \Delta, \delta(.) \) indicators, and indicator function

\( D \) AIC criterion

\( \varphi(.) \) Gaussian distribution
$h_s, h_c, M$  
EDISON parameters: spatial and color bandwidths and Min. region

$eem, nsg, of, th, et$  
NCUTS parameter: elongation parameter for edge map, number of segments, offset of the symmetric similarity matrix, symmetric similarity matrix threshold, and error tolerance in eigenesolver, respectively.

$\tau$  
execution time

$\epsilon$  
relative error

$\mathcal{M}_E$  
set of all global minima of energy $E$

$\mathfrak{A}$  
sum of all negative coefficients in $E^k$

$a_p, a_{pq}$  
real coefficients

$\mathbf{B}, \mathbf{u}$  
real numbers

$\mathcal{K}, \mathcal{R}, \mathcal{X}$  
object, background, and variability regions

$\vartheta(.)$  
Poisson distribution

$\xi$  
Poisson density rate

$\mathcal{P}_s$  
Shape image

$\mathbf{C}_{\mathcal{K}, \mathcal{X}}$  
object/variability contour

$d_p$  
normal distance from pixel $p$ to $\mathbf{C}_{\mathcal{K}, \mathcal{X}}$

$\mathbf{C}_{d_p}$  
iso-contour at $d_p$

$h_{dp}$  
histogram value at $d_p$

$M_t$  
number of training images

$I, \tilde{I}$  
stereo pair images

$\lambda$  
camera focal length

$\partial_x, \partial_y$  
displacements in image spatial domain

$Z$  
scene point’s depth
CURRICULUM VITAE

A. CONTACT INFORMATION

Asem Mohamed Ahmed Ali
May, 1976.

2202 James Guthrie ct., Apt. 7
Louisville, Kentucky, 40217 USA.
1-502-637-4526
asem@cvip.uofl.edu

B. RESEARCH INTERESTS

Image Processing, Statistical Modeling, Pattern Recognition, Computer Vision, Robotics, Learning Systems, Optimization

C. EDUCATION

University of Louisville, Louisville, Kentucky USA
Ph.D., Electrical & Computer Engineering, May, 2008,
• Dissertation Topic: “Image Labeling by Energy Minimization with Appearance and Shape Priors”
• GPA: 3.92
• Advisor: Aly A. Farag

Assiut University, Assiut, Egypt
M.Sc., Electrical & Computer Engineering, August, 2002
• Dissertation Topic: “Intelligent Tracking Control of a D-C Motor”

Assiut University, Assiut, Egypt
B.Sc., Electrical & Computer Engineering, June, 1999
• Distinction With Honor, Class Valedictorian.

D. HONORS AND AWARDS
IEEE Student Member since 2002.

E. PUBLICATIONS


F. REVIEWER

- MICCAI

G. SOFTWARE PROGRAMMING

11 years of software development experience.

- C (11 years)
- C++ (6 years)
- CORBA Platform (4 years)
H. BIOGRAPHY

Asem M. Ali has worked at the computer vision and image processing (CVIP) laboratory as a research assistant for 4 years (2004-2008). During this period, he was in charge of establishing the main infrastructure of the robotic research in CVIP. He led the CVIP robotic research team for autonomous navigation under a grant sponsored by the US DoD. Through the work in this project, he has developed an optical flow-based navigation algorithm, and a Kalman filter-based localization algorithm. This work has been approved for funding by NASA through two consecutive grants. Also, he has been assigned different projects in the lab. He was one of the members who developed medical imaging processing tools, the CVIP Lab’s CAD system. In this project, he used his new segmentation algorithms. Currently, he is in charge of building a human face recognition system. In this project he uses stereo to reconstruct the 3-D shape of human faces.

I. LANGUAGES

- **Arabic** (Mother Tongue)
- **English** Fluent (Read/Write)
- **French** Fair (Read/Write)

J. MEMBERSHIP

- The president of the Egyptian Student Association in North America (ESANA), Louisville’s Chapter (ESA), May 2006 -2007.