FUSION OF IMAGE SEGMENTATION WITH DOMAIN SPECIFIC
INFORMATION UNDER AN UNSUPERVISED MARKOV RANDOM FIELDS
MODEL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÖZGE ÖZTİMUR KARADAĞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

FEBRUARY 2014

Approval of the thesis:

FUSION OF IMAGE SEGMENTATION WITH DOMAIN SPECIFIC INFORMATION UNDER AN UNSUPERVISED MARKOV RANDOM FIELDS MODEL

submitted by **ÖZGE ÖZTİMUR KARADAĞ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Prof. Dr. Fatoş T. Yarman Vural
Supervisor, **Computer Engineering Department, METU** _____

Examining Committee Members:

Assist. Prof. Dr. Sinan Kalkan
Computer Engineering Department, METU _____

Prof. Dr. Fatoş T. Yarman Vural
Computer Engineering Department, METU _____

Assist. Prof. Dr. Ahmet Oğuz Akyüz
Computer Engineering Department, METU _____

Assist. Prof. Dr. Aykut Erdem
Computer Engineering Department, Hacettepe University _____

Assist. Prof. Dr. Pınar Duygulu Şahin
Computer Engineering Department, Bilkent University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ÖZGE ÖZTİMUR KARADAĞ

Signature   :

ABSTRACT

FUSION OF IMAGE SEGMENTATION WITH DOMAIN SPECIFIC
INFORMATION UNDER AN UNSUPERVISED MARKOV RANDOM FIELDS
MODEL

Karadağ, Özge Öztimur

Ph.D., Department of Computer Engineering

Supervisor   : Prof. Dr. Fatoş T. Yarman Vural

February 2014, 131 pages

The formulation of image segmentation problem is evolved considerably, from the early years of computer vision in 1970s to these years, in 2010s. While the initial studies offer mostly unsupervised approaches, a great deal of recent studies shift towards the supervised solutions. This is due to the advancements in the cognitive science and its influence on the computer vision research. Also, accelerated availability of computational power enables the researchers to develop complex algorithms. Despite the great effort on the image segmentation research, the state of the art techniques still fall short to satisfy the need of the further processing steps of computer vision. This study is another attempt to generate a "substantially complete" segmentation output for the consumption of object classification, recognition and detection steps. Our approach is to fuse the multiple segmentation outputs in order to achieve the "best" result with respect to a cost function. The proposed approach, called Boosted-MRF, elegantly formulates the segmentation fusion problem as a Markov Random Fields (MRF) model in an unsupervised framework. For this purpose, a set of initial segmentation outputs is obtained and the consensus among the segmentation partitions are formulated in the energy function of the Markov Random Fields model. Finally, minimization of the energy function yields the "best" consensus among the segmentation ensemble.

We proceed one step further to improve the performance of the Boosted-MRF by introducing some auxiliary domain information into the segmentation fusion process. This enhanced segmentation fusion method, called the Domain Specific MRF, updates the energy function of the MRF model by the available information which is received from a domain expert. For this purpose, a top-down segmentation method is employed to obtain a set of Domain Specific Segmentation Maps which are incomplete segmentations of a given image. Therefore, in this second segmentation fusion method, in addition to the set of bottom-up segmentation ensemble, we generate ensemble of top-down Domain Specific Segmentation Maps. Based on the bottom–up and top down segmentation ensembles a new MRF energy function is defined. Minimization of this energy function yields the "best" consensus which is consistent with the domain specific information. The experiments performed on various datasets show that the proposed segmentation fusion methods improve the performances of the segmentation outputs in the ensemble measured with various indexes, such as Probabilistic Rand Index, Mutual Information. The Boosted-MRF method is also compared to a popular segmentation fusion method, namely, Best of K. The Boosted-MRF is slightly better than the Best of K method. The suggested Domain Specific-MRF method is applied on a set of outdoor images with vegetation where vegetation information is utilized as domain specific information. A slight improvement in the performance is recorded in this experiment. The method is also applied on remotely sensed dataset of building images, where more advanced domain specific information is available. The segmentation performance is evaluated with a performance measure which is specifically defined to estimate the segmentation performance for building images. In these two experiments with the Domain Specific-MRF method, it is observed that, as long as reliable domain specific information is available, the segmentation performance improves significantly.

ÖZ

MARKOV RASGELE ALANLARI ARACILIĞI İLE ANLAM BİLGİSİ VE İMGE BÖLÜTLEMENİN BİRLEŞTİRİLMESİ

Karadağ, Özge Öztimur

Doktora, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi   : Prof. Dr. Fatoş T. Yarman Vural

Bilgisayarlı görmenin ilk yıllarından bugüne kadar imge bölütlemenin formüle edilişi büyük gelişim göstermiştir. İlk çalışmalar eğitimsiz yaklaşımlar önermiş, yakın zamanda yapılan çalışmaların büyük çoğunluğu ise eğitimli yaklaşımlara yönelmiştir. Bu durum bilişsel bilimlerdeki ilerlemeden ve bilişsel bilimlerin bilgisayarlı görme üzerindeki etkisinin artmasından kaynaklanmaktadır. Aynı zamanda, bilgisayar gücünün artması, araştırmacıların daha karmaşık yöntemler geliştirmelerine imkan sağlamıştır. İmge bölütlemedeki büyük çabaya rağmen, mevcut bölütleme yöntemleri, bilgisayarlı görmenin daha sonraki basamakları için gerekli çıktıları vermekte yeterli olamamıştır. Bu çalışmada nesne sınıflandırma, tanıma ve bulma yöntemlerinde kullanılabilecek "tam bir" bölütleme çıktısı elde etmek amaçlanmaktadır. Bu amaçla, çeşitli bölütleme çıktıları birleştirilerek bir maliyet fonksiyonuna göre "en iyi" bölütleme elde etmek hedeflenmektedir. Önerilen yöntem, Attırımlı-MRA, bölütleme birleştirme problemini bir Markov Rasgele Alanı (MRA) problemi olarak eğitimsiz yaklaşımla formüle eder. Bu amaçla, imge öncelikle fazlaca bölütlenir, elde edilen alanlar düğümleri, komşuluk ilişkileri kenarları oluşturacak şekilde Bölge Bitişiklik Grafiği (BBG) oluşturulur. Başlangıçta tüm kenar ağırlıkları bir birim olarak belirlenir. Aynı zamanda basit bir bölütleme yönteminin parametreleri değiştirilerek çok sayıda bölütleme sonucu elde edilir. Bu sonuçlar arasında bir fikir birliği oluşturacak şekilde MRA enerjisi formüle edilir. Son olarak, MRA enerjisi minimize edilerek bö-

lütleme sonucu elde edilir. Arttırımlı-MRA yönteminin performansını bir adım daha ileri götürmek amacı ile alan bilgisi bölütleme birleştirme sürecine entegre edilir. Bu yeni yöntem, Alana Yönelik MRA, alan uzmanından alınan veriyi MRA enerjisini güncellemekte kullanır. Alan bilgisi, mantıksal yüklemler aracılığı ile Alana Yönelik Haritalar (AYH) elde etmek için kullanılır. Böylelikle, Alana Yönelik MRA yöntemi, bir grup basit bölütleme sonucu ile bir grup AYH'yı kullanarak yeni bir MRA enerjisi tanımlar. Bu enerjinin minimize edimesi ile alan bilgisi ile uyumlu "en iyi" bölütleme sonucu elde edilir. Deneyler sonucunda önerilen bölüt birleştirme yöntemlerinin, birleştirilen bölütlerden her birinden daha yüksek performans elde ettiği görülmüştür. Arttırımlı-MRA yöntemi, literatürde sıklıkla kullanılan En İyi K (EİK) yöntemi ile de karşılaştılmış ve Arttırımlı-MRA'nın daha iyi performans elde ettiği gözlenmiştir. Alana Yönelik MRA yöntemi bitki örtüsü içeren dış mekan imgeleri üzerinde, bitki örtüsü alan bilgisi kullanılarak uygulanmıştır. Bu deneyde bir miktar performans artışı sağlanmıştır. Bu yöntem aynı zamanda uzaktan algılama veri kümesinde bina imgeleri üzerinde uygulanmıştır. Bina tanıma probleminde, gelişmiş alan bilgisi mevcut bulunmaktadır. Elde edilen sonuçların performansı, bina tanıma problemi için tanımlanmış bir bölütleme performans ölçütü ile değerlendirilmiştir. Bu iki deney, güvenilir alan bilgisinin mevcut olması durumunda bölütleme performansının büyük ölçüde arttığını göstermiştir.

Anahtar Kelimeler: Alana Yönelik Bölütleme, Markov Rasgele Alan, Alan Bilgisi, Bölütleme Sonuçlarının Birleştirilmesi

To Çağrı Can,

ACKNOWLEDGMENTS

TABLE OF CONTENTS

xiv

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

BF	Belief Propagation
BOEM	Best One Element Move
BOK	Best of K
CDHS	Contour Detection and Hiearchical Image Segmentation
DS-MRF	Domain Specific Markov Random Fields
DSI	Domain Specific Information
DSM	Domain Specific Map
EG	Efficient Graph Based Segmentation
ICM	Iterated Conditioanl Modes
MRF	Markov Random Fields
MS	Mean Shift
RAG	Region Adjacency Graph
RSST	Recursive Shortest Spanning Tree
SA	Simulated Annealing

CHAPTER 1

INTRODUCTION

This thesis is about image segmentation, which is one of the oldest, yet unsolved problems in computer vision. We begin with introducing the problem of image segmentation, then discuss the difficulties, the available techniques and finally, describe our contribution to resolve some of the problems in this area.

1.1 What is Image Segmentation?

Loosely speaking, image segmentation is partitioning an image into a set of "homogeneous" regions with respect to a criterion. Generally, the aim in segmentation is to simplify further processing steps of the image for a pre-defined task, such as object recognition, detection, classification, image retrieval etc. Segmentation problem arises in almost all of the image processing problems. Due to this common requirement for segmentation, starting from the early times of computer vision, various methods are proposed for segmentation. While the early segmentation studies formulate segmentation as the process of grouping similar image pixels together, recent studies formulate segmentation as grouping "semantically" similar pixels together.

1.2 The Major Difficulties in Image Segmentation

Despite extensive research, there are main problems in the segmentation process that remain to be resolved. Some common problems are listed as follows:

1. **Subjectivity:** There is not a unique solution for the segmentation problem. Even the human perception of "homogeneous regions" varies among people. This is clearly observed in the manually segmented ground-truth of the Berkeley Segmentation Dataset [84]. A sample image from this dataset and five ground-truth segmentations of the image is provided in Figure 1.1. As the images in this figure implies, there is not a unique segmentation partition. For this reason, the evaluation of the output of a segmentation method is subjective. Although various segmentation evaluation measures are proposed [114, 93, 113, 58, 122, 84], defining an objective segmentation measure for a subjective problem is itself an ill-defined problem.

(a)	(b)	(c)
(d)	(e)	(f)
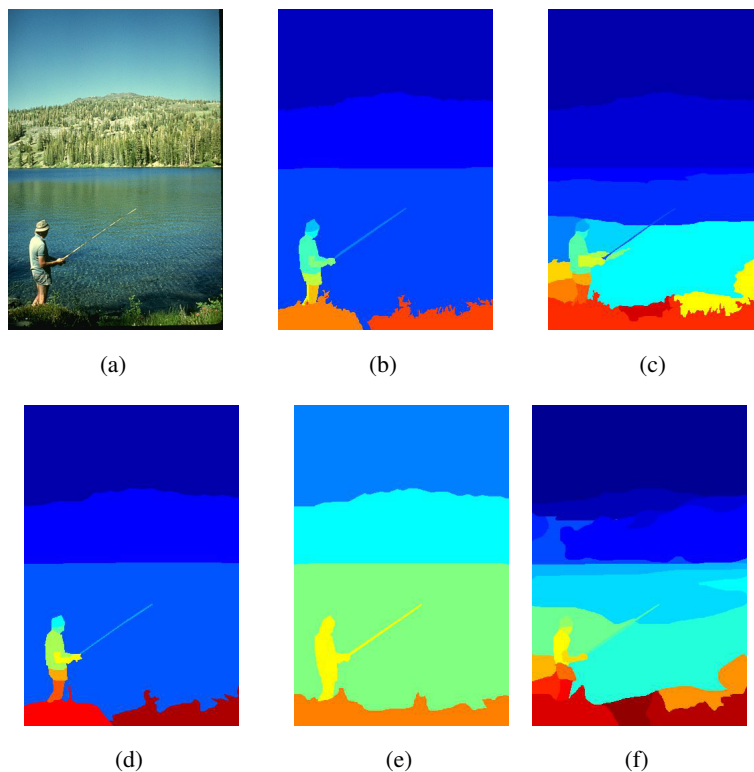
Figure 1.1: (a) A sample image from the Berkeley Segmentation Dataset [84] and (b),(c),(d),(e),(f) hand labeled ground-truth segmentations of this image. Each ground-truth segmentation is labeled by a distinct human subject.

2. **Parameter Tuning:** In the last two decades a wide range of approaches for the segmentation problem are proposed. A group of studies take up segmentation as a boundary extraction problem and propose edge based solutions, while another group of studies concentrate on the similarity of the pixels which fall into the same region and propose region based solutions. The studies modeling segmentation as the coupling of the edge and the region processes aim improving

the performance of the algorithms which are purely based on the edge process or the region process. All of these algorithms in the literature have a set of parameters which are determined by some heuristics, based on the requirements of the problem domain. Majority of the studies employ a validation set, in order to tune the parameters of the models. The sensitivity of a segmentation method to its parameters is depicted in Figure 1.2, where a sample image from [84] and two segmentations of this image obtained for two different parameter settings of a segmentation method are provided. The question 'Which one is "better"? ' remains unanswered, as long as the final goal of the process is unknown. Therefore, the problem gets complicated, since one should determine which parameter setting to use to get a "better" segmentation.

(a) (b)

(c) (d)
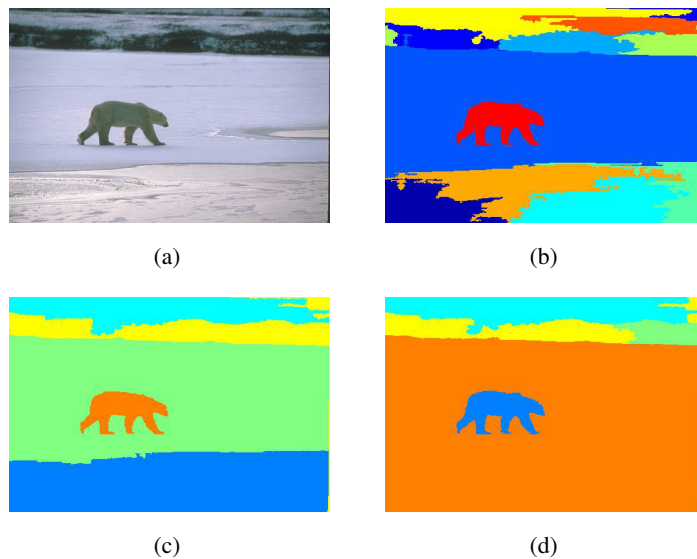
Figure 1.2: (a) A sample image from the Berkeley Segmentation Dataset [84] and (b),(c),(d) output of a segmentation method with three different parameter settings

3. **Semantic Gap:** Semantic gap is the difference between the low level image descriptors and the high level semantic information about the image regions, such as the object labels, relationship between the adjacent regions, etc. Semantic gap problem arises due to the insufficiency of the computational vision models in mimicking the human visual system. Although the vision scientists pay attention to the cognitive aspects of vision from the early years of vision research, a valid and practical model of the human visual system is not available yet. Understanding and modeling the human brain activity during vision

process is an inter-disciplinary study of many research areas such as psychology, cognitive science, neuro-science and computer science and gathering the research from all of these areas requires some more time to come up with an efficient and effective model.

Gestalt Laws of Perception, which date back to 1930s, underline the grouping attitude of human perception. Gestalt psychologists point out that the human visual system group things together to perceive the whole. Hence, there is strong evidence that a segmentation process takes place during perception. However, how the segmentation process takes place as an information representation and flow process is an unresolved problem of the brain. If the details of human visual system can be resolved, then an important step will be taken for bridging the semantic gap problem. In other words, if the modeling of high level and low level information and their interactions can be discovered in the human visual system, then its modeling in the computation vision will be much easier.

1.3 Available Image Segmentation Systems

In the rich segmentation literature, various solutions are proposed to handle the problems already mentioned in the previous section. The subjectivity problem can also be considered as a perception problem. And in the current state of the literature, an exact solution to this problem is not available.

Researchers propose to overcome the parameter tuning problem by consensus segmentation systems, which employ a set of segmentation partitions and find a final segmentation a pre-defined consensus among these segmentations. The consensus models optimize a cost function defined over the multiple segmentation outputs obtained by employing a set of segmentation methods. The segmentations are performed either by varying the parameters of a segmentation algorithm or employing different segmentation methods on the same image. The optimization of the cost function defines the "best" or the "most representative" segmentation output among the set of segmentations. Alternatively, a new segmentation output, which ensures the consensus among the set of initial segmentation outputs is constructed. In this way, the parameter tuning problem is resolved.
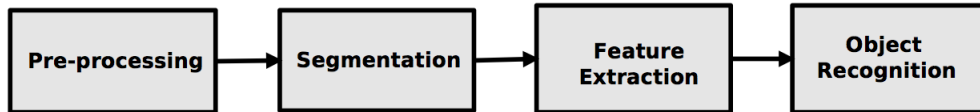
Figure 1.3: Feed-forward vision process

The major problem, which divides the segmentation studies into two groups as supervised and unsupervised segmentation, is the semantic gap problem. This problem gives rise to a new literature as "semantic segmentation". Let us comment on the evolution of segmentation methods from the unsupervised (bottom up) segmentation systems towards the semantic segmentation (top-down) systems.

In the early studies of 1980s, vision is considered as a feed-forward process as depicted in Figure 1.3. This approach is also referred as bottom-up or data driven model since it merely employs image data. In this model, for a given image initially a pre-processing step such as filtering or histogram equalization takes place. Then, image is segmented and features are extracted from the image partitions. Finally, the object recognition takes place. In this perspective of vision, it is assumed that the regions extracted by the segmentation method correspond to objects or object parts. However, this assumption may not be valid in many practical problems. Moreover, the extracted regions do not necessarily correspond to semantically meaningful regions. As the semantics gain more importance, the understanding and formulation of the segmentation problem are evolved. Cognitive aspects of the vision process gain more importance and human visual system attracts the attention of the vision scientists in an accelerating pace.

Experiments on the human visual system, reveals that the human perception of image is strongly affected by the high level cues. Inspired by the human visual system, researchers propose to utilize higher level information into the segmentation process [19, 79]. For this purpose, some machine learning methodologies are utilized to train a system using a set of labeled images for learning the features of certain objects. This approach is referred as top-down segmentation, since, high-level information related to image content is provided to the segmentation system before segmentation takes place. In other words, the segmentation system is aware of what is going to be

segmented in advance.

With the emergence of this approach, segmentation problem is defined as an image labeling process. For this purpose, pixel-wise labeled datasets are constructed [119, 84] and class labels are assigned to each image pixel at the output of a segmentation method as shown in Figure 1.4. This formulation of the segmentation problem is referred as the semantic segmentation, multi-class segmentation, image labeling or object based segmentation in the literature [19, 54, 105, 121, 78, 83].

Integration of high level cues into the segmentation process turns out to be very effective in the segmentation process. Nevertheless, more generic segmentation methods which can employ high-level cues may be required, in a specific application domain, such as medical image datasets, where a great deal of domain knowledge is available. In other words, a segmentation which is guided by high-level information may be required while a labeled dataset is not available. For those problems, segmentation approaches employing prior information is proposed. The prior information utilized in the literature is generally related to shape, appearance and location [37, 40, 45, 38, 57, 56]. Studies, usually employ one or two types of prior information during the segmentation process.

(a)　　　　　(b)
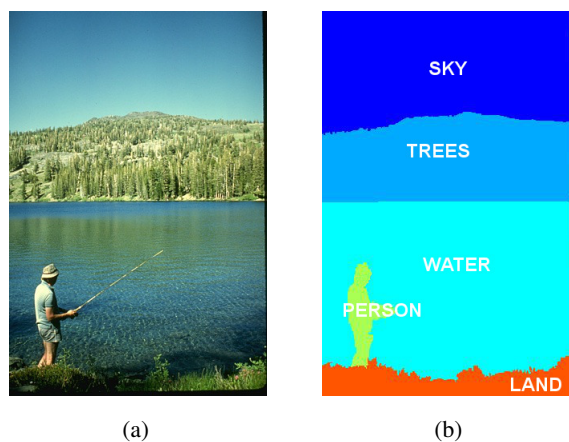
Figure 1.4: (a) A sample image from the Berkeley Segmentation Dataset [84] and (b) a realization of semantic segmentation

1.4 Our Contribution

In this thesis, we propose a segmentation fusion approach for problems where a labeled dataset is not available, while high level information related to the given problem domain is provided by a domain expert. This is the first study, which formulates the segmentation fusion problem as a Markov Random Fields Model. Therefore, our approach exploits the flexible formalism of the Markov Random Fields framework. The major contributions of this thesis can be listed as follows:

1. A new unsupervised segmentation fusion approach is proposed in this thesis. The method called, Boosted-MRF, finds the best consensus among the ensemble of multiple segmentation outputs, by minimizing the energy function of a Markov Random Fields model.

2. The Boosted-MRF method is further improved by incorporating the expert information into the energy function. This improved method called, Domain Specific-MRF segmentation optimizes the consensus with the constraints of information received from a domain expert. The proposed method finds the consensus among a set of segmentation partitions by minimizing the energy function of the Markov Random Fields Model. During the fusion of multiple segmentation outputs, the high level semantic information is incorporated to weight the smoothing term of the energy function of the Markov Random Fields model. This approach enables us to revise the energy function with respect to the available semantic information. In this study, high level semantic information is referred as domain specific information and it appears in many forms; appearance, shape, shape-relations, objects, object-relations. In this thesis, the high level cues provide more complex prior information compared to the similar studies employing prior information during segmentation in the literature.

3. Segmentation fusion problem is formulated as an energy minimization problem, where the energy function is used to ensemble the segmentation outputs together with the high level information. The main contribution of this thesis is to propose a generic segmentation method whose energy function can be

7

designed based on the problem domain

1.5 Organization of the Thesis

The thesis begins with an overview of the image segmentation methods in the literature, in Chapter 2. Then, two unsupervised segmentation fusion methods are introduced in Chapter 3 and Chapter 4. The unsupervised segmentation fusion method of Chapter 3, called Boosted MRF, does not utilize any auxiliary information about the domain or the images in the dataset. On the other hand, the segmentation fusion method of Chapter 4 employs some high level information about the domain. The suggested segmentation fusion method of Chapter 4, called Domain Specific MRF can be considered as unsupervised, since it does not require any labeled training data. The experimental studies and the analysis of the proposed methods in Chapter 3 and Chapter 4 are provided in Chapter 5. The thesis is concluded with a summary of the contributions and future work in Chapter 6.

CHAPTER 2

AN OVERVIEW OF IMAGE SEGMENTATION METHODS

2.1 Introduction

Image segmentation is one of the most widely studied research areas in the literature. This is due to it is critical position between the representation and the recognition tasks.

Segmentation process is approached in wide range of perspectives, depending on the application domain. Studies in the literature can be grouped in several ways. In the early studies, they are commonly categorized as region based and edge based systems, depending on whether the segmentation algorithm is based on similarity or dissimilarity. Recent studies categorize segmentation systems as top-down and bottom-up methods. Early studies are based on bottom-up segmentation methods, which process information extracted from the image. These systems are also referred as image based segmentation methods. The problem related to these systems is that, they lack semantics. In other words, the regions obtained by a bottom-up segmentation do not necessarily correspond to the objects or object parts. In order to solve these shortcomings, in the last decade top-down segmentation methods have emerged. These methods integrate semantic information into the segmentation process, so that a semantically meaningful segmentation is obtained. Majority of the studies employ class labels as semantic information, where each image pixel is assigned a label [10, 83, 54]. This process is referred as semantic segmentation, multi-class segmentation or image labeling [10, 54, 60] in the literature. Nevertheless, the level and the form of semantic information may vary depending on the problem domain.

The organization of the thesis is as follows; first we discuss general issues and problems about the image segmentation problem. Then we give an overview of the state of the art segmentation systems, and elaborate the topics that are closely related to the proposed system in this thesis.

2.2   Domain Dependence in the Segmentation Process

Image segmentation is formulated in various forms in the literature. Construction of the segmentation systems depends on the problem domain and the type of available information. When the segmentation is formulated as a labeling process, a large number of training data is required. However, a labeled dataset with sufficient statistics is not always available. Even if it is available, the probabilistic models may not discriminate the object regions due to large within class variances. On the other hand, some type of prior information related to a given problem is generally available. And utilization of this information would improve segmentation performance considerably. This prior information, which may be available in any form, is referred as *domain specific information* and any problem for which some domain specific information is available is referred as *domain specific problem*.

Domain specific problems arise in a wide range of areas, such as medical image processing, remote sensing, document image analysis etc. From another point of view, long time studies reveals valuable information related to a given problem domain, which can later be employed as domain specific information. For instance, in the remote sensing studies, researchers discovered that certain low level features are more informative for sea detection, while other features are more informative in land cover classification [1, 112, 22]. Moreover, if illumination information is available in some computer vision applications, it is utilized to obtain shades in the image [6, 126], which is further employed during recognition tasks such as building detection, urban detection etc. Similarly, in medical imaging, spatial information among body parts, shape information and statistical information known a-priori is utilized in the segmentation process [38, 57, 56].

Domain specific information may arise in different forms and it may be directly re-

lated to the image content. It may be related to objects in the image or background of the image. For example, in an image processing task on a production line of a company, generally background have a standard color and illumination condition. Here, the problem is to segment foreground object as accurately as possible so that, problematic products are eliminated. In such a problem, segmentation accuracy of foreground object can be increased by introducing information related to the background.

The main goal of this study is to propose a segmentation framework, which can integrate various forms of domain specific information into the segmentation process. Since we are mainly interested in the level of available information and its integration into the segmentation process, segmentation studies in the literature are categorized with respect to the type and level of the information. In the following sections the segmentation methods are overviewed under three major headings;

- Bottom-up segmentation, which does not employ any information,

- Top-down segmentation, which employs a labeled set of training data,

- Segmentation fusion, which combines multiple segmentation outputs.

2.3 Bottom-Up Segmentation

Bottom-up segmentation is also referred as *image based* approach since it determines image regions directly by processing low level image features such as intensity, color and/or texture. This low level information from image pixels and their spatial proximity is considered for determining homogeneous regions in bottom-up segmentation approaches.

Segmentation problem is formulated in various forms by bottom-up approaches. Main categories in the bottom-up approach can be listed as; region based and edge based methods [52, 106, 5, 35, 82, 55], clustering methods which estimate mixture densities [28, 4, 59, 15, 31], spectral clustering methods [103, 42, 29, 102], and methods formulating segmentation as an energy minimization process [86, 8, 62, 67, 68, 32].

2.3.1 Thresholding Based Segmentation

Thresholding based segmentation is the simplest method of image segmentation. Depending on the type of threshold or the similarity measure, thresholding methods can be categorized into several groups such as; global thresholding, local (dynamic) thresholding, multiple thresholding. The critical step in the thresholding based segmentation is determining the threshold. This is usually achieved by histogram shape analysis, mixture modeling or clustering methods. Among these methods Otsu's thresholding method, which selects a threshold that minimizes intra-class variance, is the most popular thresholding based image segmentation method [52].

2.3.2 Region Based Methods

Region based segmentation methods define a similarity criterion and group image pixels based on this criterion. Region splitting and merging and region growing are well known region based segmentation systems [52]. Major problem in region splitting and merging is determining the split and merge threshold values. On the other hand, the region growing methods suffer from the selection of the initial seeds. Various solutions to these problems are proposed and novel region based segmentation systems are introduced in the literature [5, 35].

2.3.3 Edge Based Methods

Edge based segmentation methods are based on detecting the discontinuities in the image. Initially the edges are detected then boundaries are extracted by linking the edges [52]. Gradient operators, adaptive local operators are in this group. Edge based approaches are employed in the recent studies such as [82, 55]. The main problem of the edge based methods is that the detected edges may not provide a closed boundary, hence a post processing is required.

2.3.4 Morphological Methods

Morphological image processing techniques model image as topographic data [52]. Watershed segmentation, which is a well known morphological image segmentation system, is based on detecting regional minimum, catchment basin and watershed lines by morphological erosion and dilation operations [52]. Watershed segmentation outputs closed boundaries, however it generally provides over-segmented results. Watershed segmentation is used in many application areas, such as medical image processing [56, 87, 48] and remote sensing [6]. Gaetano et al. [48] propose an Edge, Mark and Fill strategy which automatically generates markers by morphological operators in order to obtain a better initial watershed segmentation. Nguyen et al. [88] propose incorporating prior information into watershed segmentation by deriving an energy function while defining the watershed line.

Recently, Arbelaez propose formulating image segmentation as a contour detection problem by constructing a hierarchical region tree using the output of a contour detector [9]. Initially, they obtain contours by processing local and global cues together. Brightness, color and texture are employed as local cues and eigenvectors are employed as global cues for contour detection. Once contour signal is obtained, seed points with minimum contour response are determined and watershed transform is employed to find fine regions. These regions are later merged iteratively, meanwhile a hierarchy of regions is obtained. They compare the system with state of the art segmentation systems and obtain very good segmentation results. One drawback of their system is the high computational complexity due to its hierarchical structure.

2.3.5 Variational Methods

Variational methods formulate image segmentation as a minimization of a function which consists of a data term and a smoothness term. Mumford-Shah proposes a continues solution by cartoon model [86] for modeling image. Mumford-Shah functional attracts the attention of many researchers in this area [8, 62]. Similarly, active contour models, level set methods formulate segmentation by differential equation [66, 24, 125]. Variational methods are suitable for incorporating prior information

into the segmentation process, through the cost functions. As a result, this group of segmentation methods are employed simultaneously for object detection and segmentation [39, 40, 37].

2.3.6 Clustering Based Methods

Clustering methods basically estimates the mixture density of the image pixels or regions. Clustering algorithms, such as k-means [59], fuzzy c-means [15], mean-shift [28] are used for image segmentation by estimating region similarities on intensity, color or texture values. However, density estimation methods do not group pixels with respect to their spatial relations, which is a very critical criterion in segmentation. Deng and Manjunath propose solving this problem by a two stage segmentation process. They propose JSEG [31] algorithm, which first performs color quantization and then applies a spatial segmentation method. In the first stage, a color-map is obtained where each pixel is replaced with its representative color-map label. In the second stage, a homogeneity criterion is employed on color-map and J-image is obtained which is further utilized by a region growing algorithm to obtain final segmentation.

Comaniciu and Meer propose using Mean-Shift algorithm for image segmentation. They employ color values and spatial relations simultaneously for image segmentation [28]. In this method, a bandwidth for both feature values referred as range bandwidth and a bandwidth for neighborhood referred as spatial bandwidth is set initially. The algorithm is run in an expectation-maximization framework, where it starts with a set of initial means, which are updated iteratively until a pre-defined criterion. Unfortunately, Mean-Shift segmentation is quite sensitive to initial means and parameters; range bandwidth and spatial-bandwidth. Determining right values for these parameters require further observations or cross-validation which may not be possible at some problem domains. Nevertheless, Mean Shift is utilized in a large number of studies in the literature. Especially, most of the comparative segmentation studies, studies on segmentation evaluation measures, consensus segmentation studies employ Mean Shift. Sometimes, it is employed at the pre-processing step of a segmentation system to find super pixels or to find and initial partition.

14

Achanta et al. [4, 3] propose method called SLIC (Simple Linear Iterative Clustering) based on k-means clustering of super-pixels. The algorithm starts with a set of initial seeds, and iteratively moves these seeds to the lowest gradient position in a 3×3 neighborhood. It obtains compact superpixels by limiting the search space for seed centers, meanwhile keeping the computational complexity small.

2.3.7   Spectral Clustering Methods

Shi and Malik propose a spectral clustering method, called Normalized Cuts [103] which models image segmentation as a graph partitioning problem. In their formulation, the nodes of graph correspond to image pixels and the weights of edges correspond to similarity between two pixels, and segmentation is formulated as finding the minimum cut of the graph. Minimum cut problem penalizes large segments, hence the cost function of a cut is normalized by dividing the cost by segment sizes. This minimization problem is approximated as a generalized eigen value problem.

Felzenszwalb and Huttenlocher propose the Efficient Graph-Based Image Segmentation method [42] which also models image segmentation as a graph partitioning problem. Similar to Normalized Cuts, nodes correspond to pixels and edge weights correspond to dissimilarity between neighboring pixels. Each region is represented by a connected component in a graph. Quality of a segment is measured by comparing edge weights in the same component with that of between two components. A segmentation is considered as "good" if it has low weight edges between two vertices in the same component and high weight edges between vertices in different components. The algorithm starts with each pixel as a distinct region and two regions are merged if their difference (minimum weight edge between two regions) is smaller than their internal differences (largest weight in the minimum spanning tree of the component). The advantage of the method is that it provides regions that obey global properties defined by the merging criterion and too coarse or too fine regions are avoided. Moreover, the algorithm is fast in practice compared to the Normalized Cuts Segmentation.

Cour et al. propose Multiscale Normalized Cut method [29]. They construct a graph at several scales and define relations among neighboring scales. In this method, short

range and long range dependencies are modeled simultaneously. Contrary to other multi-scale approaches, this approach does not find the optimal graph cut iteratively, but it formulates segmentation as a constrained multiscale graph cut problem.

Sharon et al. propose Segmentation by Weighted Aggregation (SWA) method which formulates image segmentation as an algebraic multi grid method for solving Normalized Cut criterion [102]. In this method, pixels are represented by the nodes and coupling of pixels are represented as edges weights. Initially, half of the pixels are selected as the seed nodes and strongly coupled nodes are merged to obtain nodes of the next scale. At the new scale, properties of each node and edge weights are estimated. At each iteration, a set of seed nodes are selected and the process is continued recursively. At the output of the algorithm, a hierarchy of regions is obtained. The method has the power of extracting the salient regions at different scales, which may correspond objects of different sizes. Since a subset of nodes are selected at each iteration, the method is computationally more efficient than normalized cut image segmentation.

2.4 Top-Down Segmentation

Top-down segmentation approach is also referred as *class based* segmentation. In this approach segmentation is directly based on the properties of well defined class(es). Segmentation problems constructed in top-down approach are generally supervised problems, where initially certain image properties or system parameters are learned from a labeled dataset.

Most of the supervised segmentation studies in the literature assign a label to all the pixels in the image. For this purpose, a pixel-wise labeled dataset is required to train the segmentation system. These systems are referred as image labeling, multi-class segmentation, object based image segmentation, detection driven segmentation or class driven segmentation in the literature. In most of the methods, the algorithm is trained before the segmentation takes place.

2.4.1 Template or Fragment Based Systems

Among the top-down segmentation methods, template-based or fragment-based representation of image receive attention. For example; Ullman and Borenstein [19] proposed a top-down segmentation system where an object class is represented as a set of fragments which are first localized and then their relative position is formulated as a global constraint to obtain the final segmentation. The main problem in this approach is that, modeling a large variability of shapes within an object class may not be possible in most of the practical problems. Hence, this approach can be employed for a limited number of object classes.

2.4.2 Knowledge Based Systems

This group of studies employ high level information in the form of ontologies or description logics [10, 11]. Thus, it is assumed that the domain knowledge is available a priori in these studies. Relations between low level visual features and high level semantic concepts as well as relations among concepts are modeled by using the ontologies. Athanasiadis et al. [10] propose a graphical representation for their ontology. Their ontology consists of concepts corresponding to the nodes of the graph, relations among these concepts corresponding to the edges of the graph. Concepts are provided a priori while relations are learned from a training set. An Attributed Relation Graph (ARG), which carries both low-level visual information and high level fuzzy candidate label information in its nodes and semantic similarity information in its edges, is constructed. Once the ARG is constructed, a semantic segmentation is obtained either by semantic watershed which employs watershed segmentation on ARG or by recursive shortest spanning tree (RSST) algorithm. Berka et al. [13] employ similar representation using ARG. However, they estimate the dissimilarity between image regions using an expert system. They introduce semantic information as rules of an expert system. By means of the rules, they introduce reasoning to visual context analysis. The rules are defined on fuzzy membership values of regions. The membership of regions, the relative positions to their neighbors and their neighbors' membership values are considered to find a dissimilarity measure between two regions. Although robust systems can be built using knowledge based and rule based

systems, they require detailed expert knowledge.

2.4.3 Aspect Models

Aspect models also known as Topic or bag of words model were originally proposed for the text data. These models are introduced to computer vision literature by Fei Fei and Perona in [80]. The idea is based on the Latent Drichlet Allocation (LDA) [17] or Probabilistic Latent Semantic Analysis (pLSA). In this model, it is assumed that each image patch corresponding to a visual word is generated independently given its corresponding latent topic. Although the assumption holds for text documents, in the case of images, spatial information is essential. Hence, initial studies barely based on latent topic models [80] are later followed by studies that integrates spatial relations into the topic models [23, 123]. In [23] image is represented hierarchically and it is assumed that over-segmented and homogeneous image patches are assigned to same latent topic. While in [123] latent topics are modeled via Markov Random Fields which is widely used for representing spatial relations.

2.4.4 Graphical Models

The studies combining segmentation and recognition processes become popular in the recent years. For this purpose graphical models such as Markov Random Fields (MRF) and Conditional Random Fields (CRF) are employed in image segmentation [71, 16]. While MRF, as a generative model, is mostly utilized by bottom-up approaches [67, 68, 32], CRF, as a discriminative model, is utilized when higher level information is available. The approaches based on Markov Random Fields based approaches are explained in section 2.4.4.1, and the CRF methods are explained in section 2.4.4.2.

2.4.4.1 Markov Random Fields

In this section we review the literature, related to the segmentation fusion and domain specific segmentation systems proposed in this study. Hence, this section introduces

the foundations of this thesis. Additionally, some algorithms given in this section are adapted to the proposed systems.

Markov Random Fields are widely used for segmentation of images starting from the pioneering work of Geman and Geman [49], who are inspired by energy minimization of physical systems, which can be modeled by the Gibbs Distribution.

Early studies employed the MRF model for bottom-up segmentation [77, 120]. Recently, MRF models are also applied to the top-down segmentation problems [75, 105, 53].

Segmentation problem is defined as the process of assigning labels to a set of sites, where labels correspond to the indexes of homogenous regions in the unsupervised segmentation problem. Let s be a set of m discrete sites:

$$s = \{s_1, ..., s_m\}, \tag{2.1}$$

and labels are discrete values as follows:

$$f_{s_i} \in L = \{1, ..., M\}. \tag{2.2}$$

Let F be a family of random variables defined on s, where each site is assigned a label and the joint event $F = f$ is a realization of F where $f = \{f_{s_1}, ..., f_{s_m}\}$ is called a configuration of F.

Image segmentation can be formulated as a labeling process, where image pixels correspond to sites and region indexes correspond to labels. In other words, segmentation is a mapping $f : s \rightarrow L$. This process is formulated by Bayes Theorem as follows; let x be the observations. Then, segmentation is the the process of estimating the configuration f which maximizes the a posteriori probability (MAP)

$$f^* = \arg \max_{f \in V} P(F = f | x), \tag{2.3}$$

where V is the set of all configurations $V = L^m$. The posterior probability can be computed using the Bayes rule:

$$P(f|x) = \frac{p(x|f)P(f)}{p(x)}, \tag{2.4}$$

(a)Neighborhood   (b)Cliques
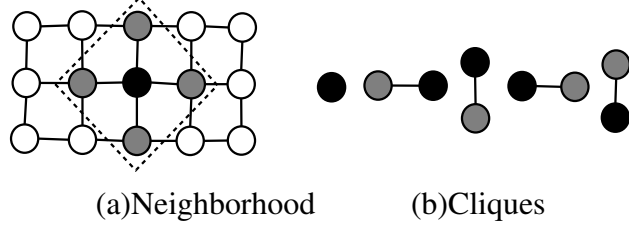
Figure 2.1: (a) Neighborhood (b)Unary and pairwise cliques defined for the neighborhood defined over the window

where $p(x)$ is the probability density function of x, $P(f)$ is the a priori probabilities of configurations f, and $p(x|f)$ is the likelihood densities of observations x. In order to find the MAP solution, one needs to derive the prior probabilities and the likelihood functions.

MRF model provides a probabilistic framework for modeling spatial and contextual dependencies. For this purpose, first a neighborhood system for s is defined as;

$$N = \{N_{s_i} | \forall s_i \in s\}, \tag{2.5}$$

where N_{s_i} is the collection of sites neighboring to s_i. The sites s and the neighborhood relation N is used to form a graph. In this graph, a set of cliques, C is defined. A clique is a subset of s. A clique of one site, c_1, is referred as *singleton clique* while clique of two sites, c_2, is referred as *doubleton clique*. The singleton and doubleton cliques for a given neighborhood is depicted in Figure 2.1. While generally, cliques consisting of maximum two sites are employed, a clique can contain more than two sites as well. The collection of all cliques defined on s and N is $C = C1 \cup C2 \cup ....$

The labeling process, f, is an MRF process if it satisfies two properties:

- Positivity: $P(F = f) > 0, \forall f \in U$.

- Markovianity:
  $P(F_{s_i} = f_{s_i} | F_{s_j} = f_{s_j}, s_j \in s, s_j \neq s_i) = P(F_{s_i} = f_{s_i} | F_{s_j} = f_{s_j}, s_j \in N_{s_i})$.
  The Markov Property states that the conditional probability of an event at site s_i, given all the remaining sites can be completely defined in terms of the conditional probability of the event at s_i given the events at the neighbors of s_i.

An MRF process can be modeled by Gibbs distribution as stated in Markov-Gibbs

equivalence by Hammersley-Clifford theorem [14]. This theorem states that, F is an MRF on s with respect to N if and only if the probability distribution $P(F = f)$ on the configurations is a Gibbs distribution with respect to N. And the Gibbs Distribution of the configurations f with respect to N is as follows:

$$P(f) = \frac{-1}{Z} \times e^{-\frac{1}{T}U(f)}, \tag{2.6}$$

where, Z is a normalizing constant, T is a global control parameter and $U(f)$ is the energy function, which has the following form:

$$U(f) = \sum_{c \in C} V_c(f) = \sum_{\{s_i\} \in C_1} V_1(f_i) + \sum_{\{s_i, s_j\} \in C_2} V_2(f_{s_i}, f_{s_j}) + ..., \tag{2.7}$$

where, the energy term is defined on the cliques of the graph defined by s and N.

MRF energy is formulated as follows:

$$U = U_{data} + U_{smooth}, \tag{2.8}$$

where generally, the data term corresponds to the unary potential V_1 and the smooth term corresponds to pairwise and higher order potentials $V2, ...$

The data term (the unary potential or the singleton potential) measures the compatibility between the observed data x_{s_i} and the assigned label f_{s_i}. It is usually assumed that the image features are drawn from a Gaussian Distribution. And the unary potential is modeled as the negative log likelihood of a label f_s being assigned to site s. Therefore, the data function is usually modeled as follows:

$$V_1(f_s) = \sum_{s \in S} ln(\sqrt{2\pi\sigma_{f_s}}) + \frac{(x_s - \mu_{f_s})^2}{2(\sigma_{f_s})^2}, \tag{2.9}$$

where, σ_{f_s} and μ_{f_s} are the model parameters of the region index f_s which is assigned to the site s.

Smoothness term ensures that the labeling process reflects the spatial consistency in the image. In other words, it is a measure of piecewise smoothness in the image. This is accomplished by defining a smoothness criteria between neighboring pixels(superpixels). In this representation, definition of the neighborhood relations is a critical issue. Majority of the studies in the literature model the MRF energy

over pairwise relations which is modeled by Potts Model [12, 95]. Potts model is the simplest discontinuity preserving model which is adapted from statistical physics, as follows:

$$V_C = \delta(f_{s_i}, f_{s_j}) = \begin{cases} +1 & \text{if } f_{s_i} \neq f_{s_j} \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

Various versions of Potts model are encountered in the literature. One direct modification is contrast sensitive Potts model [70], where the function takes a value, based on the level of difference between two neighboring sites. On the other hand, a group of studies propose higher order MRF energy by introducing a third term which is utilized for modelling long range dependencies in the image.

Image segmentation systems, which employ the MRF model consists of two major steps:

- Constructing the MRF energy function

- Minimizing the MRF energy

In the next subsection, we overview the literature for these steps.

Contruction of the MRF Energy Function

Early studies developed on the MRF based image segmentation are employed for textured images. In 1987, Won and Derin proposed a Markov Random Fields model for segmenting the noisy textured images [120]. Their system consists of three levels of random fields which they refer as region process, texture process and noise reduction. At the lowest level, some noise is introduced to the image during image formation and recording. In the middle level, texture process which is modeled by MRF takes place and texture patterns are formed. In the highest level, region process is modeled by MRF and each image pixel is assigned a region label. MRF energy formulated in this view is minimized by using simulated annealing.

Lakshmanan and Derin propose an MRF based image segmentation system for gray scale images [77]. They assume that the number of regions and corresponding mean gray-level for each region is known. Then, they propose a simultaneous parameter

estimation and segmentation system based on MRF model. They employ simulated annealing in an expectation-maximization framework, to optimize the energy function.

Most of the early studies on MRF based image segmentation methods concentrate on textured images. Panjwani and Healey propose to use color features in their MRF model [92]. They develop an agglomerative hierarchical clustering method which models image regions as Gaussian Markov Random Field model, for segmentation. Their main contribution is to model spatial relations as interactions within and between (R,G,B) color planes.

Göktepe et al. propose an unsupervised image segmentation method based on MRF model for textured images [51]. They obtain an initial coarse segmentation by hierarchical self organizing map and improve this segmentation by minimizing the MRF energy using simulated annealing. The energy function consists of two parts;

- first order gray level Potts model for modeling region geometry,

- higher order gray level MRF model for texture filling in those regions.

Kato propose a multi-layer MRF model for image segmentation at multiple resolution [67]. An MRF model is employed at each layer with a different feature and a combined MRF model is employed at a special layer, which interacts with all levels and obtain a final segmentation based on combination of features from those layers.

Deng and Clausi propose an MRF based segmentation method similar to the work of Lakshmanan and Derin [77], which simultaneously estimates MRF model parameters and obtains image segmentation [30]. They refer to the first energy term as feature modeling and the second term as region labeling and minimize this energy by using simulated annealing. Their contribution is to introduce the variable weighting parameter between the two components of the energy function. This variable weight ensures that feature modeling is emphasized at initial iterations of simulated annealing, so that MRF model parameters are learned. As the number of iterations increase, the variable weight function ensures that emphasize is given to the region labeling.

Kato et al. propose an image segmentation system based on MRF, which combines

color and texture features [68]. They propose employing CIE-L*u*v color features and Gabor filters by putting them into one multivariate normal mixture, whose modes correspond to clusters of pixels with homogeneous color and texture properties. They employ expectation maximization to optimize the MRF energy, which starts with an initial labeling and estimates model parameters during maximization step, and continue with the expectation step which finds an estimate for labeling using maximum likelihood with current model parameters.

Kohli et al. [70] propose a higher order clique potential; P^n Potts model which is defined for a clique of size n and takes a value γ_k if all the sites in the clique are assigned to same label, and it is assigned a value γ_{max} otherwise, where $\gamma_{max} > \gamma_k$. Higher order potential functions are expected to model long range dependencies in the image. Kohli et al. also proposed region based consistency potential which takes group of pixels and measure the level of consistency among the labeling of pixels in this region.

Flach and Schlesinger propose incorporating shape priors into MRF based segmentation [45]. For this purpose, they adopt probabilistic shape priors from level-set based models and introduces a shape prior function into the Potts interactions.

MRF Energy Minimization

Most of the initial segmentation studies based on MRF employ **Simulated Annealing** (SA) for minimizing the MRF energy function [51, 120, 77]. Theoretically, if provided with right parameters and iterates sufficiently, it gives global optimum. However, in practice, depending on the size of the problem, SA may provide local optimum and it is computationally prohibitive for real world problems. Derin and Elliott propose employing dynamic programming for minimizing the MRF energy function[32].

Besag [14] propose **Iterated Conditional Modes** (ICM) for minimizing the MRF energy, which is a deterministic algorithm based on the idea of optimizing local energy iteratively. ICM achieves global minimum if the energy function is convex. However, generally, the MRF energy functions may not be convex and the ICM ends up with a local minimum.
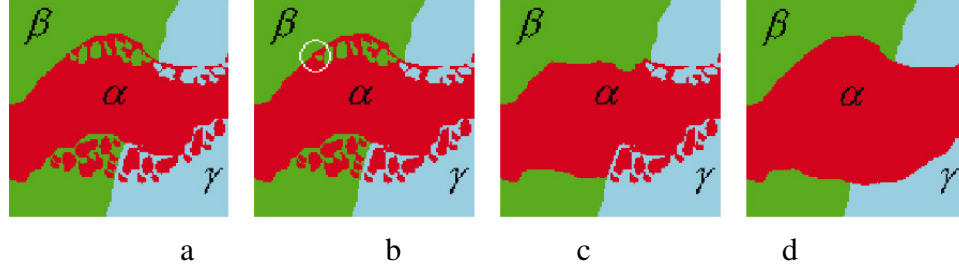
Figure 2.2: Types of moves: ($a \rightarrow b$) is standard move, ($a \rightarrow c$) is α-β-swap and ($a \rightarrow d$) is α-expansion. (Taken from [21])

Later, together with the advance in graphical methods, algorithms for minimizing MRF energy as a graph optimization problem are proposed. Boykov et al. [21] propose to minimize MRF energy function using graph cut methods; **expansion** and **swap**. These methods are computationally more efficient than the classical methods and provide approximate solutions. Contrary to the classic algorithms such as simulated annealing, where the label of only one pixel is changed at an iteration, in swap and expansion moves the label of a set of pixels is changed at once. Types of moves are explained in *Figure 2.2* which is taken from the study of Boykov et al. [21].

Steps of the Swap and Expansion algorithms are given in *Algorithm 1* and *Algorithm 2*, respectively. At each iteration of the algorithm a graph is constructed whose structure is determined dynamically by the current partition and the α (and β) label(s). The optimal expansion(swap) move is decided by the minimum cut method of [20] on this graph. The structure of the graph for finding the optimal swap move is provided in Figure 2.3. The vertices of the graph are the two terminals α and β and the set of pixels P_α and P_β which represents the set of pixels labeled as α and β respectively. Each pixel is connected to the terminals α and β by t-links and neighboring pixels are connected by n-links. Finding the minimum cut in this graph reveals the labeling with the minimum energy.

Similarly, the graph for expansion moves, which is provided in Figure 2.3, is dynamically determined. The set of vertices of this graph includes the terminals for α, $\bar{\alpha}$, the image pixels P and the auxiliary nodes, which are created at the boundaries between image partitions. Each pixel is connected to the terminals α, $\bar{\alpha}$ by t-links and to its neighboring pixels by n-links. Finding the minimum cut in this graph reveals

the labeling with the minimum energy.

The minimum cut method employed in the swap and expansion algorithms are proposed in [20]. In this method, two search trees are built, one from the α terminal and the other from the β or $\bar{\alpha}$ terminal as shown in Figure 2.4. In this graph, the internal nodes of the trees are passive nodes, while the outer nodes are the active nodes which enable the trees to grow by acquiring new children. The algorithm consists of three stages which iteratively repeats:

- "growth": search trees generated from the terminal nodes grow until they touch. As the trees touch a path p is constructed between the terminals.

- "augmentation": p is augmented and some edges become saturated and the tree(s) break into forest(s).

- "adoption": trees are restored. The nodes which are left without parents due to saturated edges are referred as the orphan nodes. They look for parents in the same tree with them. If they can not find a parent, they become free nodes. Similar case hold for the children of the orphan nodes.

The algorithm stops when the trees can not grow, which implies that the maximum flow is achieved.

Algorithm 1 Swap Algorithm

Start with an arbitrary labeling f

Set success $= 0$

for each pair of labels $\{\alpha, \beta\} \subset L$ **do**

    Find $\hat{f} = \arg\min E(f')$ among f' within one α-β swap of f

    If $E(\hat{f}) < E(f)$, set $f = \hat{f}$ and $success = 1$

end for

If $success = 1$ go to 2

Return f

Boykov et al. explained the details of the algorithm thoroughly in [21, 73, 20] and they also provide the source code on web which is also employed in our experiments.

Image segmentation based on MRF model is formulated as a **Belief Propagation**

Algorithm 2 Expansion Algorithm

Start with an arbitrary labeling f

Set success $= 0$

for each label $\alpha \subset L$ **do**

  Find $\hat{f} = argmin E(f')$ among f' within one α-expansion of f

  If $E(\hat{f}) < E(f)$, set $f = \hat{f}$ and $success = 1$

end for

If $success = 1$ go to 2

Return f

(a)Graph for finding the optimal Swap move

(b)Graph for finding the optimal Expansion move
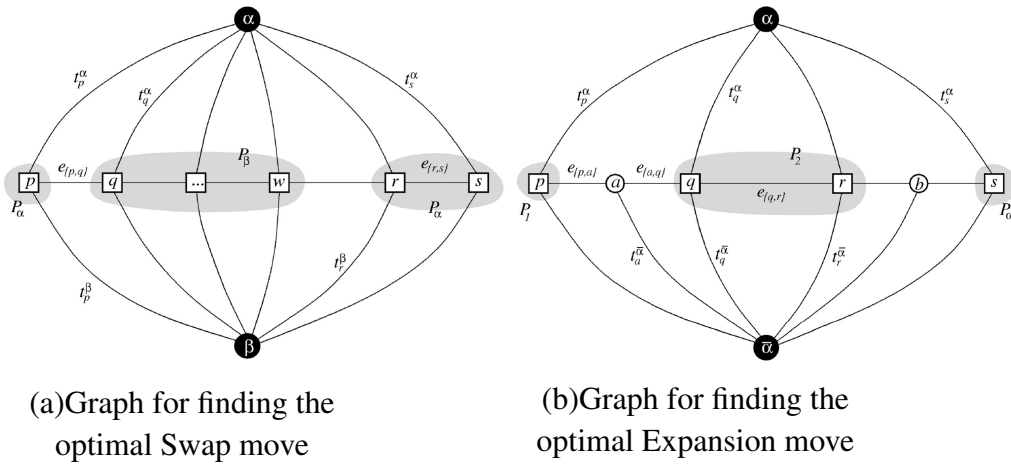
Figure 2.3: The graphs are dynamically constructed at each cycle of (a) Swap and (b) Expansion algorithm. (Taken from [21])
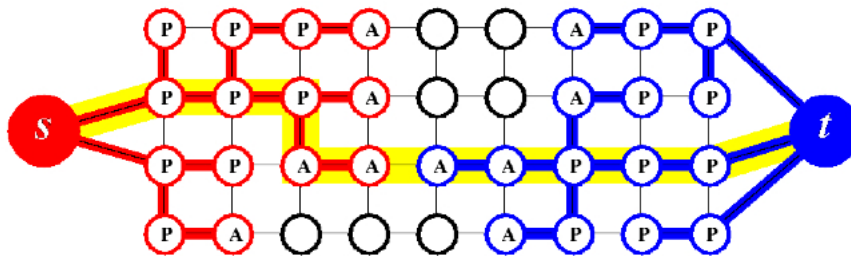
Figure 2.4: Example of search trees in the min-cut method of [20]. In this figure, s and t correspond to the terminal nodes α and $\bar{\alpha}$ in the Expansion method, and α and β in the Swap method. (Taken from [20])

(BP) problem, where messages among neighbors are computed iteratively and finally a belief vector is obtained for each site. Segmentation corresponds to the labeling which minimizes the individual belief vector at each node. Felzenswalb et al. proposed an efficient belief propagation method for minimizing MRF energy in early vision problems [44]. They propose modeling message updates over a bipartite graph which brings both memory and time gain. They also propose a hierarchical structure which reduce the number of message passing iterations.

In case of loopy graphs standard Belief Propagation may fail to converge, Kolmogorov and Wainwright propose a modified version of BP to avoid the problems caused by the loops [116, 72]. Their method, referred as **Tree-reweighted message passing**, converts a graph with loops into a spanning tree and then applies the standard BP on this tree.

Szeliski et al. analyze and compare the available energy minimization methods of graph cuts, loopy belief propagation and tree re-weighted message passing together with ICM experimentally. They compare the performance of these methods both interms of the solution quality and runtime. They point out that the modern energy minimization methods perform much better than the ICM method and they come very close to computing the global minimum. They run experiments on three benchmarks and they report that different algorithms perform better in different methods. The authors provide their codes online as mentioned in [108].

2.4.4.2   Conditional Random Fields

Conditional Random Fields model [76] is the discriminative form of Markov Random Fields. Let X be a random variable over the observed image data and F be the random variable over the labels. While MRF estimates the posterior $P(F|X)$, by first inferring the likelihood $P(X|F)$ and the prior $P(F)$, CRF directly infers $P(F|X)$. A given graphical model $G = (V, E)$ is a CRF, if, when conditioned on X, the random variable f_{s_i} obey the Markov property with respect to the graph. Hence, Markovianity in CRF is defined as follows:

$$P(f_{s_i}|X, f_{S-s_i}) = P(f_{s_i}|X, F_{N_{s_i}}). \tag{2.11}$$

28
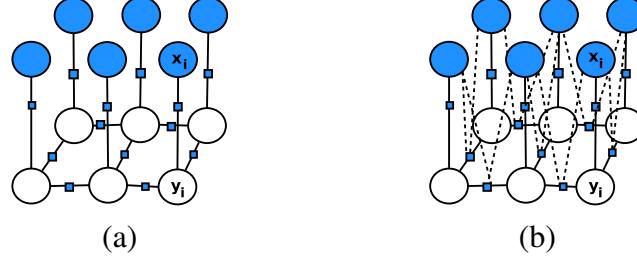
Figure 2.5: Graphical representation of a)MRF and b)CRF.

This implies that the label at each site is globally conditioned on the observation X. Graphical representation of CRF and MRF is shown in Figure 2.5. The corresponding energy function of the MRF model is as follows:

$$E(f;x) = \sum_{s \in S} \psi_s(f_{s_i}; x_{s_i}) + \sum_{s_j \in N_{s_i}} \psi_{s_i s_j}(f_{s_i}, f_{s_j}). \qquad (2.12)$$

And the corresponding energy function of CRF model is as follows:

$$E(f;x) = \sum_{s \in S} \psi_s(f_{s_i}; X) + \sum_{s_j \in N_{s_i}} \psi_{s_i s_j}(f_{s_i}, f_{s_j}; X). \qquad (2.13)$$

Shotton et al. propose a CRF based supervised segmentation method whose energy function consists of terms related to shape-texture, color, edge and location properties [105]. Shape-texture features are employed in the form of textons which are learned by a set of filter-banks. And they are boosted with appearance information and integrated into the CRF energy by a multi-class logistic classifier. Later, they proposed [104] semantic texton forests, which are ensembles of decision trees that act directly on image pixels. Semantic textons are more efficient than employing filter-banks to obtain textons.

Gould et al. propose a multi-class segmentation system which employs relative location information [54]. Relative location probability map is learned from the training data. In this map each superpixel casts a vote for where it would expect to find pixels of every class given its location and predicted label. Hence, in this approach an initial prediction is obtained by using appearance features by employing the Adaboost algorithm and then relative location probability map is utilized given the initial prediction. Gould et al. train their system both with a logistic regression function and with a CRF

model. CRF model is more advantageous compared to the logistic regression classifier since it also models pairwise relations.

In another study, Gould et al. propose a CRF model for decomposing a scene into geometrically and semantically meaningful regions [53]. For this purpose, they employ color and texture features of [105] and train a set of logistic classifiers for semantic classes and geometric classes separately. They train multi-class logistic classifiers for modeling both pairwise relations among semantic classes and pairwise relations among geometric classes. Location of the horizon is also modeled in their CRF function. In this method, model parameters for each CRF term are learned separately and the CRF energy is minimized iteratively; pixel-region association variables are updated, then, the model parameters are optimized and finally the energy is estimated. If the energy decreases, then the change is accepted and the process continues until the convergence.

Ladicky et al. [75] propose integrating co-occurrence statistics by using a higher order term in the conditional random field model for concurrent detection and segmentation. For this purpose, they introduce a higher order potential term which models association costs between labels. First, the model parameters are learned on a training set and then segmentation and detection processes are performed simultaneously by taking into account the label association costs.

2.5 Segmentation Fusion Systems

The segmentation fusion systems are analyzed in two groups. The first group is the segmentation systems which combine the top-down and the bottom-up approaches, while the second group is the segmentation systems which combine a set of initial segmentation outputs.

2.5.1 Combination of Bottom-Up and Top-Down Segmentation

In the recent years, researchers propose to combine the top-down and bottom-up segmentation methods, in order to overcome the limitations of top-down segmentation

and compensate for the lack of semantics in the bottom-up segmentation.

In order to obtain semantically meaningful segmentation Borenstein and Ullman proposed combining bottom-up and top-down segmentations [18] by defining a combination score function. As a bottom-up approach they employ a hierarchical segmentation method and obtain a hierarchy of homogeneous image regions at multiple scales. As top-down segmentation they learn shapes of certain objects from a training dataset. For this purpose, a fragment bank is constructed from a training dataset and fragments are labeled as figure or ground. Given a new image, a subset of the stored fragments are employed to find top-down segmentation. Results of the top-down and bottom-up segmentation is unified by means of maximizing a score function. The score function measures both the consistency between final segmentation and the result of top-down segmentation, and the consistency between parent-child labeling in the bottom-up segmentation hierarchy.

Kumar et. al. propose an object category specific MRF model, called OBJCUT, for simultaneous object recognition and segmentation [74]. Prior shape information is represented by pictorial structure [43], which is a probabilistic model for representing the objects as the collection of rigid parts and their relations. Information from pictorial structure is integrated to MRF energy via latent variables and segmentation is obtained by minimization of MRF energy via graph-cut.

Prasad et al. [96] propose a method which learns the class specific edges for object detection. Then, they integrate this information into the OBJCUT segmentation [74] by modifying the algorithm in two ways. The MRF energy of OBJCUT is modified to use only class specific edges and its boundary term is modified to encourage segmentation along high contrast regions.

Levin and Weiss propose to integrate both low level and high level cues into a CRF energy [79]. They suggest an energy function, which consists of two terms for top-down and bottom-up learning models. The term for top-down learning consists of a fragment mask on the image which is used to find the maximal correlation locations. The term for the bottom-up learning defines affinity weights such that the energy function is minimized if the labeling discontinuities align with the image discontinuities.

2.5.2 Consensus Segmentation Systems

For a given image, consensus segmentation systems combine a set of initial segmentation outputs to obtain a final segmentation. The multiple segmentation outputs can be obtained either by varying the parameter set of the same segmentation method or by running different segmentation algorithms.

The consensus segmentation systems can be categorized into two groups. The first group adapts the ensemble clustering methods while the second group employs graph methods to combine a set of initial segmentations.

2.5.2.1 Consensus Segmentation by Ensemble Clustering

Ensemble clustering methods are proposed to overcome the inherent problems of clustering such as the scale-invariance and the parameter selection [110]. Moreover, they have been studied commonly is the last decade [50, 117, 107].

Let $X = \{x_1, x_2, ..., x_N\}$ be a set of N objects, and $P = \{P_1, P_2, ..., P_M\}$ be a set of clusterings of X. Ensemble clustering is the process of finding a consensus clustering P^* which optimally represents the ensemble P. P^* may be directly selected from P or it may be a new clustering which satisfies some consensus criteria among the clusterings in P.

Since unsupervised segmentation bares the similar problems, ensemble clustering based consensus segmentation systems are proposed to overcome similar problems of image segmentation. General ensemble clustering methods that are employed to image segmentation are reviewed and evaluated by Franek et al. [46]. Here, ensemble clustering methods that are commonly adopted to image segmentation problem will be explained:

Best of K (BOK) selects the best or most representative partition among the initial partitions in P which maximizes (minimizes) a criterion function as follows:

$$P^* = \arg\min_P \sum_{i=1}^{M} \psi(P_i, P). \tag{2.14}$$

Here, if ψ is a distance measure, then the criterion function is minimized. If it is a similarity measure, then P which maximizes ψ is selected.

Best One Element Move (BOEM) [124] starts with an initial partitioning and iteratively test each label for each object x_n. Label of an object is updated if a pre-defined criterion function as in Eq. 2.14 is minimized (maximized). Zheng et al. proposed employing symmetric difference distance (sdd) [124]. Let π_i and π_j be two partitions, n_{00} is the number of pairs of elements co-clustered in both partitions, n_{11} is the number of pairs not co-clustered in either partition, n_{01} is the number of pairs of elements co-clustered in the first partition but not in the second and n_{10} is the number of pairs that are co-clustered in the second partition but not co-clustered in the first partition. They define symmetric difference distance as follows:

$$d(\pi_i, \pi_j) = n_{01} + n_{10}. \tag{2.15}$$

The advantage of this distance function is that it is computable in linear time. Nevertheless, other distance functions can be used as well.

2.5.2.2 Consensus Segmentation by Graphs

Graph based consensus segmentation systems initially constructs a region adjacency graph (RAG) whose nodes represent pixels (super-pixels) and edge weights represent similarity (dissimilarity) relations [25, 118]. Edge weights between two nodes are determined based on the outputs of initial segmentations. Generally, edge weights are formulated as the relative frequency of co-occurrence of two pixels (super-pixels).

Fred and Jain proposed constructing a co-association matrix whose entry (i, j) is the relative frequency of two pairs x_i and x_j occurring in the same cluster among the initial clusterings [47]. They employ this matrix as a similarity measure and apply hierarchical agglomerative clustering algorithm to obtain consensus partition.

Wattuya [118] et al. proposed employing Random Walker [55] to obtain a consensus segmentation. They initially, construct RAG whose weights are determined by Gaussian weighting:

$$w_{ij} = exp(-\beta \times (1 - \frac{n_{ij}}{M})) \tag{2.16}$$

where, n_{ij} is the number of initial segmentations that assign x_i and x_j into the same partition. In this way, edges correspond to similarity, where a high value of w_{ij} indicates that x_i and $x)j$ are very similar. Next, by means of removing a set of edges with low weight, they determine seed regions. And then, they employ Random Walks to get consensus segmentation.

Strehl and Ghosh [107] proposed hypergraph based consensus segmentation methods. In the hypergraph representation, the set of initial partitionings is represented as a hypergraph and each partition is represented by a hyperedge. Their Cluster-based Similarity Partitioning Algorithm constructs a $N \times N$ similarity matrix and employ graph partitioning algorithm METIS [65] to obtain consensus partition. The Hyper-Graph Partitioning Algorithm assumes that all hyperedges have the same weight, and formulate segmentation as a min-cut problem such that the hypergraph is cut in k connected components of approximately the same dimension. They use hypergraph partitioning package HMETIS [64]. The third algorithms of Strehl and Ghosh is the Meta-Clustering Algorithm, which considers partitions as vertices of a meta-graph. Edge weights, indicating similarity between vertices are determined by the Jaccard measure. This graph is partitioned via METIS [65] and finally each object is assigned to the most associated meta-cluster.

2.6 Benchmark Datasets for Performance Evaluation of the Segmentation Methods

Traditional bottom-up segmentation methods partition an image into the homogeneous regions, but do not provide labels for the extracted regions. On the other hand, semantic segmentation systems that employ top-down approaches assign class labels to all image pixels. Therefore, in order to evaluate the performance the bottom-up segmentation methods ground truth segmentations are sufficient, while pixel labels are required to evaluate performance of the top-down segmentation methods.

In 2001 Martin et al. constructed Berkeley Segmentation dataset which consists of 300 natural images. All images have $5 - 7$ groundtruth segmentations which are labeled by human subjects [84]. 200 images are used as training set and 100 images

are used for testing by supervised segmentation systems. Later, an additional set of images with their ground truth labels are added and currently number of images in this dataset is 500. Majority of the studies that compare segmentation or boundary detection systems and studies on segmentation evaluation criteria apply their system on the initial dataset with 300 images [114, 93, 113, 58, 122]. A benchmark together with the dataset is also available online [91].

MSRC object recognition dataset constructed by the Criminisi et al. [119] in 2004 consists of 591 natural images with ground truth segmentations. Images in this dataset belong to 23 classes. Since region labels are available in this dataset, it is employed by both segmentation [9] and labeling systems [105, 75, 121].

PASCAL VOC dataset is constructed by Everingham et al. in 2005 with four classes to be used in classification and detection problems. Later the dataset is expanded to 20 images and ground truth information for segmentation problems is included. As it provides ground truth for various problems, it is used in many studies in the literature [78, 83, 53].

2.7 Evaluation of Segmentation Systems

Evaluation of a segmentation system is a crucial task, because of many reasons: First of all there may be no ground-truth available to compare the result of a segmentation algorithm to a true segmentation. Secondly, mathematical distance metrics, such as the sum of squared error is not always consistent with our human perception of segmentation. Thirdly, segmentation results in a highly non-linear boundaries among the regions and it is not possible to define a measure of goodness for the output of segmentation.

In spite of the above mentioned difficulties, there are some probabilistic methods to measure the "quality" of a segmentation partition which will be summarized in the subsequent sections.

2.7.1 Mutual Information

In information theory, the mutual information is a measure of the mutual dependence of the two random variables. The mutual information of two discrete random variables X and Y can be defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{e \in X} p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right), \qquad (2.17)$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. In Pattern Recognition, Mutual Information is used for estimating similarity of two clusters. Wattuya et al. propose employing Normalized Mutual information to estimate segmentation performance, whose formula is as follows:

$$\phi^{(NMI)}(S_a, S_b) = \frac{\sum_{h=1}^{|S_a|} \sum_{l=1}^{|S_b|} |R_{h,l}| log \frac{n \times |R_{h,l}|}{|R_h| \times |R_l|}}{\sqrt{\sum_{h=1}^{|S_a|} |R_h| log \frac{|R_h|}{n} \sum_{l=1}^{|S_b|} |R_l| log \frac{R_l}{n}}}, \qquad (2.18)$$

where, S_a and S_b are two segments, n is the image size, R_h and R_l are regions from S_a and S_b, $R_{h,l}$ is the common part of the two regions. NMI takes values in the range $[0,1]$.

2.7.2 Variation of Information

Variation of information is a measure of distance between two clusterings. In case of evaluating segmentations, given a segmentation S and its ground truth segmentation S', variation of information (VOI) is estimated by *equation 2.19*

$$VOI(S, S') = H(S) + H(S') - 2I(S, S'), \qquad (2.19)$$

where, H and I represent entropy and mutual information respectively. Arbeleaz employ VOI and other measures like PRI and precision-recall to compare performance of various systems [9]. The application of this measure in the case of several ground truth segmentations is unclear and it is not widely used in segmentation evaluation.

2.7.3 Consistency Error

Martin et al. proposed a segmentation error measure which is based on estimating the level of consistency between two segmentations [84]. Given two segmentations, one of which may be a ground truth segmentation, consistency error measure is estimated by *Eq. 2.20*. In this equation, for each segment S_1 in the first segmentation and the segment S_2 in the second segmentation which contains the same pixel p_i are compared. Either there is a refinement relation between two segments, or there are overlapping pixels in two segments.

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \setminus R(S_2, p_i)|}{R(S_1, p_i)} \tag{2.20}$$

In Eq. 2.20, "\" is set difference operator which is non-symmetric. Using this local error measure, two error measures are defined first one is Global Consistency Error (GCE), which is provided in *equation 2.21*, and second measure is Local Consistency Error(LCE), which is given below:

$$GCE(S_1, S_2) = \frac{1}{n} min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\}, \tag{2.21}$$

$$LCE(S_1, S_2) = \frac{1}{n} \sum_i min \left\{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \right\}. \tag{2.22}$$

GCE and LCE take values in the range [0,1] where values close to 0 indicates high segmentation performance. There are only two misleading cases; first one occurs if image has only one region, in which case consistency error return 0 and the other one occurs if there are N regions, where all pixels are assigned to different regions, in which case again consistency error is 0. In other words, if there is a refinement relation between two regions then consistency error is $zero$. Consistency errors are informative if two segmentations have approximately the same number of regions.

2.7.4 Probabilistic Rand Index

Probabilistic Rand Index (PRI) takes pixels in pairs and measures the ratio of compatibly labeled pixels in segmentations S_{test} and ground truth segmentations S_k as

follows:

$$PR(S_{test}, \{S_k\}) = \frac{1}{\binom{N}{2}} \sum_{i<j} [c_{ij}p_{ij} + (1 - c_{ij})(1 - p_{ij})], \qquad (2.23)$$

where, N is the number of pixels in image, c_{ij} is the event of a pair of pixels i and j having the same label in image S_{test} and p_{ij} is the ground truth probability of two pixels having the same label estimated over all ground truth segmentations of the image. PRI takes values in the interval [0,1], where values close to 1 indicates high segmentation performance. PRI is meaningful even if two segments have different number of regions.

Unnikrishnan et al. proposed Normalized Probabilistic Rand Index [113] where the expected value for p_{ij}, which is the expected value of two pixels having the same label estimated over all the ground-truth segmentations, as follows:

$$Normalized\ index = \frac{Index - Expected\ index}{Maximum\ index - Expected\ index}. \qquad (2.24)$$

PRI can be employed with more than one ground truth segmentations and it gives meaningful results even if the segmentations have different number of regions. For this reason, it is often used in evaluation of segmentation systems [93, 9].

2.7.5 Precision, Recall and F-score

Precision and recall are boundary based segmentation evaluation measures. Given a segmentation in the form of boundaries and corresponding ground truth, precision estimates the fraction of true positives; that is the fraction of pixels that are labeled as contour which are also labeled as contour in the ground truth segmentation. Recall is defined as the fraction of ground truth contour pixels that are labeled as contour in the segmentation result. F-score is the harmonic mean of precision and recall, which is defines as,

$$F = 2 \times \frac{precision \times recall}{precision + recall}. \qquad (2.25)$$

38

Precision and recall measures are extensively used in the classification and recognition literature. They are, also, used for evaluating segmentation systems, especially, for boundary based methods [41, 9].

2.8   Chapter Summary

In this chapter, an overview of the image segmentation systems is provided with respect to the level and type of the available information about the data sets. Image segmentation systems are categorized based on the methodologies to integrate the information into the segmentation process. In this respect, the segmentation studies are categorized as top-down, bottom-up and fusion based systems and various methods in these categories are reviewed. Later, the segmentation bechmark datasets are explained. The chapter is concluded by reviewing the performance measures for image segmentation.

CHAPTER 3

SEGMENTATION FUSION UNDER UNSUPERVISED MARKOV RANDOM FIELDS

In the previous chapter, the application of the MRF model on both the supervised and unsupervised segmentation problems are reviewed [120, 77, 92, 51, 68, 69, 30, 70, 45, 54, 105, 75]. It is monitored that the MRF energy function is very convenient for integrating available information in a quantitative way into the segmentation process. In this chapter, it is assumed that for a given segmentation problem, no prior information and no labeled dataset is available. Hence, only the bottom-up methods are employed in this chapter, and an unsupervised segmentation method which utilizes the MRF energy for fusing a set of segmentation outputs is proposed.

3.1 An Unsupervised MRF Model For Segmentation Fusion:Boosted-MRF

In this thesis, the MRF model is utilized to construct a segmentation fusion system, called *Boosted-MRF*. In this system, information fusion is realized by re-defining the pairwise potentials of the MRF energy function to model the consensus among a set of segmentation outputs.

The Boosted-MRF segmentation system consists of four major modules as provided in Figure 3.1. First a region adjacency graph(RAG) is formed in the RAG Construction module to represent the image. Then, the edge weights of the RAG is determined in the Segmentation Fusion module to achieve the consensus among a set of segmentation outputs, and they are integrated into the system in the RAG Update module. Finally, MRF energy of the RAG is minimized in the Boosted-MRF Energy Mini-
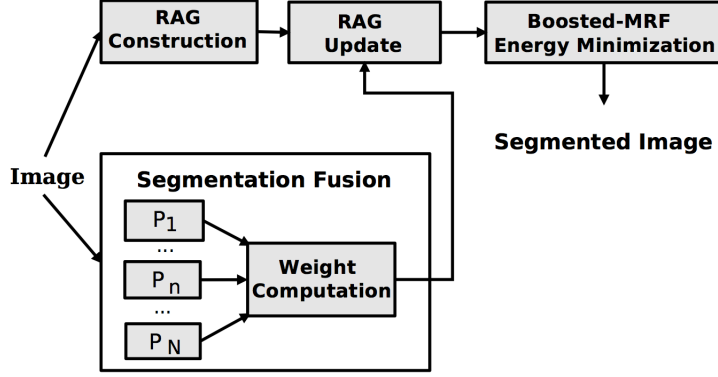
Figure 3.1: System Architecture of Boosted-MRF

mization module to obtain a final segmentation.

3.1.1 RAG Construction

The image is first over segmented and the set of super-pixels $S = \{s_i\}$ is obtained. Then, this set is used to form the Region Adjacency Graph (RAG) $G = (V, E)$. The nodes of RAG, V, are the super-pixels S and first order neighborhood system is applied for determining the edges, E. Thus, the nodes corresponding to a pair of super-pixels, s_i and s_j, which share a boundary are connected by an edge with weight $w_{ij} = 1$.

3.1.2 Segmentation Fusion and Boosted RAG Construction

In the Segmentation Fusion module, for a given image, an ensemble $\lambda = \{P_1, ..., P_N\}$ of N segmentations is obtained. The segmentation maps can be extracted by either at the output of different segmentation methods or by varying the parameter settings of the same method. Then, at each segmentation P_n, each super-pixel of the RAG is labeled by the index of the segment that it belongs. During the labeling process, the label of a super-pixel is determined by the label of the most overlapping region. A sample image and labeling of its super pixels is depicted in Figure 3.2.

The edge weight between a pair of adjacent super-pixels is updated based on the
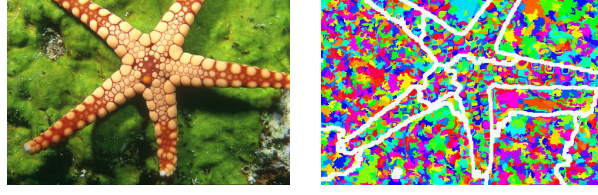
Figure 3.2: Labeling Process of super pixels: Each super pixel is assigned to the label of the maximum overlapping region with that super pixel. First image is a sample image from Berkeley Segmentation Benchmark [84]. Super pixels are depicted in colors while the boundaries of a segmentation is depicted in white.

relative frequency of co-occurrence of this pair in the ensemble of segmentations. The relative frequency of co-occurrence is estimated by counting the number of co-occurrence if this super-pixel pair which falls into the same segment in all of the segmentation maps, which is estimated with the following equation:

$$f(s_i, s_j) = \frac{1}{N} \sum_{n=1}^{N} \delta(l_{n_{s_i}}, l_{n_{s_j}}), \tag{3.1}$$

where δ is the Kronecker delta function, $l_{n_{s_i}}$ is the label of super pixel s_i in the initial segmentation P_n. Finally, the edge weights in the initial RAG is updated to obtain the *Boosted-RAG* as follows,

$$w_{ij} = \alpha \times f(s_i, s_j), \tag{3.2}$$

where α is a free parameter of the proposed system that is used to scale edge weights in the range $[0, \alpha]$.

Note that, the edges in RAG are updated by estimating the relative frequency of the co-occurrence of neighboring super pixels having the same label over N segmentations. Therefore, when there is a high frequency of co-occurrences, this corresponds to a high consensus among the segmentation methods, whereas, low frequency of co-occurrences indicate conflicting results of different segmentation methods. Incorporating the degree of consensus into the pair-wise potentials emphasizes the regions with high consensus, while suppressing the conflicting results. Based upon the updated weights, a new RAG is obtained whose edge weights posses information from ensemble of segmentations, hence this new RAG is referred as Boosted-RAG.

3.1.3 MRF Energy Minimization

Final segmentation is obtained by minimizing the following energy function:

$$E(x) = \sum_{s_i \in S} \psi_i(l_{s_i}) + \sum_{s_i \in S, s_j \in N_{s_i}} w_{ij}, \times \psi_{ij}(l_{s_i}, l_{s_j}), \tag{3.3}$$

where the first term, singleton potential, is modeled by assuming that the image features are drawn from Gaussian Distribution and color features are employed for this term. The second term, pairwise potential is modeled as the weighted Potts model, where the weight function is estimated by equation 3.2 and ψ is the Potts model.

Steps of the algorithm are given in Algorithm 4. First, a set of baseline segmentations, $\{P_1, ..., P_N\}$ is obtained. Then, a set of super pixels, S, are obtained by over-segmentation, and RAG is constructed. The first baseline segmentation, P_1, is employed to estimate model parameters, $(\mu_{l_i}, \Sigma_{l_i})$ for each region index $l_i \in L$. After that, equation 3.2 is utilized to obtain Boosted-RAG. Finally, MRF energy minimization starts at line 8 of the algorithm by initializing a random labeling and minimizing MRF energy by α expansions of [21] at lines $10 - 13$.

3.2 Complexity Analysis

The complexity of the proposed segmentation fusion system is determined by the MRF energy minimization method. The steps of the Expansion method are the lines 9-14. The lines between 9 and 14 is a cycle of the Expansion method. At each cycle the algorithm iterates $|L|$ times, one for each label $l_i \in \{L\}$. The worst case complexity of line 11 is reported as $O(mn^2|C|)$ in [20] where $|C|$ is the cost of the minimum cut, n is the number of nodes and m is the number of edges in the graph. The expansion algorithm is guaranteed to terminate in $O(|P|)$ cycles, where P is the number of the image pixels. In practice, the algorithm terminates in a few cycles. So, the computational complexity of Boosted-MRF is $O(|L|mn^2|C|)$.

On the other hand, the complexity of the BOK method depends on the distance metric employed. If Symmetric Difference Distance (SDD), as explained in Section 2.5.2.1,

---
Algorithm 3 Steps of Boosted-MRF

1: Obtain all $\{P_1, ..., P_N\}$

2: Obtain super pixels, S,

3: Construct RAG for S

4: **for** each region index l_i **do**

5:     estimate model parameters (μ, Σ) from P_1

6: **end for**

7: Estimate edge weights by eq. 3.2 to obtain Boosted-RAG

8: Initialize a random labeling l

9: $success = 0$

10: **for** each label $\alpha \in L$ **do**

11:     find $\hat{l}$=arg min $E(l')$ among l' within one α expansion of l

12:     if $E(\hat{l}) < E(l)$, set $l = \hat{l}$ and $success = 1$

13: **end for**

14: if $success = 1$ go to 10

15: return l

---

is employed, then the complexity of the BOK method is $O(K^2 MN)$, where K is the number of initial segmentations, and $M \times N$ is the number of image pixels.

In practice the running times of BOK is only slightly better than the Boosted-MRF method. This is mainly because, the Boosted-MRF constructs the RAG over super-pixels while the BOK method estimates SDD over the pixels.

3.3 Chapter Summary

In this chapter, a segmentation fusion system based on the MRF model is proposed. The proposed system called the Boosted-MRF, first obtains a set of segmentation outputs by a set of bottom-up segmentation methods. Then, these outputs are used to estimate the relative frequency of co-occurrence for the neighboring super-pixels which have the same label in the ensemble. This information is incorporated in the smoothness term of the MRF energy. Finally, segmentation output is obtained by minimizing the MRF energy by the graph-cut based energy minimization method,

expansion.

CHAPTER 4

A DOMAIN SPECIFIC IMAGE SEGMENTATION SYSTEM WITH MARKOV RANDOM FIELDS

In the previous chapter, it is assumed that no prior information for a given segmentation problem is available and an unsupervised segmentation scheme is proposed. In this chapter, it is assumed that, prior information about the segmentation problem is available. The prior information, which is referred as domain specific information, is not in the form of a labeled dataset, but provides more general information for the given segmentation problem. The MRF model is utilized to incorporate various prior contextual information or constraints in a quantitative way into the segmentation process. In this chapter, a new segmentation scheme, which is referred as domain specific segmentation, is introduced. The chapter starts with the explanation of the domain specific information. Then a new segmentation fusion method which is enhanced with the domain specific information is introduced.

4.1   Domain Specific Information

Domain Specific Information (DSI) is prior information for a given problem which is related to the characteristic properties of the problem domain. The characteristic properties may be related to the appearance, shape(s), shape(s)' relations, location, object(s) or object(s)' relations. The most critical part of the suggested segmentation fusion method is to formally define and integrate the domain specific information into the segmentation process.

47

4.1.1 Information Utilization in Segmentation Problem

In the image processing literature, the level of information is somehow quantified between high and low, where the low level information refers to the image features and high level information refers to pixel labels corresponding generally to semantic classes. There is a large number of studies dealing with the problem of relating low level and high level information in the literature. This problem is called the *semantic gap* problem and general attitude in bridging this gap is to train a segmentation system by introducing the semantic labels corresponding to low level image features. For this purpose, top-down and bottom-up approaches are combined and segmentation and recognition tasks are taken up cooperatively as described in Chapter 2.

The information that is integrated into the segmentation process varies enormously. At one end, a labeled training set is employed to learn model parameters of a segmentation method. At the other end, unsupervised approach is taken and no prior information is employed. In between these two approaches, there is another group of segmentation studies which employ prior information. The prior information is usually related to shape, appearance or location, and the studies generally employ one or two types of prior information. Nevertheless, depending on the problem domain, prior information shows great diversity, and for a given problem more than two types of prior information may be available. Moreover, higher level information than shape, appearance and locality may be available. For example, if expert knowledge for a given problem provides information about the existence of certain objects and their relative positions, corresponding object detectors can be utilized and neighborhood relations in the MRF energy function can be defined such that the segmentation process is encouraged to segment the related regions as neighbors.

In this study, an unsupervised segmentation fusion method, which employs the MRF model for utilizing information from a collection of DSI simultaneously to guide the segmentation process, is proposed.

4.1.2 Types of DSI

DSI arises in various forms in the real world problems. Depending on the available information, DSI may be related to a whole image or part of an image. Forms of DSI can be categorized into four groups as follows:

- **DSI Related to Low Level Region Features:** DSI related to color, texture, intensity information is in this group. Since, this information is directly related to the low level image features, it is also referred as the low level DSI. Expert knowledge regarding illumination conditions or surface reflectance can be considered in this group. Similarly, information related to the colors of objects or color of the image background is an example of DSI in this group.

- **DSI Related to Primitive Shapes:** Prior information stating that certain shapes are expected in a given image is in this category. Specific shapes are extracted by shape detectors to segment related image parts. This group of information may be considered as mid level DSI. It is one step closer to the high level information compared to the low level DSI.

- **DSI Related to Object(s) in the Image:** Prior information related to existence of the objects in a given image is in this group. In order to represent this type of information, domain expert provides a higher level information about the existence of the object(s) in the given problem domain and object detectors are employed to segment corresponding image parts. This type of DSI is referred as the high level DSI.

- **DSI Related to Spatial Constraints:** DSI related to location and relative position are in this group. If DSI is related to location, then it may be employed in specific problems where segmentation with locally varying scales is required. This information is utilized to obtain segmentation outputs whose scale spatially changes. If DSI is related to relative positions of shapes or objects in the image, then DSI is utilized to define neighborhood relations. DSI related to location information is considered as low-level DSI, while DSI related to pair-wise relations is considered as high-level DSI.

49

4.1.3 Formalization of Domain Specific Information

Let $DS = \{ds_m\}$ be a set of DSI related to a given problem domain. Each ds_m is a DSI in one of the four forms, as described in the previous section. Each DSI related to low level region features, primitive shapes and objects are utilized separately to obtain an incomplete partition of the image. This partition is referred as *domain specific map*, DSM_m .

Initially, each ds_m is represented by a predicate as follows:

$$P(\cup s_i) = \begin{cases} TRUE & \text{if } \forall s_i \text{ satisfies property } \wp \\ FALSE & otherwise \end{cases} \tag{4.1}$$

where, the property $\wp$ poses DSI about either low level region features, primitive shapes or objects and $S = \{s_i\}$ is the set of image pixels(super-pixels). If DSI is related to the low level region features then then the definition of the predicate is straightforward. If DSI is related to primitive shapes or objects, then the property is defined as the response of image pixels to a shape or object detector. In this way, two regions can be obtained as satisfying or unsatisfying the property $\wp$. A set of predicates may be utilized simultaneously to segment an image into more than two regions. In that case, a set of predicates $\{P_m\}$ are applied on the image to obtain the corresponding DSM, as follows:

$$DSM_m = \{\{r_t, l_m\}|\ r_t = (\cup s_i)\ s.t.\ P_m(\cup s_i) = TRUE, \forall r_t \in I\}, \tag{4.2}$$

where, l_m is the label assigned to the set of super-pixels r_t which satisfy the predicate P_m. Once, the DSMs are obtained, DSI related to spatial constraints are utilized to integrate information gathered from the DSMs into the segmentation process, as will be explained in section 4.2.1.2.

4.1.4 Constraints and Flexibilities

DSI refers to a wide range of prior information and it can be applied to broad areas of image processing. However, the critical issue in the application of DSI on various

areas is the sound mathematical definition of DSI, so that corresponding DSM can be gathered. A DSM is an incomplete partitioning of the image where certain parts of the image is labeled as having specific characteristic which is defined by a predicate. For example a predicate may be defined as "The super-pixel belongs to a face." and a corresponding DSM can be obtained by partitioning the image into two regions where the first region consists of super-pixels which satisfy the predicate while the second region consists of super-pixels which do not satisfy the predicate. This example points out that the available DSI should be represented as a predicate which can be employed to obtain a DSM. For example, prior information indicating that 'A given dataset consists of funny pictures.' may not be represented mathematically, hence it may not be employed as DSI. On the other hand, prior information related to the intensity, texture, shape, location and object can all be formulated mathematically and corresponding DSMs can be obtained.

4.2 Domain Specific Image Segmentation with Markov Random Fields

Domain specific image segmentation is a new scheme in image segmentation which is based on the incorporation of available domain information into the segmentation process. Integration of domain information into a segmentation/classification process is not novel. However, studies with domain information either utilize one or two of the shape, appearance and location information during segmentation [37, 38, 39, 56, 57], or they employ DSI in a pre-processing or a post-processing step to eliminate the regions with certain characteristics before feeding the segmentation output to a classifier [101, 100]. In the latter case, sometimes certain bias is introduced into the system since the regions are directly eliminated, which in the further processes causes decrease in the system performance. On the other hand, in this study a domain aware segmentation is proposed by constraining the segmentation process using DSI such that the segmentation output can be directly fed to a classifier.

In this chapter, a novel segmentation system, which integrates domain specific information in various forms into the segmentation process, is introduced. In this system, domain information is introduced in a fuzzy form and the steps of the segmentation process is adjusted to extract regions which comply with the DSI. In order to inte-

grate the DSI into the segmentation process in a fuzzy form, MRF is employed, and an MRF based domain specific image segmentation system, DS-MRF, is introduced. The proposed system is an unsupervised segmentation system which integrates semantic information into the segmentation process by means of available DSI. DS-MRF, not only integrates DSI but it also fuses information from the output of a set of segmentation methods. So, DS-MRF performs information fusion in two perspectives:

- *Bottom-up Information Fusion* is realized by the fusion of information from the output of a set of bottom-up segmentation methods.

- *Top-down Information Fusion* is realized by the fusion of information from a set of DSMs which are obtained by employing available DSI.

4.2.1 System Architecture of the DS-MRF Model

The system architecture of the DS-MRF method is depicted in the block diagram of Figure 4.1. As it is seen from the figure, the image is first over-segmented to create a Region Adjacency Graph. Next, a set of bottom up segmentation (BUS) maps and domain specific maps (DSM) are created. Information from BUS and DSM are used to update RAG as domain specific (DS-RAG). Finally, the MRF energy function of the domain specific RAG is minimized to obtain the segmented image.

4.2.1.1 RAG Construction

A bottom-up segmentation method from the literature is employed to over-segment the image. A region adjacency graph(RAG) is obtained. Each segment (super-pixel) forms a node of the RAG. Super-pixels having a boundary are connected by an arc with weight $w_{ij} = 1$ in the RAG.

4.2.1.2 Segmentation Fusion for Domain Specific RAG Construction

Segmentation fusion module consists of two submodules:
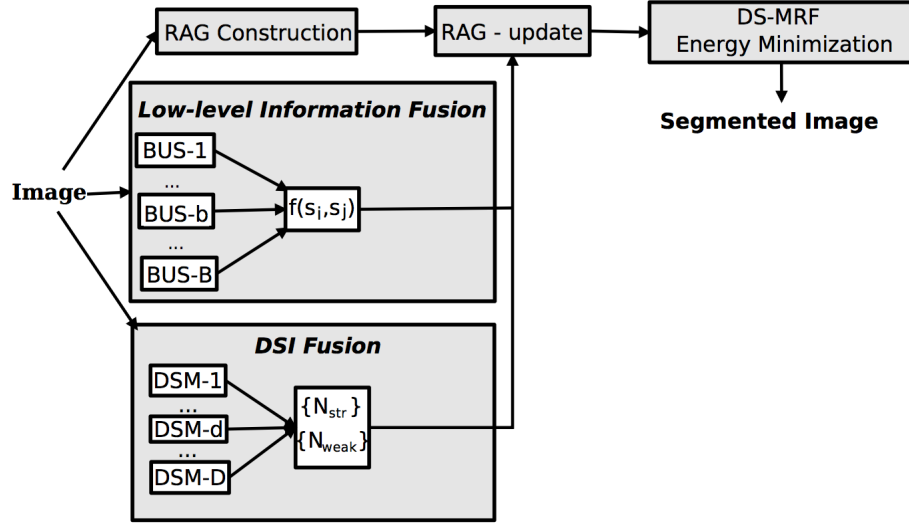
Figure 4.1: System Architecture of DS-MRF.

- *Low Level Information Fusion*

- *DSI Fusion*

which operate cooperatively to update the RAG as domain specific RAG (DS-RAG). Low Level Information Fusion module employs a set of bottom-up segmentations(BUS) and estimates the relative frequency of co-occurrence of each pair of super-pixels in the same region over the outputs of these bottom-up segmentations. While, the DSI Fusion module generates DSMs as described in section 4.1.3 and two sets of DSM pairs are defined; $DSM_{connect}$ and $DSM_{disconnect}$. The $DSM_{connect}$ set consists of pairs of DSMs which indicate constraints that are utilized for connecting neighboring regions, while the $DSM_{disconnect}$ set consists of pairs of DSMs which indicate constraints that are utilized for dis-connecting a pair of neighboring regions. The Segmentation Fusion module defines three types of neighborhood relations using these maps and the set of DSM pairs, and updates the unit weights of the RAG to generate a domain specific RAG:

- **Standard Neighborhood** is defined between two spatially adjacent super-pixels which are not labeled as having a domain specific property by the Domain Specific Maps. If one of the neighboring super-pixels is labeled by some DSM while the other neighbor is not labeled by any DSM, then they are considered

as standard neighbors. The set of standard neighbors is represented as N_{std}. The edge weights in N_{str} are kept as $w_{ij} = 1$.

- **Strong Neighborhood** is defined between two adjacent super-pixels, which are labeled by a pair of DSM pairs in the set of $DSM_{connect}$. The set of strong neighbors is represented as N_{str}. The edge weights between the strong neighbors are updated by the relative frequency of s_i and s_j having the same label in the set of BUS maps and it is computed as,

$$f(s_i, s_j) = \frac{1}{B} \sum_{b=1}^{B} \delta(l_{b_{s_i}}, l_{b_{s_j}}), \tag{4.3}$$

where δ is the Kronecker delta function which is equal to 1 when s_i and s_j have the same labels, $l_{b_{s_i}}$ and $l_{j_{s_j}}$ in the bottom-up segmentation map BUS_b. In this way, the super-pixels with strong neighborhood relations are encouraged strongly to have similar labels.

- **Weak Neighborhood** is defined if DSI related to spatial relations which indicates relative positions between shapes and objects is available. Two adjacent super-pixels, which are labeled by a pair of DSMs in the set of $DSM_{disconnect}$ are defined as weak neighbors. Definition of weak neighborhood enables the system to integrate shape relations and object relations into the segmentation process. Certain shapes or objects known to appear as neighbors are defined by corresponding DSMs and they are defined as weak neighbors. The set of weak neighbors is represented as N_{weak} and the weight w_{ij} is set as $\beta \leq 0$.

Note that, the above information fusion procedure creates three types of neighborhood relations for the domain specific RAG; $\{N_{std}\}, \{N_{str}\}$ and $\{N_{weak}\}$.

4.2.1.3 Domain Specific Energy Minimization

Recall from the Eq. 2.8 that there are two terms in the MRF energy function: The first term is defined as,

$$U_{data} = \sum_{s_i \in S} \psi_i(l_{s_i}) \tag{4.4}$$

and it is modeled as negative log likelihood using color features.

The second term is the smoothness term which is defined as the weighted Potts model. The weight function for a pair of super-pixels s_i and s_j is set as follows:

$$w_{ij} = \begin{cases} f(s_i, s_j) & \text{if } (s_i, s_j) \in N_{str} \\ \beta & \text{if } (s_i, s_j) \in N_{weak} \\ 1 & otherwise \end{cases} \qquad (4.5)$$

where, s_i represents super-pixels as MRF sites. β is a constant such that $\beta \leq 0$. The first condition of the U_{smooth} function estimates the weight of pairwise energy between a pair of strong neighbors, while the second condition estimates the weight of energy for a pair of weak neighbors and the third condition estimates the weight of energy for a pair of standard neighbors. So, the smoothness term of the MRF energy function is estimated with the following formula:

$$U_{smooth} = \alpha \times (\sum_{s_i \in S, s_j \in N_{str_{s_i}}} f(s_i, s_j) \times \psi_{ij}(l_{s_i}, l_{s_j}) + \sum_{s_i \in S, s_j \in N_{weak_{s_i}}} \beta \times \psi_{ij}(l_{s_i}, l_{s_j}) + \sum_{s_i \in S, s_j \in N_{std_{s_i}}} \psi_{ij}(l_{s_i}, l_{s_j}))$$

$$(4.6)$$

where, α is a constant which scales the pairwise potential weights in the range $[0, \alpha]$ and ψ is the Potts model. Finally, the energy function is minimized by the graph-cut based method of Boykov et al. [21]. They formulate the energy function for a configuration $(l_{s_1}, ..., l_{s_n})$ as follows:

$$U(l_{s_1}, ..., l_{s_n}) = \sum_{s_i} U^{s_i}(l_{s_i}) + \sum_{s_j < s_i} U^{s_i, s_j}(l_{s_i}, l_{s_j}), \qquad (4.7)$$

where $U^{s_i}(l_{s_i})$ is the data term for a pixel (super-pixel) s_i which is assigned to label l_{s_i} and $U^{s_i, s_j}(l_{s_i}, l_{s_j})$ is the smoothness term defined for neighboring pixels (super-pixels) s_i and s_j whose labels are l_{s_i} and l_{s_j} respectively. Kolmogorov and Zabih [73] state that the necessary and sufficient condition for the energy function for graph representability is regularity, and this energy function is regular if and only if each term of the pairwise potential satisfies the following condition:

$$U^{s_i, s_j}(0, 0) + U^{s_i, s_j}(1, 1) \leq U^{s_i, s_j}(0, 1) + U^{s_i, s_j}(1, 0). \qquad (4.8)$$

In our energy function, the strong neighborhood relation and the standard neighborhood relation satisfy the regularity condition while the weak neighborhood relation

55

satisfies the condition only if the β variable is set as 0. For negative values of β, the expansion algorithm is not guaranteed to find the global optimum. On the other hand, the data term is usually sufficient for ensuring that the different labels are assigned to a pair of weak neighbors. Hence, β is set as 0, and the MRF energy function of DS-MRF is minimized by the Expansion algorithm of Boykov et al.

The steps of the DS-MRF method is provided in Algorithm 4.

Algorithm 4 The Steps of DS-MRF

1: Obtain all $\{BUS_i\}$

2: Obtain DSM(s) by utilizing DSI

3: Obtain superpixels, S,

4: Construct RAG for S

5: **for** each region index $l_i \in \{L\}$ **do**

6:     estimate model parameters (μ,Σ) from BUS_1

7: **end for**

8: Obtain DS-RAG from RAG

9: Initialize a random labeling l

10: $success = 0$

11: **for** each label $l_i \in \{L\}$ **do**

12:     find $\hat{l} = argmin E(l')$ among l' within one α expansion of l

13:     if $E(\hat{l}) < E(l)$, set $l = \hat{l}$ and $success = 1$

14: **end for**

15: if $success = 1$ go to 10

16: return l

4.2.2 Complexity Analysis

The computational complexity of the DS-MRF method is determined by the energy minimization steps, which correspond to lines $10 - 15$ in Algorithm 4. Lines $10 - 15$ is a cycle of the Swap/Expansion algorithm and Lines $11 - 14$ is an iteration. The computational complexity of the Expansion algorithms is explained in the Chapter 3 as $O(|L|mn^2|C|)$, where L is the number of labels $l_i \in \{L\}$, n is the number of nodes, m is the number of edges and C is the cost of the minimum cut. On the other hand, the

computational complexity of the Swap algorithm is $O(|L|^2 mn^2 |C|)$, since the number of cycles depends on the number of $\alpha - \beta$ pairs. Swap and expansion algorithms are very efficient compared to the classic methods Simulated Annealing(SA) and Iterated Conditional Moves(ICM). This is mainly because, swap and expansion algorithms make large moves by changing the labels of a set of pixels at each step, while SA and ICM make simple moves by changing the label of a single pixel at each step.

4.3 Chapter Summary

In this chapter, domain specific information is introduced and its types are explained. Constraints and the flexibilities of domain specific information is discussed. An MRF based domain specific system, DS-MRF, is introduced. Formulation of the segmentation problem in the domain specific perspective is explained and the domain specific energy function is optimized to obtain a partition.

CHAPTER 5

EXPERIMENTS

5.1 Experiments on the Boosted-MRF Method

The performance of the Boosted-MRF method is evaluated on a commonly used segmentation benchmark. The behavior of the method on different settings is monitored and the observations are provided.

5.1.1 Dataset

The proposed system is employed on the Berkeley Segmentation Dataset and Benchmark [84], which is widely used in image segmentation literature [118]. The dataset consists of 300 color images and each image has 5-7 ground truth segmentations.

5.1.1.1 First Experimental Setup

In the first experimental setup, the Efficient Graph Based Segmentation [42] is employed as baseline segmentation method. The algorithm has three parameters: smoothing parameter (σ), threshold function (k) which determines the scale of the segmentation where larger values of k results in larger regions and minimum area size ($minArea$). These parameters are set based on the study of Wattuya et al. [118]. Hence, 24 baseline segmentations are obtained for the following parameter settings; $\sigma = \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, $k = \{150, 300, 500, 700\}$ and $minArea = 1500$. Super pixels are obtained by Mean Shift segmentation [28] by setting parameters

(a)   (b)

Figure 5.1: Performance comparison of the Boosted-MRF method and the baseline EG segmentation methods by PRI and GCE: Performance of baseline methods are plotted in green, performance of the Boosted-MRF method plotted in red. Images are ranked in (a) ascending PRI, (b) ascending GCE values for each method.

$h_r = 2$, $h_s = 2$, $minArea = 25$, so that the super-pixels are very small. And initial segmentation is obtained by Mean Shift segmentation for $h_r = 2$, $h_s = 2$, $minArea = 500$. The sensitivity of the method to the initial segmentation is analyzed in subsequent sections.

The only parameter of the proposed method, α, is set as $\{2, 5, 8\}$. For each image, segmentation result which maximizes the ANMI over 24 initial segmentations is selected as the output segmentation.

The contribution of the proposed system is observed by comparing its performance with the performances of the baseline segmentation methods. For this purpose, Probabilistic Rand Index(PRI) and Global Consistency Error (GCE) of all images are sorted in the ascending order for each segmentation system and displayed in Figure 5.1. In this figure, the performance of the EG segmentations are plotted in green while the performance of the Boosted-MRF segmentation is plotted in red. Performance increase achieved by the Boosted-MRF segmentation is observed both by the increase in PRI values and by the decrease in GCE values in these graphs.

Similarly, Normalized Mutual Information (NMI) for each image is estimated for both the proposed system and for the baseline segmentation systems. The sorted NMI for 300 images are obtained for each baseline segmentation and for the Boosted-MRF segmentation. The ranked images vs. NMI graph is given in Figure 5.2 (a).

60

(a)   (b)
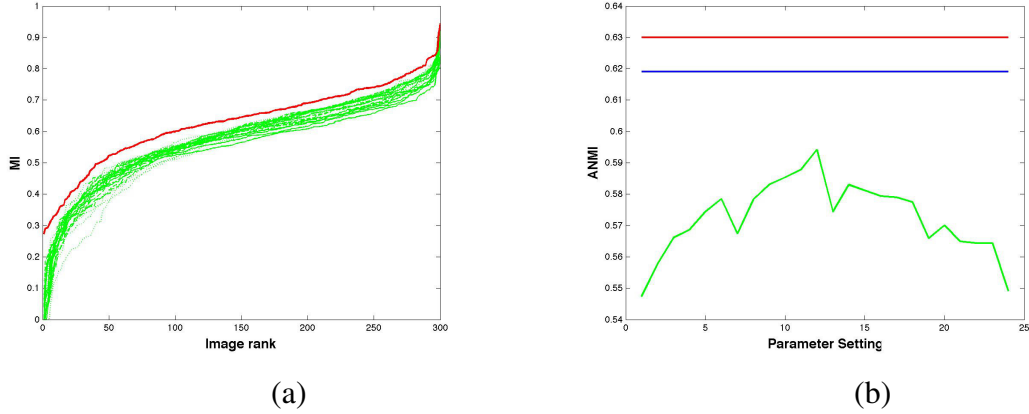
Figure 5.2: Performance comparison of the Boosted-MRF method and the baseline EG segmentation methods by MI and ANMI: Performance of the baseline methods are plotted in green, performance of the Boosted-MRF method plotted in red. (a) MI vs. image-rank for baseline segmentations and the Boosted-MRF segmentation. (b) parameter settings (each parameter setting corresponds to a distinct baseline segmentation method) vs. ANMI. In this graph, performance of the Random Walker Based segmentation fusion [118] in plotted in blue.

As a final comparison in this setup, performance of the proposed system is compared to the Random Walker based combination of segmentations [118]. For this purpose, Average Normalized Mutual Information (ANMI) is estimated over 300 images and the results are as depicted in Figure 5.2 (b). In this graph, the performances of 24 baseline segmentations are plotted in green, the performance of Random Walker Based system is plotted in blue and the performance of the Boosted-MRF segmentation is plotted in red.

5.1.1.2   Second Experimental Setup

In the second experimental setup, Mean Shift segmentation [28] is employed as baseline segmentation method. Its range and spatial bandwidth parameters are set as $\{2, 3, 4, 5, 6\}$ and $minArea$ is again set as 1500. Hence, 25 baseline segmentations are employed. The super-pixels and the initial segmentation parameters are set as in the first experimental setup. α is again set as $\{2, 5, 8\}$ and for each image, segmentation result which maximizes ANMI over 25 initial segmentations is selected as the output segmentation.
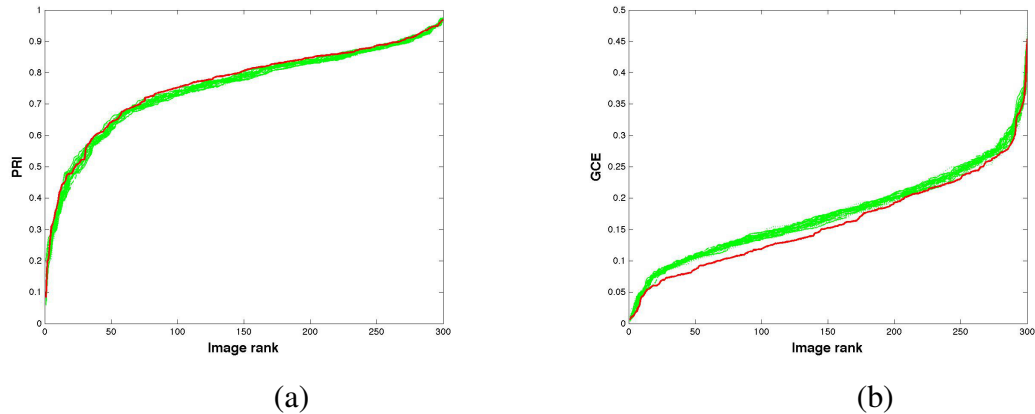
Figure 5.3: Performance comparison of the Boosted-MRF method and the MS baseline segmentation methods by PRI and GCE: Performance of baseline segmentation methods are plotted in green, performance of the Boosted-MRF method plotted in red. Images are ranked in (a) ascending PRI, (b) ascending GCE values for each method.

.

In this experiment, performance of the Boosted-MRF segmentation is compared to the performance of baseline segmentations and to the performance of classic MRF model, which assigns $w_{ij} = 1$ for all neighbors. ANMI is estimated for the Boosted-MRF segmentation, for each baseline Mean Shift segmentation and for the Classic MRF segmentation over 300 images and the results are provided graphically in Figure 5.4.

5.1.2 Visual Inspection of the Sample Segmentation Results

Sample segmentation results are provided in Figure 5.5 for visual inspection. The images in the first column are the baseline segmentation results with the lowest ANMI and the images in the second column are the baseline segmentation results with the highest ANMI. The images in the third column are the segmentation results of the Boosted-MRF method. Corresponding PRI, GCE and ANMI values are also provided for each image. It is observed in the first, third and fifth sample images that performance of the Boosted-MRF segmentation is at least as high as the performance of the best baseline segmentation result. On the other hand, in case of weak baseline segmentations as in the second sample image, considerable performance increase is achieved by the Boosted-MRF segmentation. Also, it is worth pointing out that, the
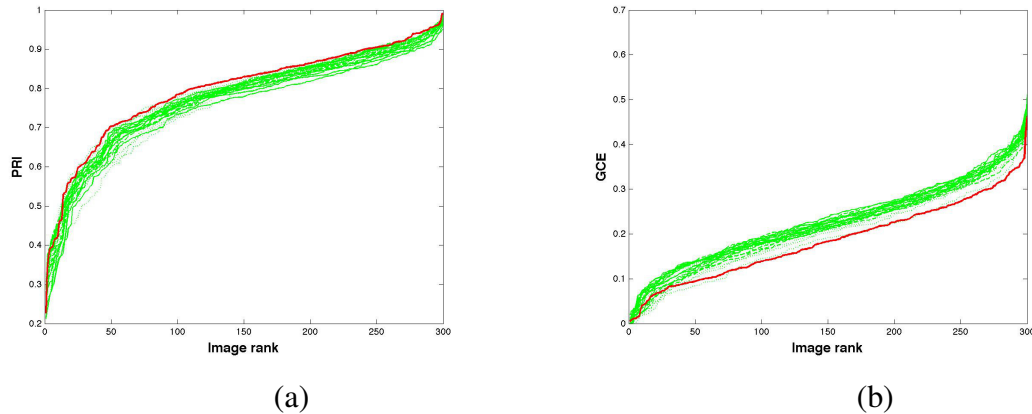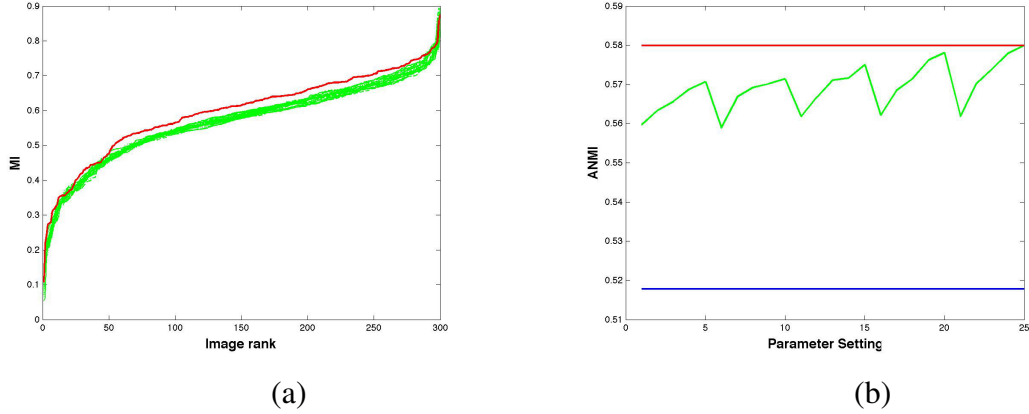
Figure 5.4: Performance comparison of the Boosted-MRF method and the baseline MS segmentation methods by MI and ANMI: Performance of the baseline methods are plotted in green, performance of the Boosted-MRF method plotted in red. (a) MI vs. image-rank for baseline segmentations and the Boosted-MRF segmentation. (b) parameter settings (each parameter setting corresponds to a distinct baseline segmentation method) vs. ANMI. In this graph, performance of the calssic MRF method is plotted in blue.

performance measures may not be sensitive enough in case of very small regions, such as the tree branches in the fourth image. Although, PRI and ANMI indicates lower performance for the Boosted-MRF segmentation, visual inspection reveals that the Boosted-MRF method segments the image at least as accurate as the baseline segmentation with the highest ANMI.

5.1.3 Comparison of the Boosted-MRF Method with a Consensus Segmentation System

Performance of the Boosted-MRF method is compared with the state of the art consensus segmentation system, namely Best of K (BOK). BOK selects the best baseline segmentation, $P^* \in \{P_1, ..., P_K\}$ by minimizing the *symmetric difference distance*, SDD [124] which is defined as follows:

$$d(P_i, P_j) = n_{01} + n_{10} \tag{5.1}$$

where n_{01} is the number of pair of pixels that are in the same region in P_i but not in

PRI=0.94, GCE=0.21, ANMI=0.78   PRI=0.96, GCE=0.19, ANMI=0.83   PRI=0.97, GCE=0.16, ANMI=0.84

PRI=0.45, GCE=0.03, ANMI=0.29   PRI=0.97, GCE=0.01, ANMI=0.76   PRI=0.99, GCE=0.01, ANMI=0.84

PRI=0.91, GCE=0.17, ANMI=0.67   PRI=0.96, GCE=0.08, ANMI=0.83   PRI=0.96, GCE=0.08, ANMI=0.83

PRI=0.94, GCE=0.13, ANMI=0.79   PRI=0.97, GCE=0.08, ANMI=0.88   PRI=0.91, GCE=0.06, ANMI=0.83

PRI=0.93, GCE=0.34, ANMI=0.73   PRI=0.94, GCE=0.23, ANMI=0.78   PRI=0.95, GCE=0.21, ANMI=0.80
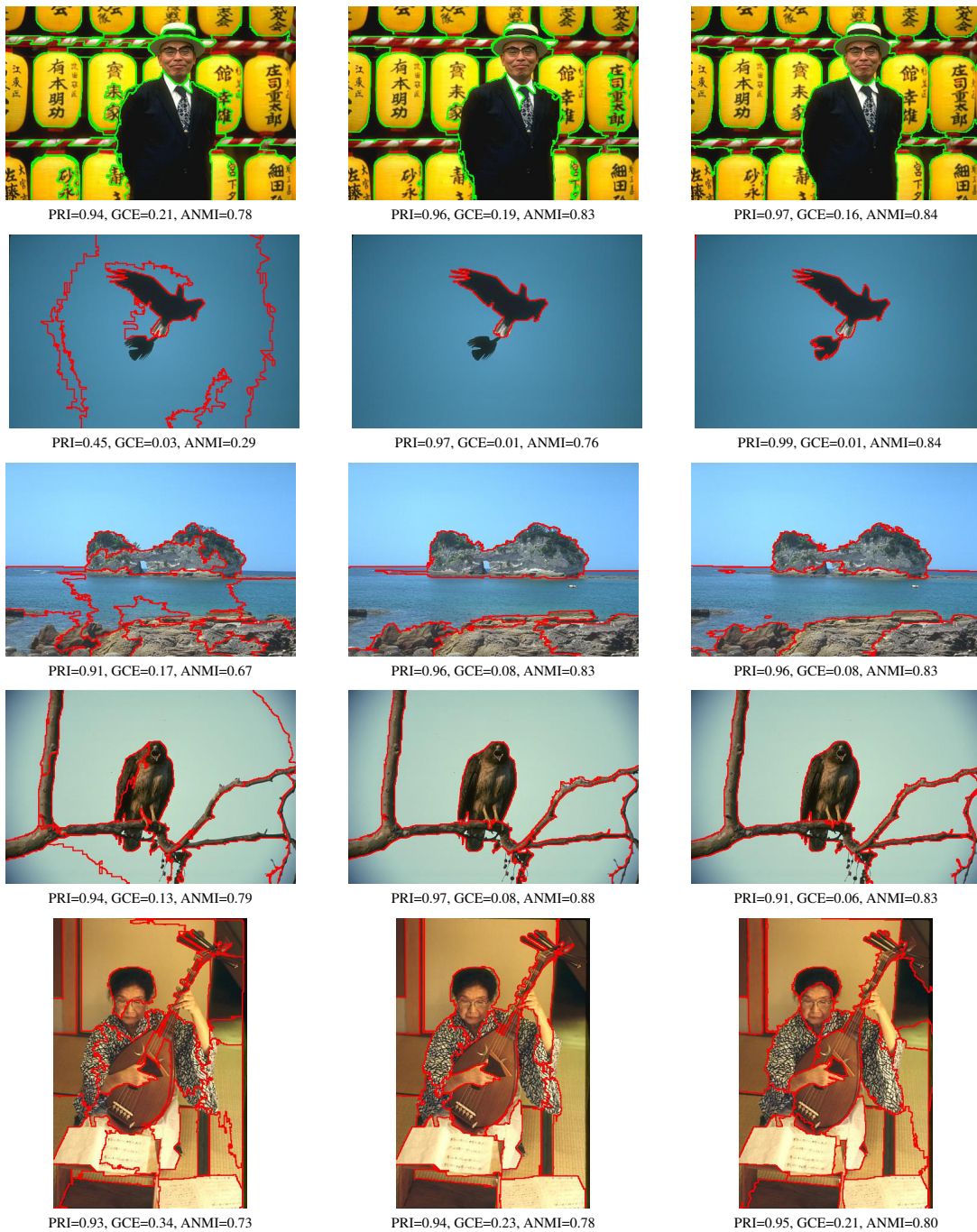
Figure 5.5: Images in the first and the second column are the baseline segmentations with the lowest and the highest ANMI respectively and the images in the third column are the Boosted-MRF segmentations.

(a)   (b)
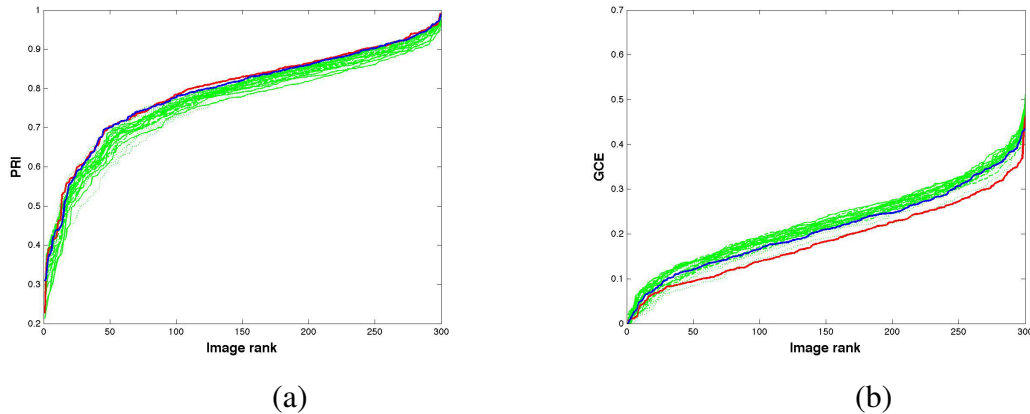
Figure 5.6: Comparison of baseline EG segmentation methods, corresponding Boosted-MRF segmentation and the BOK algorithms by PRI and GCE: The performance of baseline segmentations are plotted in green, while the performance of the Boosted-MRF method is plotted in red and the performance of BOK is plotted in blue.

.

P_j and n_{10} is the number of pair of pixels that are in the same region in P_j but not in P_i.

Performance of the BOK segmentation is compared to the Boosted-MRF segmentation both in the first and in the second experimental setup. Performance of the Efficient Graph Based baseline segmentations, corresponding Boosted-MRF segmentation and the BOK segmentation are compared via PRI and GCE in Figure 5.6. In this figures, PRI values of Boosted-MRF and BOK are very close while the GCE value for Boosted-MRF indicates higher performance compared to the BOK segmentation. Performance comparisons for the same set of segmentations by NMI and ANMI are provided in Figure 5.7, where the performance increase achieved by Boosted-MRF is monitored more clearly.

Performance of the baseline Mean-Shift segmentations, corresponding Boosted-MRF segmentation and the BOK segmentation are compared via PRI and GCE in Figure 5.8 and via Normalized Mutual Information and ANMI in Figure 5.9. Comparisons in this experimental setup are consistent with the comparisons in the first experimental setup. Again the PRI values are very close achieving the highest baseline segmentation result, while the GCE and NMI of Boosted-MRF are slightly better than those of the BOK. Although, average normalized mutual information of BOK is not

65
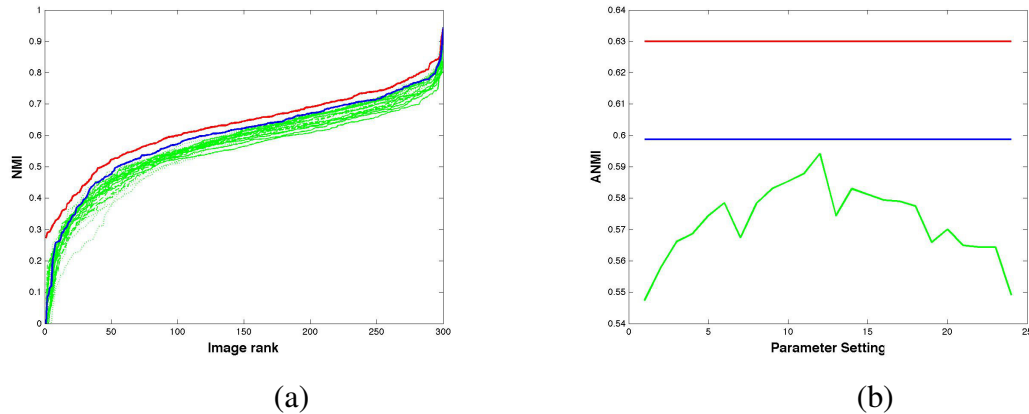
Figure 5.7: Comparison of baseline EG segmentations, corresponding Boosted-MRF segmentation and the BOK algorithms by NMI and ANMI: The performance of baseline segmentations are plotted in green, while the performance of the Boosted-MRF method is plotted in red and the performance of BOK is plotted in blue.

.

higher than all the baseline segmentations, Boosted-MRF achieves a better ANMI value compared to the BOK and the baseline segmentations.

BOK is selected as the state of the art consensus segmentation system, since it is commonly used in the literature and it is faster compared to the other methods such as Best One Element Move (BOEM). Main disadvantage of BOK is that, it selects the "best" segmentation among an ensemble of segmentations. But, it does not provide a better segmentation than the best segmentation in the ensemble. On the other hand, BOEM can provide a better segmentation, but it is computationally too expensive. Besides, the Boosted-MRF method gets a better solution than the best ensemble segmentation without a very high computational cost. The computational complexity analysis for the BOK and the Boosted-MRF are provided in the subsequent sections.

5.1.4 Sensitivity of the Boosted-MRF Method to α

Recall that, α defines the neighborhood weights in the range $[0, \alpha]$, hence its a critical parameter of Boosted-MRF. In this experiment, sensitivity of Boosted-MRF to the α parameter is analyzed. For this purpose, Boosted-MRF is run by setting α as $\{2, 5, 8\}$ and the PRI and GCE values for the corresponding segmentations are reported in Table 5.1. In a fourth run of Boosted-MRF, for each image three segmentation outputs
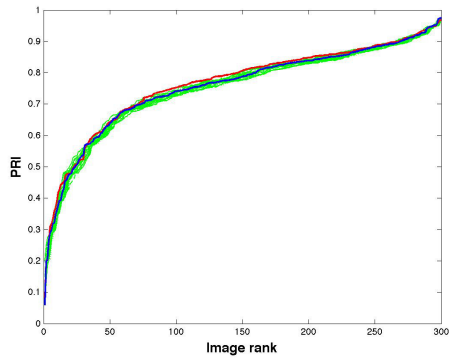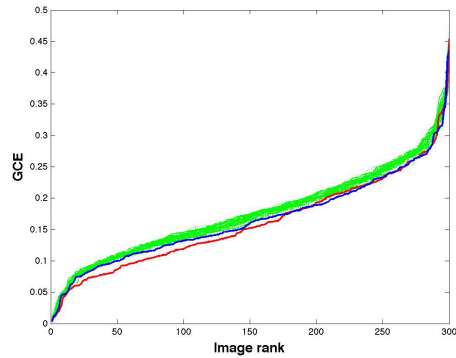
| (a) | (b) |

Figure 5.8: Comparison of baseline MS segmentation methods, corresponding Boosted-MRF segmentation and the BOK algorithms by PRI and GCE: The performance of baseline segmentations are plotted in green, while the performance of Boosted-MRF is plotted in red and the performance of the BOK is plotted in blue.

.
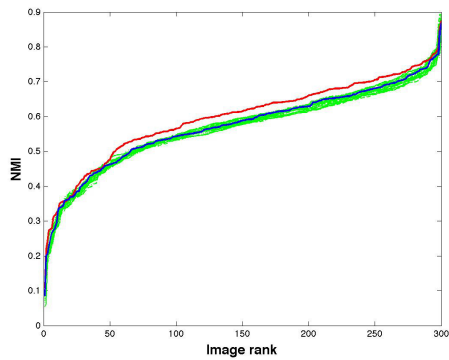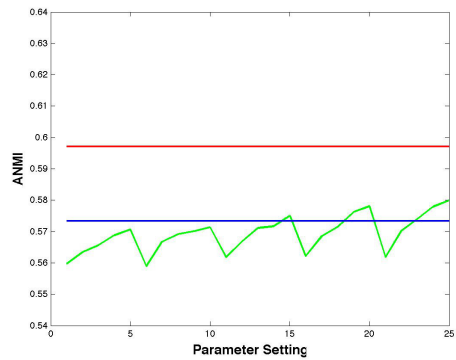
| (a) | (b) |

Figure 5.9: Comparison of baseline MS segmentation methods, corresponding Boosted-MRF segmentation and the BOK algorithms by NMI and ANMI: The performance of baseline segmentations are plotted in green, while the performance of the Boosted-MRF method is plotted in red and the performance of the BOK is plotted in blue.

.

Table 5.1: Performance of the Boosted-MRF for three different values of α

α	PRI	GCE
2	0.7755	0.1882
5	0.7967	0.1939
8	0.7875	0.1945
α selected by Mutual Information	**0.7982**	**0.1853**

for each value of α is obtained and the one which maximizes the mutual information is selected. For this purpose, mutual information between each segmentation output and the set of baseline segmentations is estimated and then the segmentation which has the maximum sum of mutual information is selected. The highest performance is obtained by selection based on mutual information.

5.1.5 Sensitivity of the Boosted-MRF Method to the Initial Segmentation

In this experiment, the performance of the Boosted-MRF method is examined for various initial segmentations. If the method starts with an over-segmentation, it reduces the number of regions during energy minimization by swap/expansion. The parameters of the initial segmentation should ensure that the initial segmentation has at least two partitions. Hence, if the parameters of the baseline segmentations are selected such that at least two regions are extracted for each image, then one of the baseline segmentations can be employed as the initial segmentation. Otherwise, another segmentation whose parameters ensures that each image is segmented into two regions, is employed as the initial segmentation. In the second experimental setup, where Mean Shift segmentation method is employed for baseline, setting $minArea$ parameter as 1000 resulted in having a segmentation output with only one region for at least one image in the dataset. Therefore, the $minArea$ parameter is set as 500, and h_r and h_s are set as $\{2, 4, 6\}$ to obtain 9 initial segmentations. Segmentation performance for each initial segmentation is provided in Table 5.2. In this table, it is monitored that the Boosted-MRF method is not sensitive to the initial segmentation.

Table 5.2: Performance of the Boosted-MRF for nine initial conditions

Parameters for initial segmentation	PRI	GCE
$h_r = 2, h_s = 2, minArea = 500$	0.7491	0.1691
$h_r = 2, h_s = 4, minArea = 500$	0.7503	0.1762
$h_r = 2, h_s = 6, minArea = 500$	0.7520	0.1730
$h_r = 4, h_s = 2, minArea = 500$	0.7493	0.1676
$h_r = 4, h_s = 4, minArea = 500$	0.7520	0.1743
$h_r = 4, h_s = 6, minArea = 500$	0.7522	0.1758
$h_r = 6, h_s = 2, minArea = 500$	0.7502	0.1672
$h_r = 6, h_s = 4, minArea = 500$	0.7547	0.1708
$h_r = 6, h_s = 6, minArea = 500$	0.7585	0.1738

5.1.6 Comparison of the Boosted-MRF with the State of the Art Segmentation Systems

Segmentation performance of the Boosted-MRF method is compared with the state of the art segmentation systems by PRI in Table 5.3. The segmentation performances of the state of the art methods are reported in the study of Kampa et al. [63]. The Boosted-MRF outperforms most of the state of the art systems, only the Contour Detection and Hierarchical Segmentation method [9] performs better than the Boosted-MRF. However, due to the high computational complexity of the contour detection based segmentation system its application on certain problems such as for high resolution images is not efficient. On the other hand, Boosted-MRF can be efficiently employed in large size images. Note that, the performance of the Boosted-MRF method is sensitive to the performance of baseline segmentation outputs. The reported performance of the Boosted-MRF in Table 5.3 is obtained by employing a set of EG segmentation outputs as described in Section 5.1.1.1.

5.2 Experiments on the DS-MRF Method

The DS-MRF method is proposed for the problems where class labels are not available, but a generic set of domain specific information can be received from a domain expert. Hence, effectiveness of the proposed system can be observed only on problems for which some domain specific information is available. In this thesis, the

Table 5.3: Comparison of the Boosted-MRF to the state of the art segmentation systems.

Method	PRI
meDDT [63]	0.77
CDHS [9]	0.81
Mean Shift [28]	0.76
NCuts [103]	0.72
EG [42]	0.78
Boosted-MRF	**0.80**

DS-MRF method is applied on two such problems; first one is a dataset consisting of outdoor images with vegetation and the other one is the building detection problem from remote sensing.

5.2.1 An Application on Outdoor Images with Vegetation

The DS-MRF method is employed on a dataset consisting of outdoor images with vegetation. Vegetation detection is a common problem in the computer vision, which may arise in various application areas, such as outdoor navigation in robotics, or image retrieval system on an animal or plant dataset. In this problem, color band information is utilized as domain specific information to distinguish vegetation regions which are chlorophyll-rich.

5.2.1.1 Dataset

A subset of the Microsoft Research Cambridge Object Recognition Image Database (MSRC) [119] is utilized in this experiment. All images in this dataset have pixel-wise ground truth segmentations. The dataset is widely used for recognition and segmentation problems in the literature. The images in the dataset are categorized as belonging to a certain object class such as building, grass, tree, cow, horse etc. In our experiments, 115 outdoor images consisting of animal (cow, horse, sheep) and flower images are used. Sample images are provided at *Figure 5.10*. Common characteristic of these images is that they all have vegetation in the background.
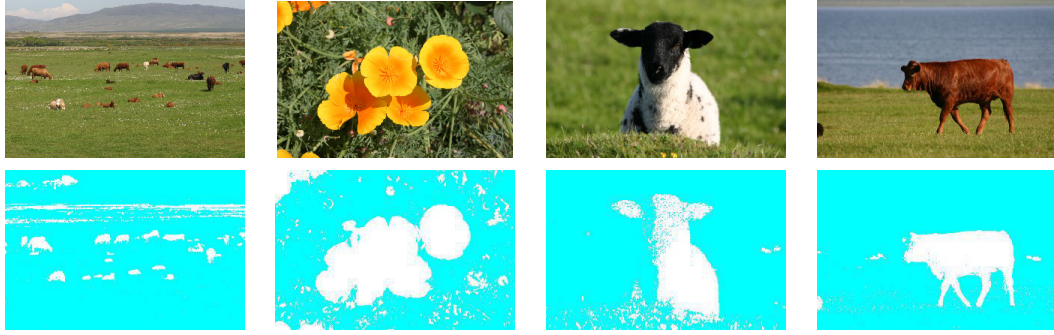
Figure 5.10: First row: Sample images from the MSRC dataset [119], Second row: NDI thresholding results of the sample images.

5.2.1.2 Domain Specific Information

Normalized Difference Vegetation Index (NDVI) is used to detect vegetation regions, [1]. This index is a normalized ratio of red and near-infrared reflectance. Similar to NDVI, [85] propose Green-Red Vegetation Index (GRVI) as follows;

$$GRVI = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red}} \tag{5.2}$$

where, ρ_{green} and ρ_{red} are reflectances of visible green and visible red respectively. Motohka et al. propose that the types of land covers can be categorized into three by GRVI thresholding; green vegetation has $GRVI > 0$ while soils have $GRVI < 0$ and water/snow has $GRVI \approx 0$. GRVI is also referred as Normalized Difference Index (NDI) in the literature and it is widely used for vegetation detection, [97]. In this study, GRVI thresholding similar to Motohka's method is applied to segment background regions of outdoor images. GRVI thresholding results for sample images are given in *Figure 5.10*. In our problem, exact detection of land cover is not crucial but an approximate clustering of image pixels belonging to similar land covers can be very beneficial. In *Figure 5.10*, thresholding of all four images provide valuable information. Hence, GRVI thresholding is employed to obtain the domain specific map (DSM).

5.2.1.3 Experimental Setup and Quantitative Results

The DS-MRF method is applied on the dataset and its segmentation performance is compared with the state of the art segmentation systems; Mean Shift Segmentation (MS) [28], Contour Detection and Hierarchical Image Segmentation (CDHS) [9], Multiscale Normalized Cuts (Mult. NCut)[29], Efficient Graph Based Segmentation (EG) [42] and ROI-SEG [35]. DS-MRF is compared with these systems from the the literature via two measures; global consistency error (GCE) and probabilistic rand index (PRI).

During the RAG construction for the *DS-MRF*, super-pixels are obtained by the Mean Shift segmentation method. For this purpose, the range bandwidth and spatial bandwidth parameters are set as 2. In the Segmentation Fusion Module, BUSs are obtained by Mean Shift segmentation. The parameters of the MS method is set based on [93, 113]. Spatial bandwidth and range bandwidth parameters are varied with {3,7,11,15}. Total of 16 different segmentation results are obtained. The DSI related to this problem states that 'Each image has vegetation.' DSM corresponding to this DSI is obtained by thresholding the NDVI and strong neighborhood relation is defined between the adjacent super-pixels that are partitioned into the same region by DSM. There is no DSI related to the spatial constraints. Hence, two types of neighborhood relations appear in the MRF framework; standard neighborhood and strong neighborhood, but weak neighborhood relation does not appear in this experiment.

The DS-MRF system is compared with the state of the art unsupervised segmentation approaches; Mean Shift Segmentation (MS) of [28], Contour Detection and Hierarchical Image Segmentation (CDHS) of [9], Multiscale Normalized Cuts of [29], Efficient Graph Based Segmentation (EG) of [42] and ROI-SEG of [35]. Parameters of each algorithm are adjusted to have segmentations with at least 3 and at most 5 regions, so that a fair comparison can be done among these methods.

Mean Shift, [26], is run with various values of spatial bandwidth and range bandwidth parameters by adopting the study of [93]. Minimum region size parameter of MS is adjusted for each image such that the number of regions remains in the interval of [3,5]. Cour's Multiscale Normalized Cut method, [29, 33] is run by setting its only

parameter, number of regions, as {3,4,5}. Efficient Graph Based Segmentation, [34] is employed, whose merge threshold coefficient, k, is set based on the size and the resolution of the image. Large values of k results in larger regions. In our experiments k is set as {5,8,11}. Contour Detection and Hierarchical Image Segmentation (CDHS) is employed by adjusting its contour threshold parameter, which takes values in the range [0,1], to obtain coarse or fine segmentation results. In our experiments, threshold is adjusted to have segmentations with number of regions in [3,5] for each image. Publicly available code for ROI-SEG method of Donoser and Bischof is employed. They propose setting Maximally Stable Extremal Region Detection distance to 25, hence we employed the algorithm for three values of MSER detection distance; {15,25,30}. Each algorithm is run with various parameters and the segmentation results with the highest PRI are reported in *Table 5.6*.

Additional to the state of the art methods, proposed method is also employed without DSI, which is also reported in the same table as DS-MRF w/o DS info. Finally, the proposed system is compared with the supervised segmentation method of [69]. They proposed a supervised segmentation method, which takes supervision interactively. For each image, a set of pixels are labeled for 3-5 classes. The PRI and GCE values are reported as 'Supervised Seg.' in Table 5.6. In this table, considering both PRI and GCE, best segmentation performance is obtained by DS-MRF.

Performance comparisons are also plotted in Figure 5.11. In Figure 5.11 (a) is the PRI values sorted in ascending order for all algorithms and Figure 5.11 (b) is the GCE values sorted in ascending order. In these figures, the contribution of the DS-MRF method is observed both by the increase PRI and by the decrease in GCE.

5.2.1.4 Visual Inspection of the Sample Segmentation Examples

While considering the performance of various segmentation systems, visual inspection of the segmentation results is very important, since the performance measures may be insufficient to evaluate the visually meaningful results. Segmentation results for sample images from the MSRC dataset are given in Figures 5.12-5.15. In these figures, the original images are provided at the first row. Segmentation results for EG, Multiscale NCut, Mean Shift, ROI-SEG, CDHS and DS-MRF are given at the

73

Table 5.4: Comparison of EG, Multiscale NCut, Mean Shift, ROI-SEG, CDHS, Classic MRF and DS-MRF performances on MSRC images

Method	PRI	GCE
EG	0.65	0.22
Multiscale NCut	0.62	0.30
Mean Shift	0.67	0.20
ROI-SEG	0.72	0.17
CDHS	0.75	0.17
DS-MRF w/o DS info (Boosted-MRF)	0.76	0.17
Supervised Seg.	0.76	0.18
DS-MRF	**0.78**	**0.15**
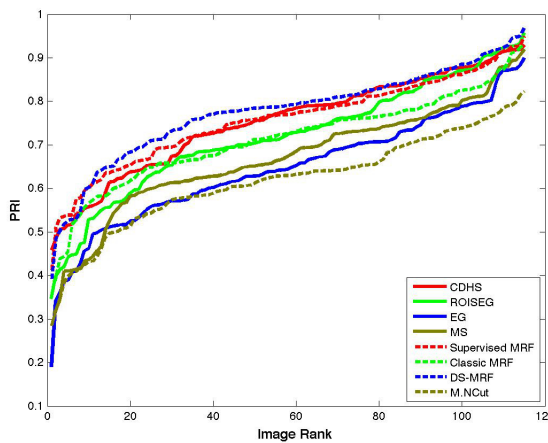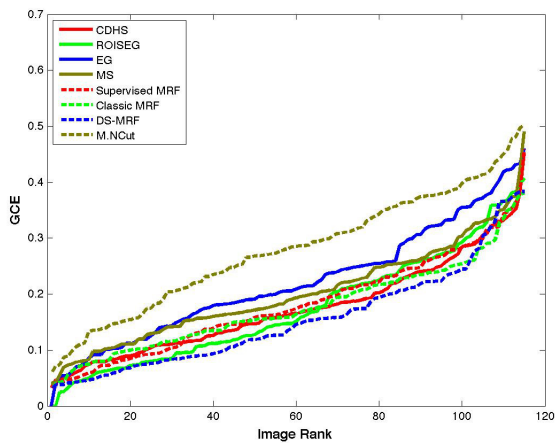
(a)   (b)

Figure 5.11: Performance comparison of DS-MRF with state of the art segmentation method by PRI and GCE: (a) PRI vs. ranked image and (b) GCE vs. ranked image on MSRC image dataset.

second, third, fourth, fifth, sixth and seventh rows, respectively. PRI and GCE values for each image are also provided.

Depending on the level of information provided by the domain specific information, increase in the performance changes. In the MSRC dataset, the domain specific information provides information about the background and this brings about an increase in the segmentation performance of the whole image. If the background can be segmented accurately by the domain specific information, foreground object(s) can also be segmented more accurately. This is observed clearly in the first image of Figure 5.12, second and fourth image of Figure 5.13, first and third images of Figure 5.14 and the first and third images of Figure 5.15, where background is segmented as a single region.

Visual inspection reveals the insufficient conditions in performance estimation by PRI and GCE. These two measures are not sensitive to tiny objects in the foreground. For example second image in *Figure 5.12* contains very small animals on a smooth background. Most of the segmentation techniques do not segment animals at all, only EG and DS-MRF segment this detailed image accurately. However, PRI and GCE values are not informative in these examples. In other words, in case of small objects which correspond to a small number of pixels PRI and GCE are not sensitive enough for evaluating segmentation performances.

5.2.2 Application of DS-MRF on the Airport Segmentation Problem

Remote sensing has been a popular research area in the last decade. Images captured by satellites such as Geoeye, Ikonos, QuickBird provide multi spectral high resolution images. Image processing methods in the computer vision literature turn out to be insufficient on remote sensing applications. This is due to the fact that objects in remote sensing images are complex and they have heterogeneous properties in terms of size, form, spectral behavior and etc.

EG PRI=0.89
GCE=0.07

PRI=0.69
GCE=0.14

PRI=0.79
GCE=0.16

PRI=0.64
GCE=0.14

M. NCut PRI=0.68
GCE=0.09

PRI=0.71
GCE=0.17

PRI=0.63
GCE=0.22

PRI=0.54
GCE=0.40

MS PRI=0.51
GCE=0.15

PRI=0.76
GCE=0.34

PRI=0.77
GCE=0.18

PRI=0.83
GCE=0.16

ROI-SEG PRI=0.57
GCE=0.05

PRI=0.72
GCE=0.12

PRI=0.87
GCE=0.12

PRI=0.82
GCE=0.17

CDHS PRI=0.84
GCE=0.09

PRI=0.67
GCE=0.08

PRI=0.85
GCE=0.14

PRI=0.82
GCE=0.16

DS-MRF PRI=0.68
GCE=0.08

PRI=0.72
GCE=0.15

PRI=0.81
GCE=0.16

PRI=0.85
GCE=0.08

Figure 5.12: Sample Segmentation Results, from the MSRC Dataset.

EG PRI= 0.79
GCE=0.10

PRI= 0.76
GCE=0.01

PRI= 0.58
GCE=0.23

PRI=0.74
GCE=0.01

M. NCut PRI=0.70
GCE=0.19

PRI=0.59
GCE=0.11

PRI=0.66
GCE=0.47

PRI=0.57
GCE=0.14

MS PRI=0.66
GCE=0.10

PRI=0.80
GCE=0.10

PRI=0.79
GCE=0.25

PRI=0.44
GCE=0.14

ROI-SEG PRI=0.73
GCE=0.23

PRI=0.85
GCE=0.07

PRI=0.74
GCE=0.32

PRI=0.55
GCE=0.11

CDHS PRI=0.84
GCE=0.10

PRI=0.92
GCE=0.07

PRI=0.65
GCE=0.23

PRI=0.89
GCE=0.05

DS-MRF PRI= 0.92
GCE=0.07

PRI=0.92
GCE=0.04

PRI=0.65
GCE=0.21

PRI=0.89
GCE=0.05
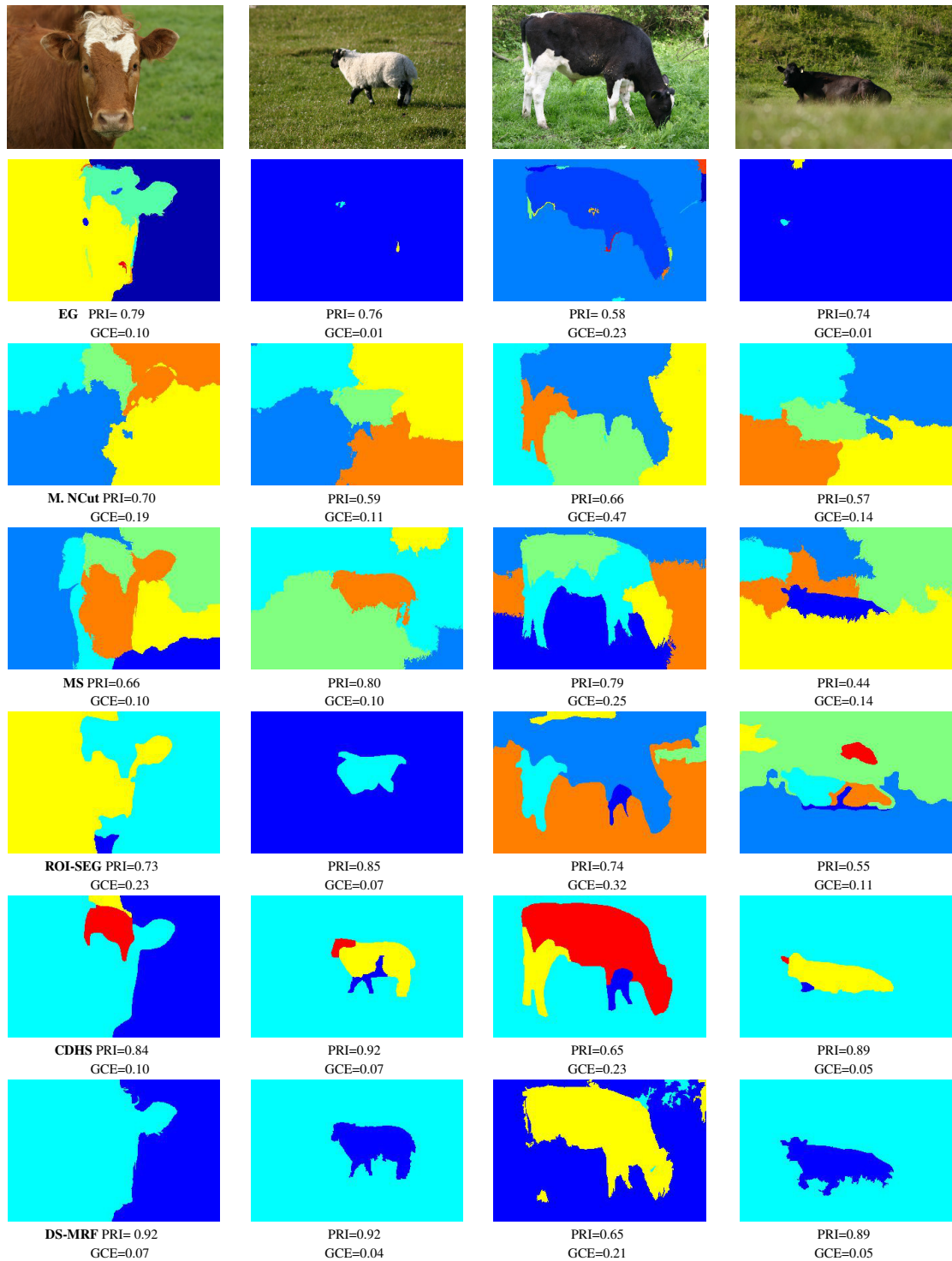
Figure 5.13: Sample Segmentation Results, from the MSRC Dataset.
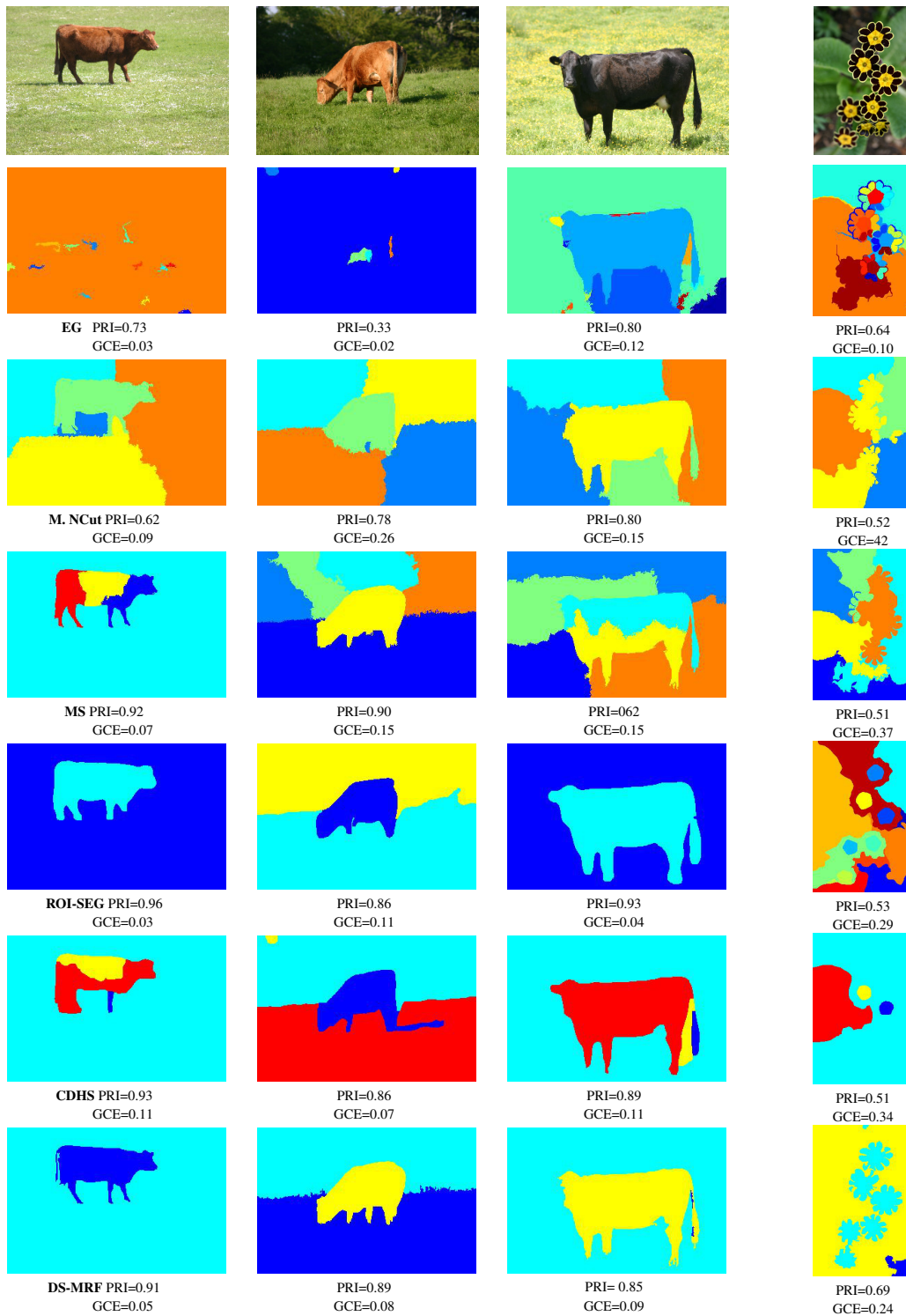
Figure 5.14: Sample Segmentation Results, from the MSRC Dataset.
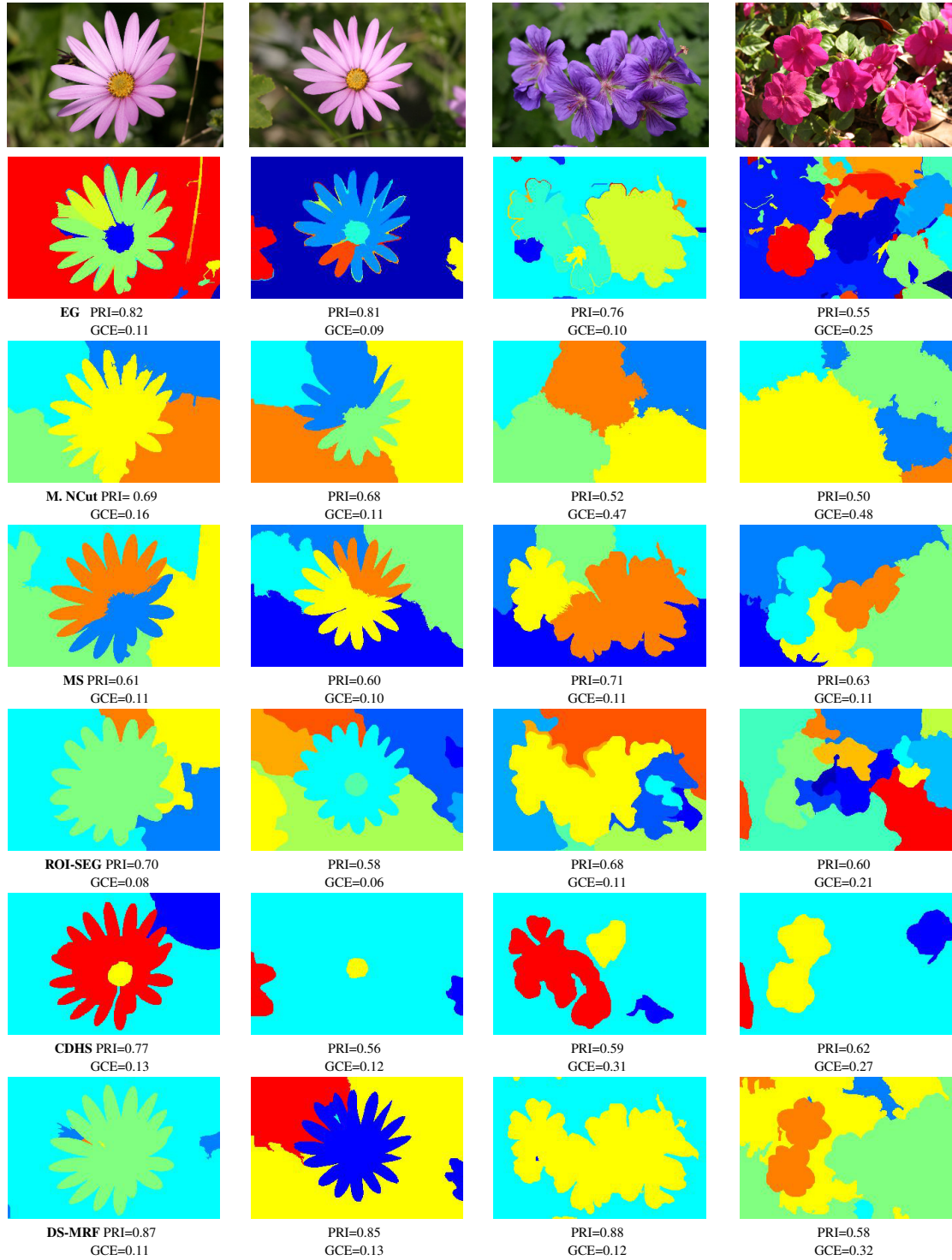
Figure 5.15: Sample Segmentation Results, from the MSRC Dataset.

5.2.2.1 Dataset

The airport dataset used in this study consists of 10 images from Geoeye and Ikonos. They are in 4m resolution and multi spectral images of 4 bands; R,G,B and NIR. Each image has a ground-truth where airport is segmented as region of interest (ROI) in which runway(s) and taxiroute is segmented as another region. A sample image and its ground-truth segmentation is provided in *Figure 5.16 (a)* and *(b)* repsectively.

5.2.2.2 Domain Specific Information

In this set of experiment, we employ two DSIs. The first one has the property,

$$p_1 = there - is - at - least - one - runway - at - airport.$$

Runways can be represented by regions bounded by parallel lines. Parallel line bounded region (PLBR) detection result for a sample image is given in *Figure 5.16 (c)* and this result is utilized as domain specific map DSM_{plbr}.

The second DSI has the property,

$$p_2 = there - is - vegetation - area - adjacent - to - runway.$$

For vegetation detection, Normalized Difference Vegetation Index (NDVI) is utilized [1]. his index is based on the fact that, plants absorb visible light in the range [400,700] nanometers while, they reflect near-infrared light in [700-1100] nanometers. NDVI is calculated as follows:

$$NDVI = \frac{NIR - VIS}{NIR + VIS},\tag{5.3}$$

where VIS and NIR stand for the spectral reflectance measurements acquired in the visible (red) and near-infrared regions respectively. These reflectance values are rations of the reflected radiation over incoming radiation in each spectral band. Hence they are in the range [0,1] and therefore, NDVI varies between -1 and 1. For a given image, its NDVI is obtained and a threshold is determined by Otsu method [2]. Pixels having an NDVI value greater than this threshold is labeled as vegetation. NDVI thresholding result of a sample airport image is provided at *Figure 5.16 (d)* and this

| (a) Original Image | (b) Groundtruth | (c) PLBR Detection | (d) NDVI thresholding |
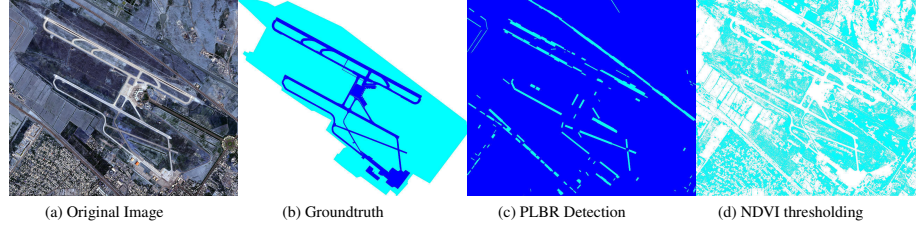
Figure 5.16: (a) An airport image and (b) its groundtruth segmentation, (c) parallel line bounded region detection result and (d) ndvi thresholding result.

output is utilized as domain specific map $DSM_{vegetation}$.

Segmentation fusion module determines the types of neighborhood relations between each pair of neighboring super-pixels to generate a domain specific RAG by fusing the information received from the bottom-up segmentation maps and domain specific maps. Two sets are defined on the available domain specific maps, DSM_{plbr} and $DSM_{vegetation}$ as follows:

$$DSM_{connect} = (DSM_{plbr}, DSM_{plbr}), \qquad (5.4)$$

$$DSM_{disconnect} = (DSM_{plbr}, DSM_{vegetation}). \qquad (5.5)$$

Each neighboring pair of super-pixels si and sj in a bottom-up segmentation map is assigned by one of the following neighborhood relations:

- **Strong Neighborhood** is defined between a pair of super-pixels which are both labeled as belonging to a parallel line bounded region by DSM_{plbr}.

- **Weak Neighborhood** is defined between a pair of super-pixels one of which is labeled as belonging to a parallel line bounded region by DSM_{plbr} and the other one is labeled as vegetation by $DSM_{vegetation}$.

- **Standard Neighborhood** is defined between a pair of super-pixels which are spatial adjacent and are not labeled as strong or weak neighbors.

5.2.2.3 Domain Specific MRF Energy

The CIELab color features are employed in the data term of the MRF energy function, which is modeled assuming that the features are drawn from Gaussian distribution. The smoothness term for a pair of super-pixels s_i and s_j is weighted by w_{ij}, which is defined as follows:

$$w_{ij} = \begin{cases} f(s_i, s_j) & \text{if } (s_i, s_j) \in N_{str} \\ \beta & \text{if } (s_i, s_j) \in N_{weak} \\ 1 & otherwise \end{cases} \tag{5.6}$$

where, s_i represents super-pixels as MRF sites and l_{s_i} is the label assigned to that super-pixel. β is constant such that $\beta \leq 0$. The first condition of the U_{smooth} function estimates the pairwise energy between a pair of strong neighbors, while the second condition estimates the energy for a pair of weak neighbors and the third condition estimates the energy for a pair of standard neighbors. So, the smoothness term of the MRF energy function is estimated with the following formula:

$$U_{smooth} = \alpha \times (\sum_{s_i \in S, s_j \in N_{str_{s_i}}} f(s_i, s_j) \times \psi_{ij}(l_{s_i}, l_{s_j}) + \sum_{s_i \in S, s_j \in N_{weak_{s_i}}} \beta \times \psi_{ij}(l_{s_i}, l_{s_j}) + \sum_{s_i \in S, s_j \in N_{std_{s_i}}} \psi_{ij}(l_{s_i}, l_{s_j}))$$

$$\tag{5.7}$$

ψ is the Potts model which decreases energy when neighboring super-pixels have the same label. β is set as 0, hence the smoothness term is estimated over the set of standard and strong neighbors. Finally, the energy function is minimized by the expansion algorithm of [21].

5.2.2.4 Experimental Setup

The segmentation performance of DS-MRF is compared with state of the art segmentation methods; MS, M-NCut, EG and ROI-SEG. Since the extraction of the contour map is very expensive on the high resolution images, CDHS is not employed in this experiment.

MS segmentation is employed by setting spatial and range bandwidth parameters as $\{3, 7, 11, 15\}$ and 16 segmentation outputs are obtained. A set of EG segmentation

outputs is obtained by setting the smoothing parameter as $\sigma = \{0.4, 0.5, 0.6, 0.7\}$, k value for the merging threshold as $\{150, 300, 450, 600\}$. $minArea$ parameter is set as 500 in both MS and EG segmentation methods. For each image, an MS and an EG segmentation output is selected among various outputs by employing Normalized Mutual Information (NMI). For this purpose, NMI among all MS segmentation outputs is estimated and the output which has the minimum sum of NMI over all MS segmentation outputs is selected. The EG segmentation output is selected similarly. The number of regions parameter of M-NCut segmentation is set for each image as the average number of regions of MS and EG segmentation outputs of that image. ROI-SEG's MSER detection distance is set as $\{15, 25\}$ and its minimum region area size parameter is set such that each image has approximately the same number of regions as the average number of regions of MS and EG segmentation outputs of that image.

In DS-MRF segmentation, MS segmentation is employed for obtaining BUS outputs. Spatial and range bandwidth parameters are assigned to values {3,7,11,15} and 16 segmentation outputs are obtained for Low-Level Information Fusion in the segmentation fusion module.

DS-MRF segmentation results for sample airport images, together with segmentation results of four other systems MS, ROI-SEG, Multiscale NCut, EG are given in Figures 5.18-5.26.

5.2.2.5 Experimental Results and Quantitative Evaluation

The performance metric employed in this experiment is referred as ***precision in region of interest*** and it measures the accuracy of segmentation in the region of interest. For a given segmentation S_1 and its ground-truth segmentation S_2, precision in region of interest is estimated as follows; for each region in S_2 set of corresponding pixels $\{p_a\}$ in S_1 and their corresponding labels $\{l_a\}$ in S_1 are determined. For each label in $\{l_a\}$, its precision is estimated. Mean precision of all regions in S_2 is estimated as the segmentation precision of that image. Since the metric is directly predefined on region of interest, recall is very close to 1. Precision and recall provide an estimate that is close to human perception and F-score is the harmonic mean of precision and

Method	F-score	Mean Precision
EG	0.44	0.29
Mean Shift	0.47	0.32
Multiscale NCut	0.48	0.32
ROI-SEG	0.31	0.19
DS-MRF	**0.53**	**0.36**

Table 5.5: Comparison of EG, Multiscale NCut, Mean Shift, ROI-SEG and DS-MRF performances on airport images

recall. F-score and mean precision values for each algorithm are provided in *Table 5.5*. DS-MRF segmentation results for sample airport images, together with segmentation results of four other systems MS, ROI-SEG, Multiscale NCut, EG are given in *Figure 5.25*. In this figure, only region of interest that is, image regions belonging to airport by ground-truth is displayed to simplify visual inspection.

It is expected that an ideal segmentation algorithm obtains regions corresponding to objects or object parts. It is also expected that the image is not over-segmented or under-segmented. Visual inspection of segmentation examples shows that the segmentation result of DS-MRF is semantically more meaningful than other segmentation results. For example, in the first image of *Figure 5.25* although precision of Multiscale N-Cut segmentation is the highest, it is clearly observed that airport regions are represented with less number of regions by DS-MRF and regions of DS-MRF are semantically more informative.

Segmentations with larger regions of airports are considered as more advantageous. Hence, the number of regions corresponding to airports are also monitored at each image. Although, the performance of DS-MRF and Multiscale N-Cut are close in *Table 5.5*, DS-MRF partitions the airport into a smaller number of regions than Multiscale N-Cut does.

5.2.3 Segmentation for Building Detection

The vast amount of building detection and segmentation studies can be categorized with respect to many criteria, such as the type of sensors, the dimension of input
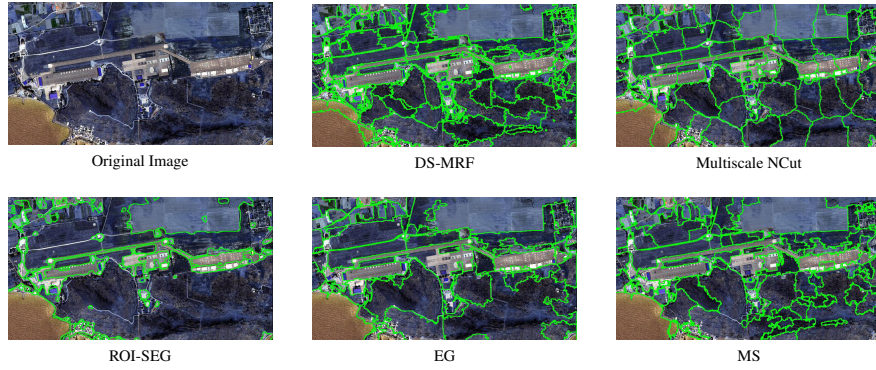
Figure 5.17: A sample airport image and its segmentation by five algorithms.
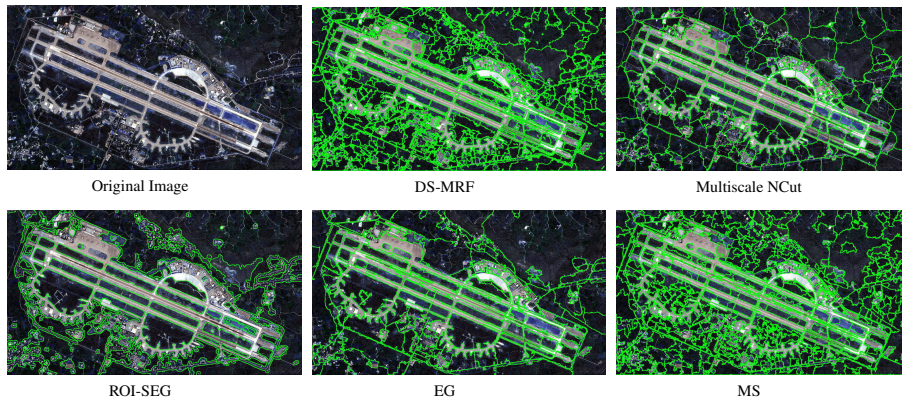
Figure 5.18: A sample airport image and its segmentation by five algorithms.
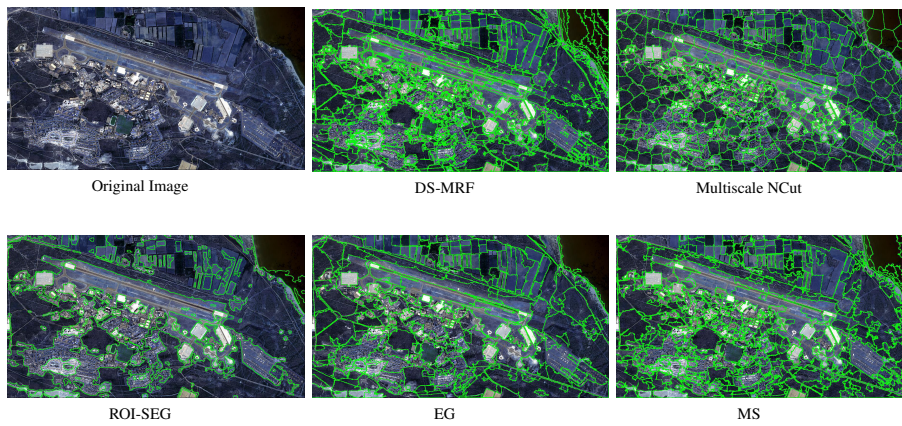
Figure 5.19: A sample airport image and its segmentation by five algorithms.
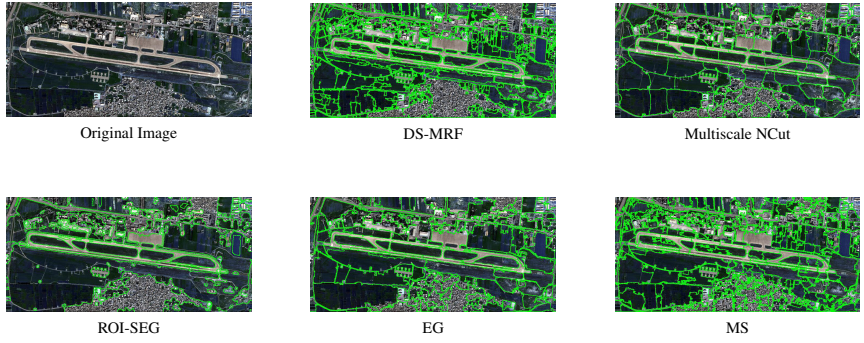
Figure 5.20: A sample airport image and its segmentation by five algorithms.
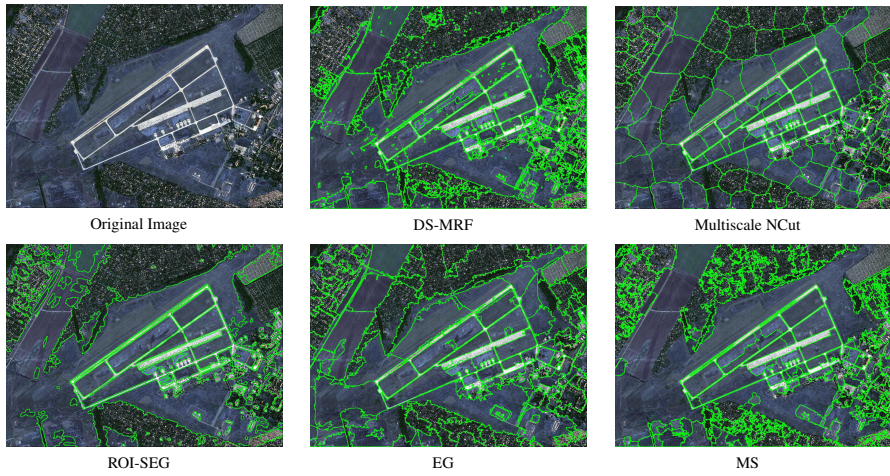
Figure 5.21: A sample airport image and its segmentation by five algorithms.
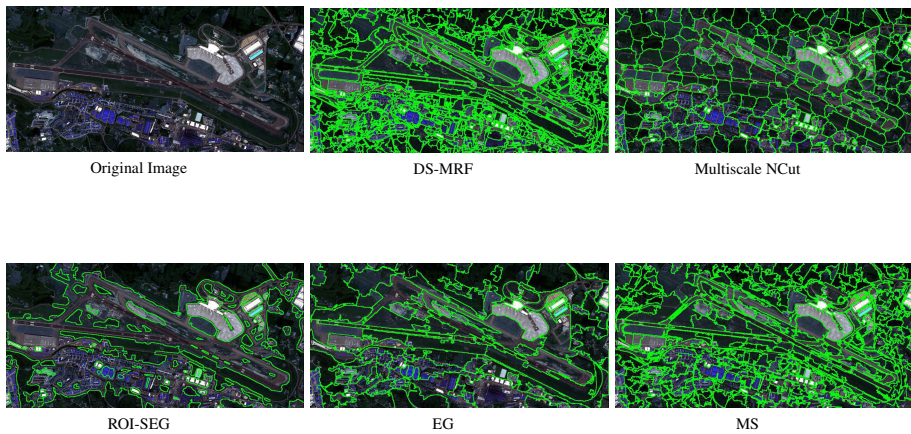
Figure 5.22: A sample airport image and its segmentation by five algorithms.
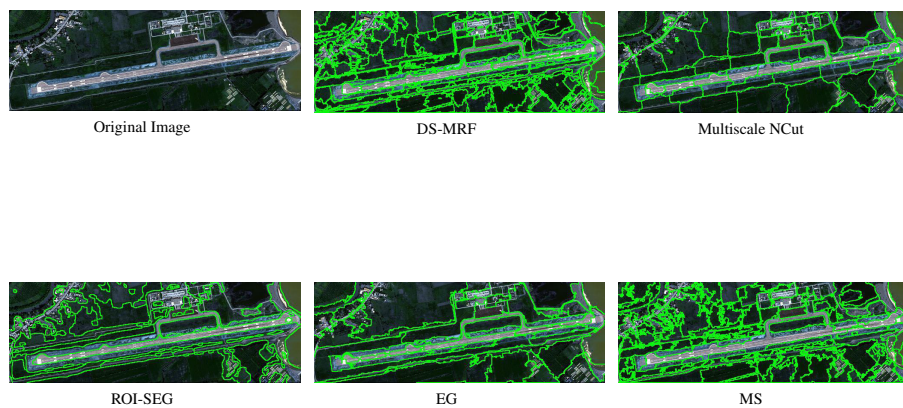
Figure 5.23: A sample airport image and its segmentation by five algorithms.
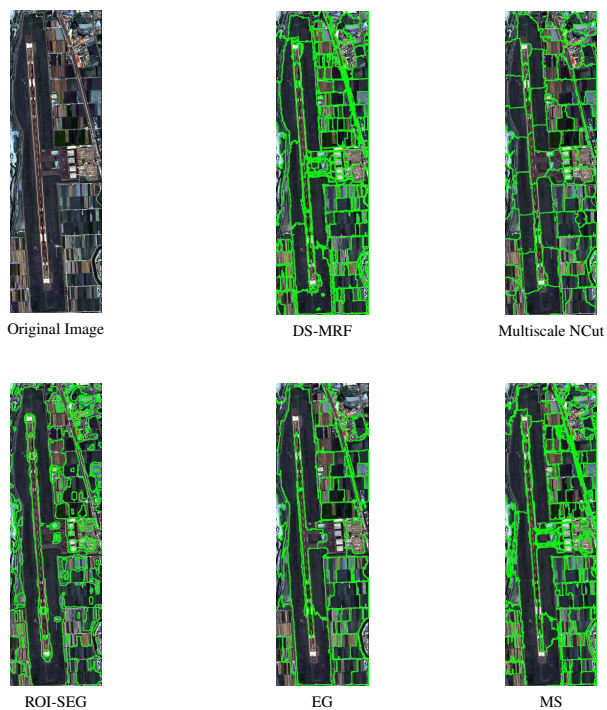
Figure 5.24: A sample airport image and its segmentation by five algorithms.

Original Image   DS-MRF   Multiscale NCut

ROI-SEG   EG   MS
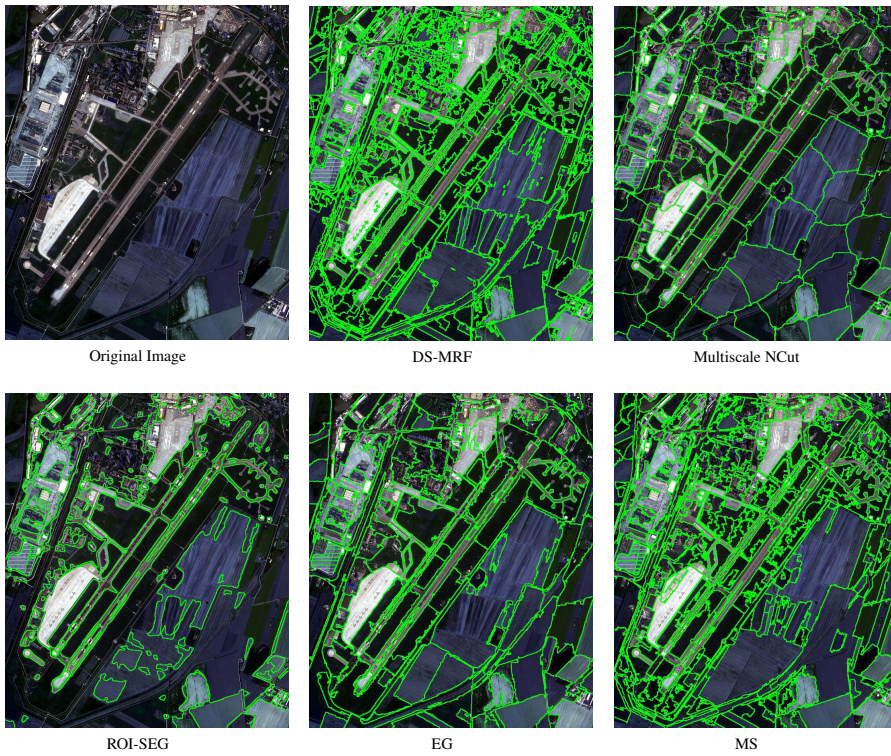
Figure 5.25: A sample airport image and its segmentation by five algorithms.
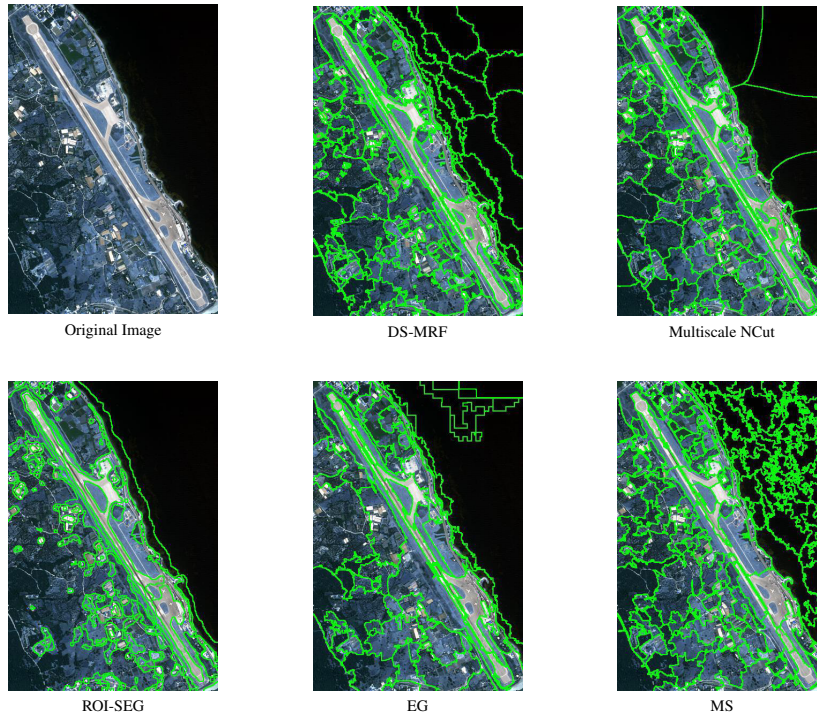
Figure 5.26: A sample airport image and its segmentation by five algorithms.

data, the degree of supervision etc. In this study, since an unsupervised segmentation algorithm for a single monocular optical image is proposed, we focus on the state of the art unsupervised segmentation methods, which employ characteristic features for buildings, such as color and features which capture rectangular shape.

In the building detection applications various domain cues are available. Generally, rectangularity of the buildings [98, 61] and shadow information [89, 61, 6] is utilized during the feature extraction step. In majority of these studies, first, the image is over-segmented and then the available information is processed on the super-pixels to detect buildings [126]. The initial segmentation is usually obtained by Watershed or Mean Shift segmentation methods [6, 126].

Domain specific information is usually employed in a stepwise manner, where certain image parts are eliminated and the rest is processed for detection of urban areas. That is, domain specific information is employed for elimination of certain regions such as shadows or vegetations in many studies [126, 89]. However, domain information does not necessarily provide hundred percent reliable information. In other words,

89

thresholding techniques for vegetation detection or shadow detection methods do not provide exact information. Hence, elimination of certain image parts may result in loss of information.

In our approach domain specific information is represented in a fuzzy form by means of MRF energy. In this way, stepwise elimination of image regions is avoided. Instead of processing available information step-by-step, all information is unified via MRF energy while obtaining an initial segmentation before detection takes place. Generally, in a detection problem, image is first over-segmented and then regions are labeled as belonging to certain object(s) or background. Hence, quality of initial partitioning is very important. If semantically meaningful regions are obtained in the first segmentation then performance of object detection is increased. Therefore, the goal in this experiment is to obtain semantically meaningful over-segmentations.

5.2.3.1 Dataset

The dataset consists of 18 images from two satellites; QuickBird(61cm) and Geoeye-1(50cm). All images have four multispectral bands (R,G,B and NIR) and each band has a resolution of 11 bits. Sample images from the dataset are given in Figure 5.27. In this dataset which is also utilized by Şenaras [101] in a building detection system, various illumination conditions are covered. Each image in this dataset has a corresponding ground-truth segmentation which consists of two regions; building and background. Since only buildings are segmented in the ground-truth segmentation, *precision in region of interest* and *Overall Segmentation Quality* (OSQ) [100] is employed for evaluating segmentation performance.

5.2.3.2 Domain Specific Information

Visual cues; rectangularity, shadow and vegetation are employed as domain specific information in the proposed DS-MRF system. Initially all DSI is employed separately to obtain a corresponding DSM. Later, information from all DSMs are processed together in RAG update module.

Figure 5.27: Sample images from the building image dataset.

As the first DSI, *rectangularity* is used since rectangular regions are strong candidates for buildings. For this purpose, Senaras's robust rectangle detection method [101] is utilized. In this method, first all the lines in the image are detected by Line Segment Detector [115]. Next, for each pair of lines a perpendicularity check is done. Later each perpendicular line pair, $line_1$ and $line_2$, is fit to a rectangle $R = \{P_1, P_2, P_3, P_4\}$ by finding the fourth corner, P_4 by Eq. 5.8 and 5.9..

$$P_4(x) = (m_2 \times P_1(x) - P_1(y)) - \frac{m_1 \times P_2(x) - P_2(y)}{m_2 - m_1} \tag{5.8}$$

$$P_4(y) = m_1 \times \frac{P_1(y) - m_2 \times P_1(x)}{m_1 - m_2} - m_2 \times \frac{P_2(y) - m_1 \times P_2(x)}{m_1 - m_2} \tag{5.9}$$

where, m_1 and m_2 are slopes of $line_1$ and $line_2$ respectively, P_3 is the intersection of the lines and P_1 and P_2 are the end points of $line_1$ and $line_2$ respectively.

At the end of this process, a binary map is obtained which is employed as the domain specific map of rectangles, $DSM_{rectangles}$. A sample image together with its rectangle detection result is provided at *Figure 5.28*.

Vegetation information is utilized as the second DSI. For vegetation detection, Normalized Difference Vegetation Index is used. This index is based on the fact that, plants absorb visible light in the range [400,700] nanometers while, they reflect near-infrared light in [700-1100] nanometers. NDVI is calculated as follows:

91
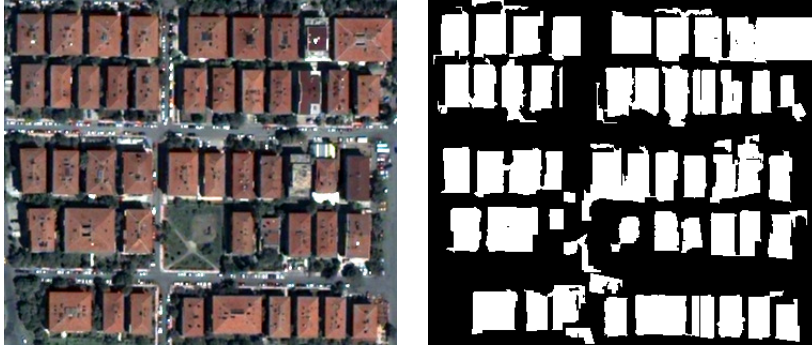
Figure 5.28: Sample image and its $DSM_{rectangles}$.

Figure 5.29: Sample image and its $DSM_{vegetation}$.

$$NDVI = \frac{NIR - VIS}{NIR + VIS},\qquad(5.10)$$

where VIS and NIR stand for the spectral reflectance measurements acquired in the visible (red) and near-infrared regions respectively. These reflectance values are rations of the reflected radiation over incoming radiation in each spectral band. Hence, they are in the range [0,1] and NDVI varies between -1 and 1.

For a given image, first histogram of NDVI values are computed. Then, a thresholding method such as Otsu [2] is used to partition the image into vegetation and non-vegetation regions. This partition is used as a domain specific map $DSM_{vegetation}$. A sample image and its $DSM_{vegetation}$ is provided in *Figure 5.29*.

Figure 5.30: Sample image and its DSM_{shadow}.

The third DSI is *shadow*. It is well-known that objects with some elevation such as buildings, trees, bridges etc. has an adjacent shadow region. Therefore, the regions adjacent to shadow regions are candidates of buildings. The shadows are illuminated as a result of atmospheric scattering of sunlight. Rayleigh scattering is considered as the most effective phenomena for modeling shadow [94]. Majority of the studies employ thresholding methods on color or NIR bands [99], [111], [27] while other studies form false color images to detect shadowed regions [7], [109]. In this study, a robust multi-spectral false color shadow detection method, suggested by Teke et. al. [109] is used to generate the Domain Specific Map for the shadow property. In this algorithm, NIR, R and G bands of the image is used to generate a false color image. Then, the image is converted to HSI color space and *ratiomap* is obtained as follows:

$$Ratiomap = \frac{S - I}{S + I},\tag{5.11}$$

where S is the normalized saturation and I is the normalized intensity. $Ratiomap$ is binarized by Otsu thresholding to obtain the domain specific map. This map includes both shadows and vegetation, hence vegetation areas are subtracted to obtain the shadow map which is utilized as domain specific map, DSM_{shadow}. A sample image and its DSM_{shadow} is depicted in *Figure 5.30*.

Segmentation fusion module determines the types of neighborhood relations between each pair of neighboring super-pixels to generate a domain specific RAG by fusing the information received from the bottom-up segmentation maps and domain specific

maps. For this purpose, the regions obtained at the output of bottom up segmentation algorithms are labeled if they overlap with the labeled regions on a DSM. Recall that the region labels in the DSMs correspond to the properties of rectangularity, shadow and vegetation. The set of $DSM_{connect}$ and $DSM_{disconnect}$ are defined as follows:

$$DSM_{connect} = \{(DSM_{rectangular}, DSM_{rectangular})\}, \qquad (5.12)$$

$$DSM_{disconnect} = \{(DSM_{rectangular}, DSM_{shadow}), (DSM_{rectangular}, DSM_{vegetation})\}. \qquad (5.13)$$

Each neighboring pair of super-pixels s_i and s_j in a bottom-up segmentation map is assigned by one of the following neighborhood relations:

- **Standard Neighborhood** is defined between two spatially adjacent super-pixels which do not have any labels obtained from the Domain Specific Maps. If one of the neighboring super-pixels is labeled by some DSM, m, while the other neighbor is not labeled by some DSM, n, which appears as (m, n) in the $DSM_{connect}$ or $DSM_{disconnect}$, then they are considered as standard neighbors. The set of standard neighbors is represented as N_{std}. The edge weights in N_{str} are kept as $w_{ij} = 1$ and the smoothness term is estimated by the Potts model.

- **Strong Neighborhood** is defined between two adjacent super-pixels, which are labeled by two DSMs that appear in the set of $DSM_{connect}$. For example, if several super-pixels remain inside of a rectangle, then, they are likely to be a part of a rectangle, forming a strong neighborhood. The set of strong neighbors is represented as N_{str}. The edge weights between the strong neighbors are updated by the relative frequency of s_i and s_j having the same label in the set of BUS maps and it is computed as,

$$f(s_i, s_j) = \frac{1}{B} \sum_{b=1}^{B} \delta(s_i, s_j), \qquad (5.14)$$

where δ is the Kronecker delta function which is equal to 1 when s_i and s_j have the same labels. The smoothness term for a pair of strong neighbors is estimated by the weighted Potts model. In this way, the super-pixels with strong neighborhood relations are encouraged strongly to have similar labels.

94

- **Weak Neighborhood** is defined between two adjacent super-pixels, which are labeled by two DSMs that appear as pairs in the set of $DSM_{disconnect}$. In the building segmentation problem, two neighboring super-pixels one of which is detected as belonging to a rectangle region and the other one is detected within a vegetation or shadow region, are in this group. The set of weak neighbors is represented as N_{weak}. The smoothness term between a pair of weak neighbors is estimated by the Kronecker delta function.

Note that the above information fusion procedure creates three types of smoothness terms for the domain specific RAG, each of which is defined over one of the neighborhood sets of $\{N_{std}\}$, $\{N_{str}\}$ and $\{N_{neg}\}$.

5.2.3.3 Domain Specific MRF Energy

The first term is defined as,

$$U_{data}(x) = \sum_{s_i \in S} \psi_i(l_{s_i}) \tag{5.15}$$

and it is modeled as negative log likelihood using CIELab color features.

Smoothness term for a pair of super-pixels s_i and s_j is weighted as follows:

$$w_{ij} = \begin{cases} f(s_i, s_j) & \text{if } (s_i, s_j) \in N_{str} \\ \beta & \text{if } (s_i, s_j) \in N_{weak} \\ 1 & otherwise \end{cases} \tag{5.16}$$

where, s_i represents super-pixels as MRF sites and l_{s_i} is the label assigned to that super-pixel. β is a constant such that $\beta \leq 0$. The first condition of the U_{smooth} function estimates the pairwise energy between a pair of strong neighbors, while the second condition estimates the energy for a pair of weak neighbors and the third condition estimates the energy for a pair of standard neighbors. So, the smoothness term of the MRF energy function is estimated with the following formula:

$$U_{smooth} = \alpha \times (\sum_{s_i \in S, s_j \in N_{str_{s_i}}} f(s_i, s_j) \times \psi_{ij}(l_{s_i}, l_{s_j}) + \sum_{s_i \in S, s_j \in N_{weak_{s_i}}} \beta \times \psi_{ij}(l_{s_i}, l_{s_j}) + \sum_{s_i \in S, s_j \in N_{std_{s_i}}} \psi_{ij}(l_{s_i}, l_{s_j}))$$

$$\tag{5.17}$$

Here, ψ is the Potts model which decreases energy when neighboring super-pixels have the same label.β is set as 0, hence the smoothness term is estimated over the set of standard and strong neighborhood pairs. Finally, the energy function is minimized by the expansion algorithm of [21].

5.2.3.4 Experimental Setup

In order to observe the contribution of the domain specific information on the performance of the unsupervised segmentation methods, the proposed domain specific building segmentation system is compared with the popular unsupervised segmentation methods, namely, watershed segmentation (WS), SLIC (Simple Linear Iterative Clustering) [4], Efficient Graph Based Segmentation (EG)[42] and Mean Shift segmentation (MS) [28]. In our experiments 10 satellite images from the QuickBird are employed. The size of the images are 500×500 with 60 cm spatial resolution and consist of 4 bands; red, green, blue and near infra-red.

Since segmentation is considered as a pre-processing step of a classification task, it is important to minimize the information loss due to the smoothing affect of the segmentation method. This requires over-segmentation with the cost of fragmentation of the objects in the scene.

MS's spatial and range bandwidth parameters are set as $\{3, 4, 5, 6, 7\}$ and minimum region size ($minArea$) is set as $\{100, 250, 400\}$. A set of EG segmentations is obtained by setting the smoothing parameter as $\sigma = \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, k value for the threshold function as $\{150, 300, 450, 600\}$ and the $minArea = \{100, 250, 400\}$. For each $minArea$ value, a corresponding MS and EG segmentation is selected among various outputs by employing Normalized Mutual Information (NMI) [36, 118]. NMI among all MS segmentations with same $minArea$ value, NMI is estimated and the segmentation which maximizes mutual information is selected. Parameter selection for EG is handled similarly.

Parameters for SLIC and WS are determined based on the results obtained by MS, EG and DS-MRF. SLIC has only one parameter, $numReg$ which is determined based on the average number of regions obtained by the other three algorithms, MS, EG

and DS-MRF. Watershed segmentation has a threshold parameter, which determines the level of suppression for shallow minima to avoid the over segmentation. This threshold value is set to $\{20, 30, 40, 50, 60, 70, 80, 90, 100\}$ and a set of segmentations is obtained. Later, the performance of WS is monitored for threshold values of $\{20, 50, 100\}$ which are also selected based on the average number of regions.

In the proposed domain specific building segmentation method, bottom-up segmentation partitions are obtained by the Mean Shift method, which is run with spatial and range bandwidth set as $\{3, 4, 5, 6, 7\}$, $minArea = \{100, 250, 400\}$ and α is set as 3.

5.2.3.5 Experimental Results and Quantitative Evaluation

In this experiment, the objective is to obtain an over-segmentation of a given image which can further be used for building detection. Therefore, representation of building regions is the critical issue in this problem. For features, to be highly representative of a building, they should be extracted from various parts of the building. Hence, buildings should be represented as large regions as possible without increasing false alarm.

In our experiments, two metrics are employed for performance evaluation. The first metric is the Precision in Region of Interest, $P - roi$, which is defined based on the precision value, commonly used in the image processing literature. The second metric is the Overall Segmentation Quality, OSQ, which is specifically defined for evaluating building segmentation performance by Senaras et. al. [100].

Precision in region of interest, $P - roi$, measures the accuracy of segmentation in the region of interest. For a given segmentation S_1 and its ground truth segmentation S_2, precision in region of interest is estimated as follows; for each region in S_2 , the set of corresponding pixels $\{p_a\}$ in S_1 and their corresponding labels $\{l_a\}$ in S_1 are determined. For each label in $\{l_a\}$, its precision is estimated as follows;

$$Precision = \frac{tp}{tp + fp},$$

(5.18)

where tp is the number of true positive pixels and fp is the number of false positive pixels.

For the building segmentation problem, each building region in the ground truth data is considered separately while estimating the precision. Then, $P - roi$ is estimated as the mean value of the precisions over all the buildings covered under the region of interest.

Overall Segmentation Quality, OSQ, is a boundary based metric, which is defined by two equations. The first equation, which is referred as Segmentation Fitting, is as follows:

$$SF = \frac{|\phi(D^b) \bigcap \vartheta^G|}{|\phi(D^b)|}, \tag{5.19}$$

where $D = \{s_i\}_{i=1}^{N}$ is a segmentation of a given image, I. In the ground truth of I, pixels are marked as building or background. The number of pixels in a segment s_i is represented as M_i, and the number of pixels corresponding to buildings in s_i are represented as M_i^b. And building segments are defined as $D^b = \{s_i : M_i^b > \lambda M_i\}$, where $\lambda \in \mathbf{R}$. $\phi(.)$ defines the set of pixels in the boundaries of segments and ϑ^G is the set of boundary pixels in the ground truth image and $|.|$ is the set cardinality.

The second equation is the Ground Truth Coverage, GTC which is defined as follows:

$$GTC = \frac{|\phi(D^b) \bigcap \vartheta^G|}{|\vartheta^G|}. \tag{5.20}$$

SF measures the accuracy of regions in representing buildings, while GTC measures the quality of representing buildings in the ground truth. In other words, SF is a measure of precision in representing building boundaries and GTC is a measure of recall for building boundaries in the ground truth. Using these two functions, OSQ is defined as follows:

$$OSQ = 2\frac{SF \times GTC}{SF + GTC}. \tag{5.21}$$

The segmentation results are observed in three groups, based on $minArea$ values. $minArea$ determines the scale of the segmentation result, and for a fair comparison the segmentation algorithms should be in approximately similar scales. $P - roi$ and OSQ for all algorithms at three scales are reported in Table 5.6.

Table 5.6: Comparison of DS, MS, WS, SLIC

Method	$minArea = 100$		$minArea = 250$		$minArea = 400$	
	P-roi	OSQ	P-roi	OSQ	P-roi	OSQ
WS	0.06	0.67	0.12	0.58	0.16	0.41
SLIC	0.40	0.79	0.31	0.66	0.18	0.42
EG	0.28	0.83	0.22	0.77	0.18	0.70
MS	0.44	0.92	0.31	0.81	0.24	0.75
Classic-MRF	0.18	0.84	0.13	0.75	0.11	0.69
Boosted-MRF	0.49	0.97	0.35	0.89	0.27	0.82
DS-MRF	**0.52**	**0.99**	**0.38**	**0.93**	**0.31**	**0.90**

Careful inspection of this table, reveals that regardless of the value assigned to $minArea$, the DS-MRF method achieves an OSQ of at least 0.90. It is also observed that the segmentation performance is higher for smaller values of $minArea$, which implies that over-segmentation is favored. Nevertheless, the objective of a segmentation algorithm is to obtain a semantically meaningful partitioning by avoiding over segmentation as much as possible. Moreover, increase in the number of regions obtained in the segmentation process also increases the complexity of a following recognition/detection process.

5.2.3.6 Sensitivity of the DS-MRF Method on the DSI

In this experiment, the effect of DSI on the DS-MRF is monitored. For this purpose, the DS-MRF is run several times, each time with a different set of DSI. First, it is assumed that no DSI is available, then it is assumed only one of rectangular shape, shadow and vegetation is available. Later, it is assumed that DSIs are available in pairs, and finally it is assumed that all three DSI is available. In all the runs, Mean Shift segmentation is employed for bottom-up segmentation with $minArea = 250$ and α is set as 3. Corresponding P-roi and OSQ values are reported in Table 5.7.

In these runs, no significant contribution is observed if only one DSI is available. Even, compared to the no DSI case, a decrease in the P-roi is observed in the runs with only one DSI. On the other hand, in these runs of DS-MRF, weak neighborhood
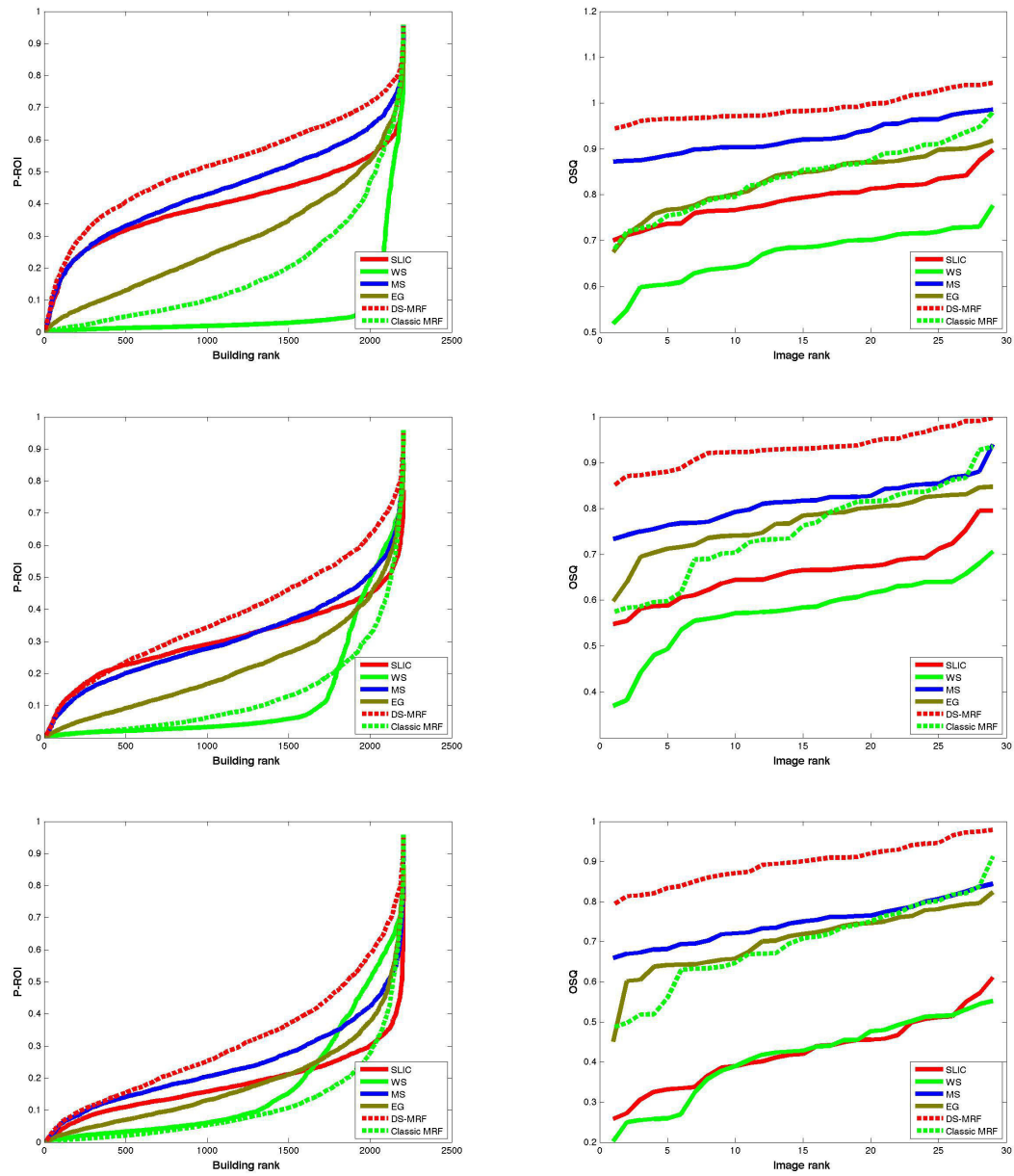
Figure 5.31: Performance comparisons by P-roi and OSQ.

Table 5.7: The effect of available DSI on the performance of DS-MRF

DSM	P-roi	OSQ
{}	0.3499	0.8878
$\{DSM_{rectangles}\}$	0.3275	0.8875
$\{DSM_{shadow}\}$	0.3320	0.8846
$\{DSM_{vegetation}\}$	0.3248	0.8716
$\{DSM_{shadow}, DSM_{vegetation}\}$	0.3461	0.8895
$\{DSM_{rectangles}, DSM_{shadow}\}$	0.3658	0.9180
$\{DSM_{rectangles}, DSM_{vegetation}\}$	0.3493	0.8917
$\{DSM_{rectangles}, DSM_{shadow}, DSM_{vegetation}\}$	0.3858	0.9313

is included only if rectangular shape information is available together with at least one of the shadow or the vegetation information. Hence, in the last three rows of Table 5.7, weak neighborhood relation is utilized, while in the other rows it is not employed. The contribution of weak neighborhood relation is significantly observed in the fifth and the seventh rows, where both rectangular shape and shadow information is available.

5.2.3.7 Visual Inspection of the Sample Segmentation Examples

Sample segmentation results of the DS-MRF method and the other five methods MS, EG, WS, SLIC and Classic-MRF are provided in Figures 5.32, 5.33, 5.34, 5.35, 5.36, 5.37 and 5.38 for visual inspection.

Figures 5.32, 5.33 and 5.34 are the segmentation outputs of the same image with three different values of the $minArea$ parameter. In these figures, segmentation results of the six algorithms are provided for $minArea$ values of 100, 250 and 400 respectively. The results in Figure 5.32 are highly over-segmented. For this reason, the contribution of DS-MRF in terms of improving the semantical meaning can not be monitored. On the other hand, the contribution of the DS-MRF method can be clearly observed in Figure 5.33 and 5.34. The main objective of a building segmentation method is to obtain a segmentation which partitions buildings into small number of regions (so that the building regions are not partitioned into multiple regions), while keeping false positives as small as possible. It is also desired that the non-building parts are
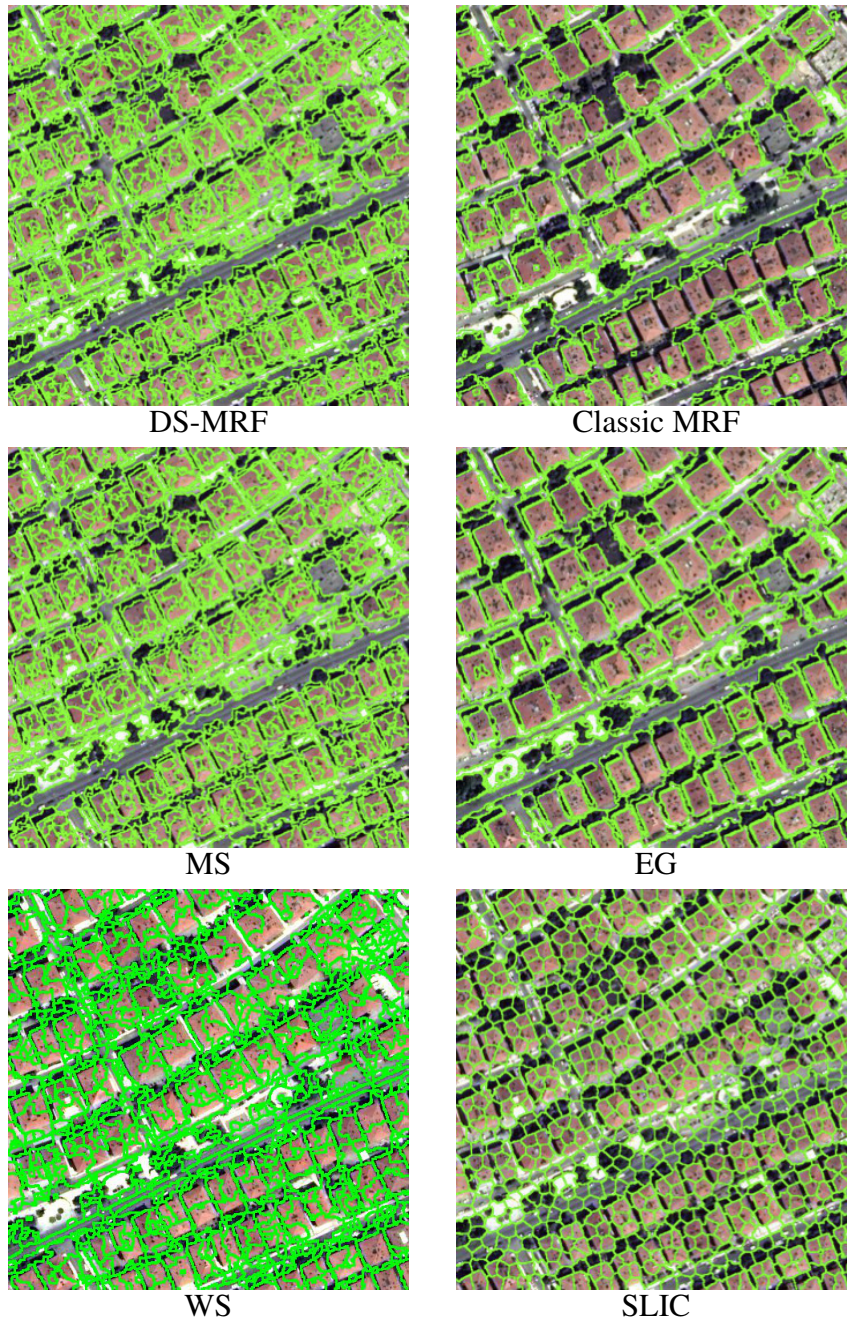
Figure 5.32: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 100$ on sample image 1.
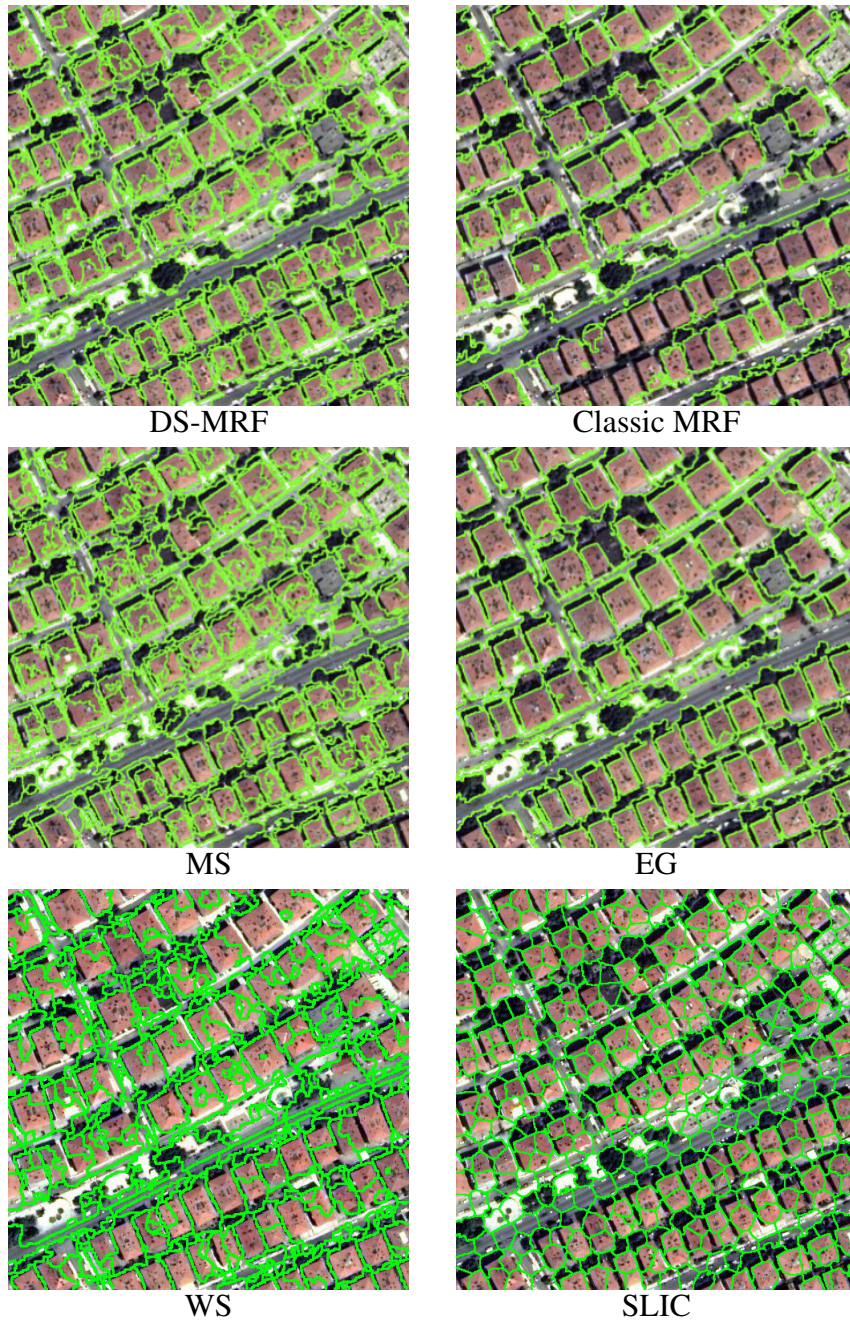
Figure 5.33: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 250$ on sample image 1.
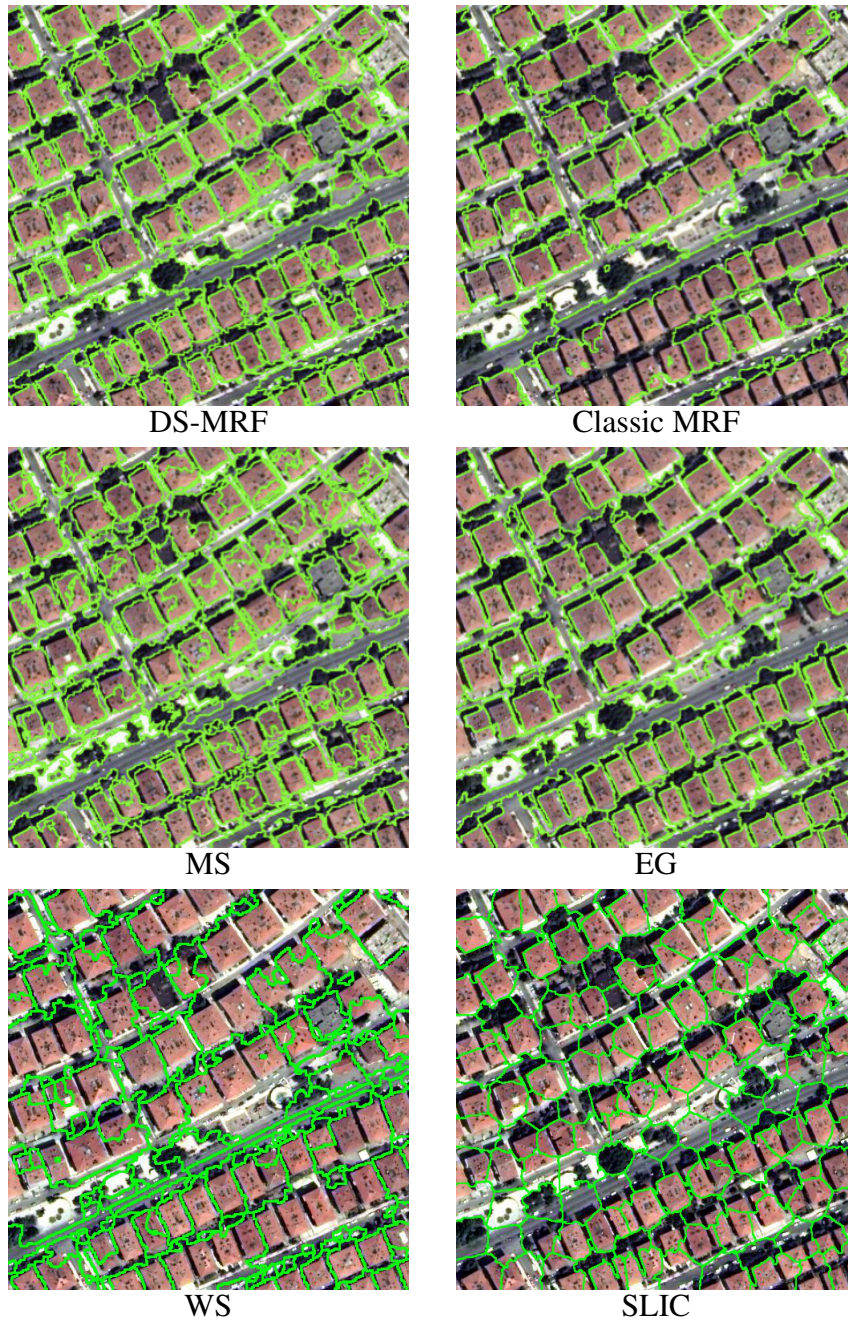
Figure 5.34: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 400$ on sample image 1.

DS-MRF   Classic MRF

MS   EG

WS   SLIC
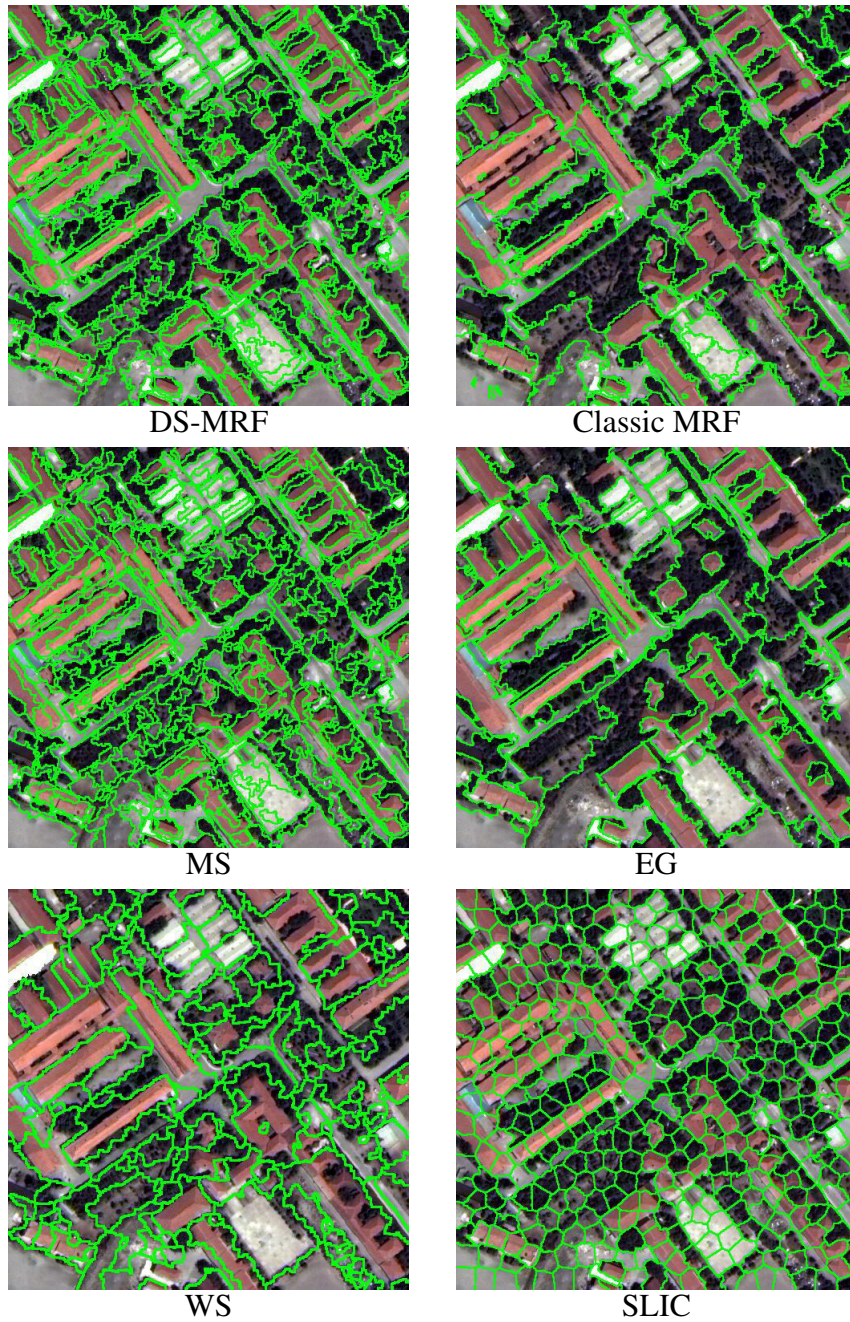
Figure 5.35: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 250$ on sample image 2.
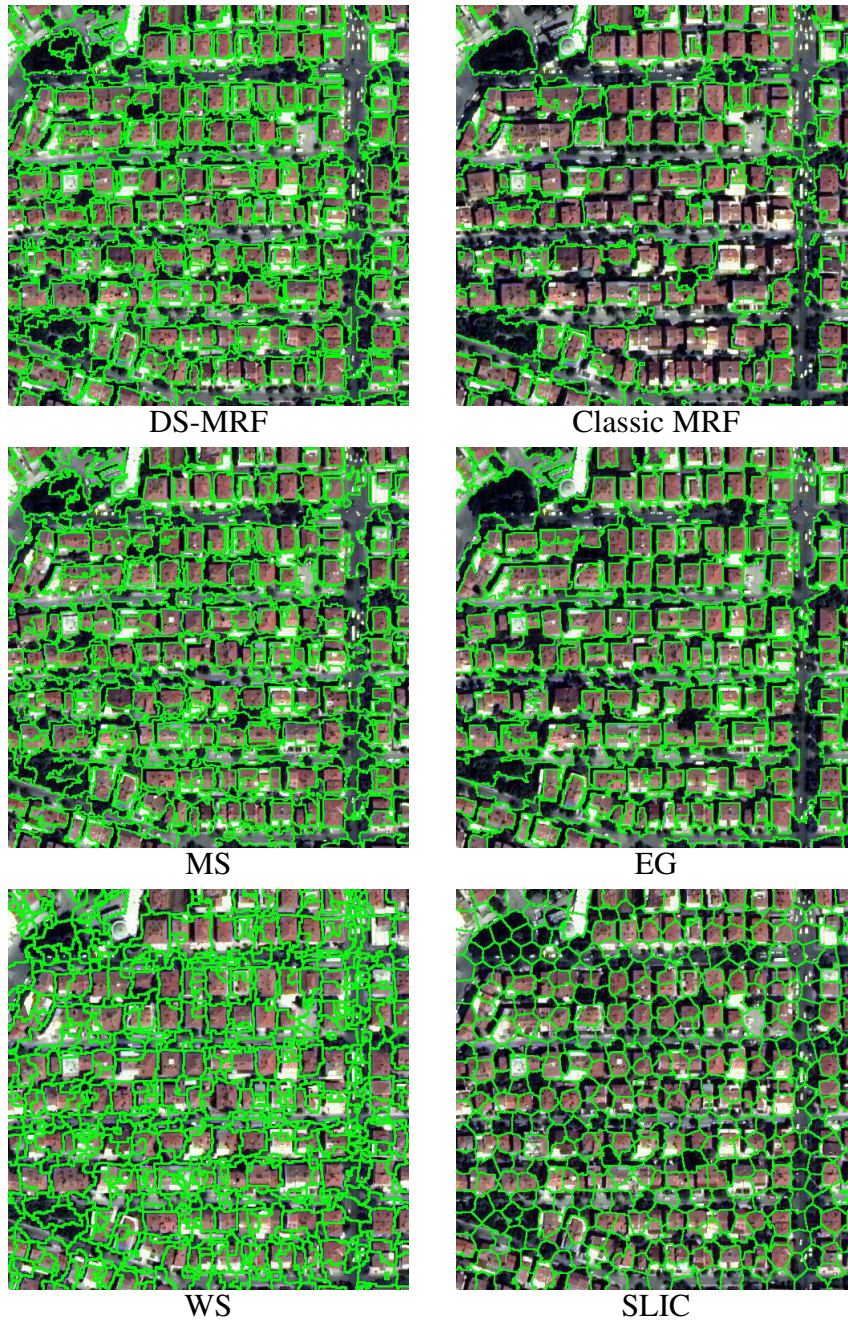
Figure 5.36: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 250$ on sample image 3.
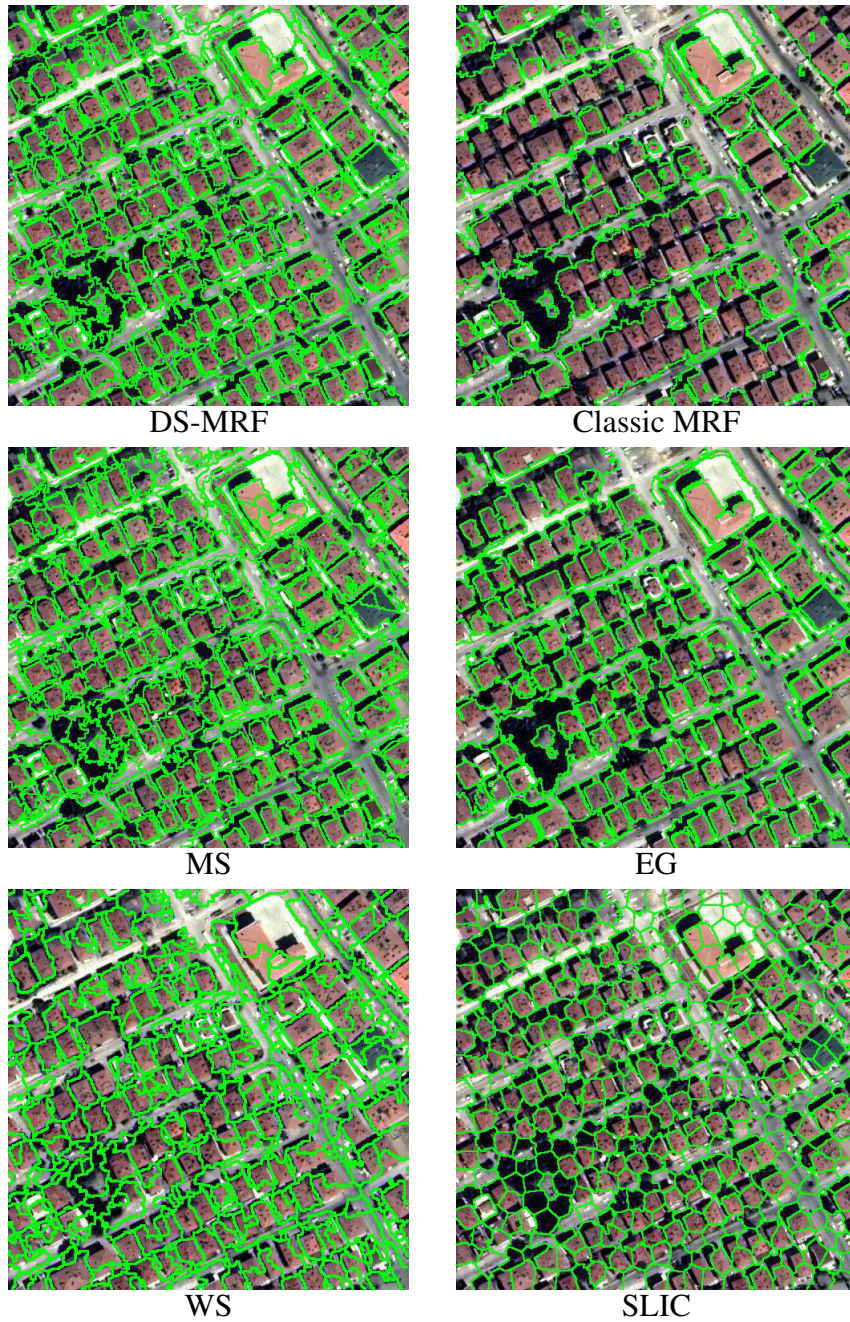
Figure 5.37: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 250$ on sample image 4.
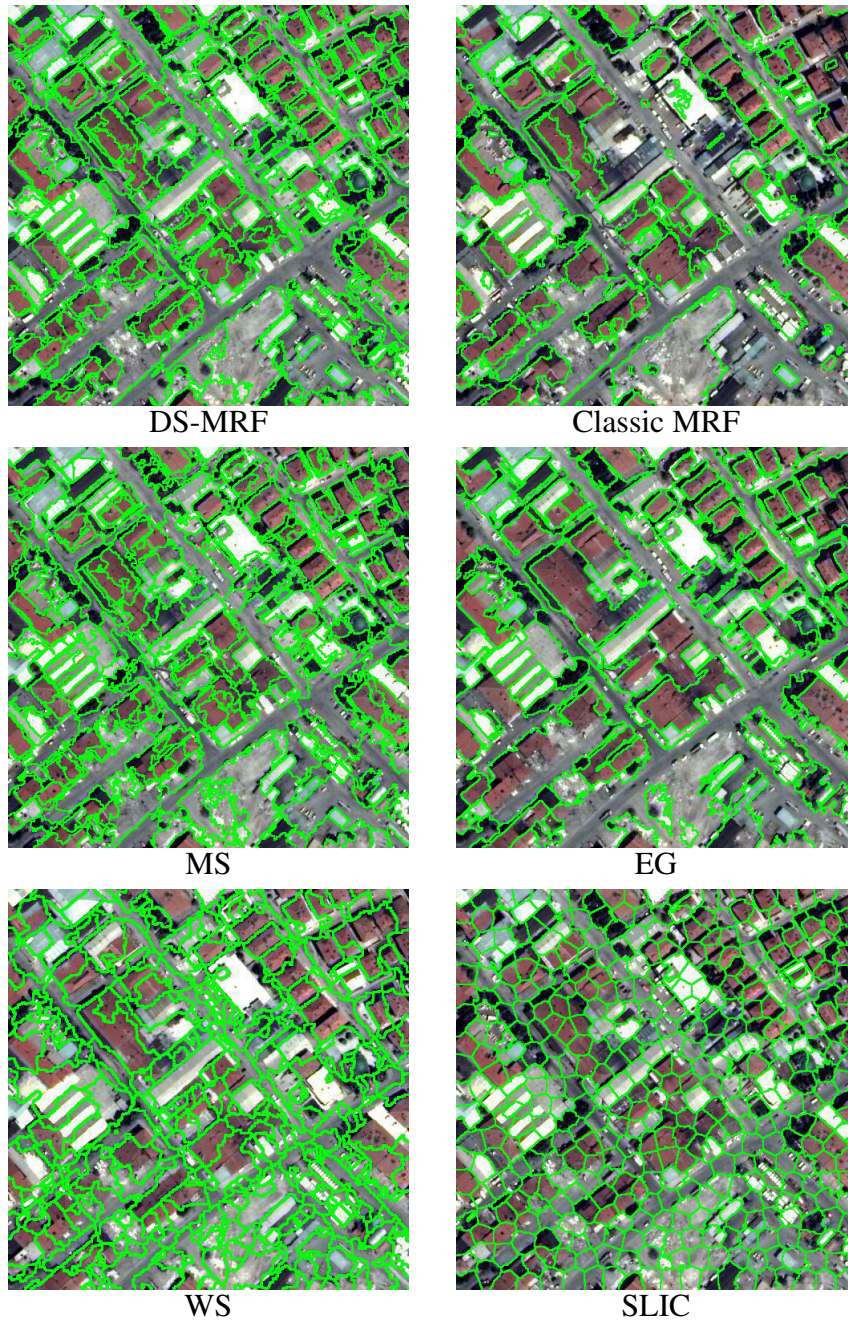
Figure 5.38: Comparison of DS-MRF, Classic MRF, MS, EG, WS and SLIC for $minArea = 250$ on sample image 5.

not partitioned into many regions. Considering these objectives, the advantage of the DS-MRF method can be most clearly observed in Figure 5.34. In this figure, MS or EG generates under-segmented images for $minArea = 400$, while DS-MRF obtains a less under-segmented image which can be further employed for building classification/recognition.

More examples are provided in Figures 5.35, 5.36, 5.37 and 5.38 for $minArea = 250$ on other images from the dataset.

5.3   Chapter Summary

Performance of DS-MRF is compared to state of the art segmentation systems through three experimental setups. In the first experiment, DS-MRF, EG, Multiscale NCut, MS, ROI-SEG, CDHS and DS-MRF algorithms are applied on a subset of MSRC dataset, which consists of outdoor images with vegetation. Thus, NDI is employed as DSI for vegetation detection. Algorithms are compared through PRI and GCE measures and satisfactory results are obtained. In the second experiment; DS-MRF, EG, Multiscale NCut, MS and ROI-SEG are applied on airport images. In this experiment, the goal is to obtain an oversegmentation which can later be used for airport detection. For this application domain, NDVI is employed as DSI on vegetation and PLBR is employed as DSI on runway. Performance of the algorithms are compared through precision in region of interest and f-score measures, and it is quantitatively observed that DS-MRF performed better than other algorithms. Besides, performance increase by DS-MRF is also observed in visual analysis. In the third experiment, the proposed system is applied on urban area images, where the goal is to obtain superpixels which can later be used for building detection. Three distinct DSI that is employed in this domain are; rectangularity of buildings, NDVI related to vegetation and shadow. In this experiment, performance of DS-MRF is compared with EG, SLIC and MS which are commonly used for obtaining superpixels. Here again, performances are compared via precision in region of interest and f-score and it is observed that DS-MRF outperforms state of the art algorithms both quantitatively and visually.

CHAPTER 6

CONCLUSION

In this thesis, two segmentation fusion methods, namely, Boosted-MRF and Domain Specific Segmentation (DS-MRF), are proposed using the Markov Random Fields (MRF) framework. Basically, Boosted MRF fuses the outputs of multiple segmentation outputs under the smoothing term of an unsupervised MRF model. On the other hand, DS-MRF employs expert information about the domain in order to fuse the multiple segmentation outputs. Both methods do not employ any labeled training data. Therefore, the suggested methods are considered as unsupervised segmentation. However, since DS-MRF incorporates some generic domain specific information, it may be considered as a special case of semantic segmentation.

The suggested segmentation fusion methods utilizes the MRF model is slightly different then the available studies in the segmentation literature, such as [70, 54, 53, 105, 75, 78, 120, 51, 92, 30, 32, 67, 68, 69, 96, 60]. In these studies, the image is either partitioned without assigning the class labels, or the pixels are labeled by a predefined index. The latter problem is referred as the semantic segmentation, and studies in this group assume that a labeled dataset is available. However, the available information does not necessarily appear in the form of a class label, but sometimes it appears in a more generic form, such as "there is a sky region at the top of all the images". There are many segmentation studies which incorporate this type of prior information into the segmentation process [37, 56, 57, 38]. However, majority of these studies incorporate one or two types of prior information mostly related to shape, appearance and location into a single segmentation scheme to improve the semantic content of the segmentation outputs. The approach suggested in this study, enables us to improve

the semantic content of the segmentation by fusing the outputs with respect to the available information. Depending on the type and degree of the available Domain Specific Information, the segmentation fusion technique is adjusted. If no information is available then, the fusion is achieved my maximizing the consensus among the outputs, defined over the MRF energy function.

On the other hand, if some priori information is available, then, this information is incorporated in the MRF energy function by generating a set of Domain Specific Maps. In this scheme, it is assumed that prior information about the given problem domain is available. This prior information is referred as domain specific information (DSI), and provides more general information than specific class labels. Before introducing the proposed domain specific segmentation, a segmentation fusion method, called Boosted-MRF is introduced. This method constitutes the basics of the proposed domain specific segmentation method.

The Boosted-MRF is an unsupervised segmentation method based on the fusion of a set of segmentation outputs, is proposed. Segmentation fusion is commonly invoked to resolve the parameter selection problem in the unsupervised segmentation problems, and various methods for segmentation fusion are proposed [25, 55, 47]. A group of these methods employ ensemble clustering methods while another group of studies employ graphs for the fusion of segmentation outputs. Although the advantage of the MRF model in integrating various information into the segmentation process by means of the MRF energy function is emphasized in the literature [81], as far as we know the MRF energy function is not employed for the segmentation fusion problem. In the proposed Boosted-MRF method the relative frequency of co-occurrences of neighboring pixels in the same region over a set of initial segmentations is estimated and this information is integrated into the MRF energy through the smoothness term. The performance of the Boosted-MRF method is analyzed by using two different types of segmentation methods in the baseline and the performances are compared to the performance of an ensemble clustering method, called Best Of K (BOK). For the purpose of comparison, three measures; the Probabilistic Rand Index (PRI), the Global Consistency Error (GCE) and the Mutual Information are employed. It is observed that the Boosted-MRF is a powerful method for combining unsupervised segmentation outputs. It is observed that the resulting segmentation performs better

than the individual segmentation outputs in the ensemble. Also, the Boosted-MRF does not require any parameter selection or optimization.

As we mentioned earlier DS-MRF requires a set of Domain Specific Information (DSI). This information is used to fuse the segmentation outputs in a new MRF energy function. For this purpose, the architecture and the MRF energy function of the Boosted-MRF method is updated such that the segmentation process is constrained by the available DSI. This is accomplished by, defining a logical predicate for the DSI and generating a corresponding domain specific map (DSM) which is an incomplete partition of the image. Then, the DSMs are utilized for defining three types of neighborhood relations; the standard neighborhood, the strong neighborhood and the negative neighborhood. Finally, the MRF energy function is updated to include three types of relations with different smoothing functions.

The proposed DS-MRF methods is employed in two problem domains. The first domain is a dataset of outdoor images with vegetation background and DSI related to vegetation detection is employed in this domain. The second domain is a remote sensing application for building detection. In this problem, more advanced DSI is available from the remote sensing literature. Prior information related to the rectangular shape of buildings, the shadow and the vegetation information are all well defined DSI which are commonly used in the building detection problem. Hence, this type of information is employed for defining the neighborhood relations. It is observed experimentally by using specific evaluation metrics; precision in region of interest and the overall segmentation quality, that the integration of DSI in the MRF model significantly improves the segmentation performance.

The major advantage of the DS-MRF method is that it can be extended to include various forms of DSI. Nevertheless, it is very critical that the provided DSI is to be formulated correctly by means of the predicates and corresponding DSMs can be obtained properly, as the method is very sensitive to the DSI. A representation of DSM may significantly improve the overall performance of the segmentation if it reflects the DSI to represent the context of a subset of regions in the map. However, it may decrease the performance, if the DSI is not valid in the entire image dataset. Therefore, the design of the DSI, together with the extraction of the DSMs lie at the

heart of the suggested segmentation fusion method.

6.1 Future Work

The Boosted-MRF and the DS-MRF methods can be extended by incorporating the segmentation output as a feed-back in the form of a new baseline segmentation output. In this way, a successive increase in the segmentation performance can be observed.

We are planning to employ the DS-MRF method on other problems in different domains, such as medical image analysis. The application of the DS-MRF method on different problem domains will probably require the definition of new DSI types and may be new neighborhood relations.

The application of the DS-MRF method on the building segmentation problem is extended in [90] to cover the detection performance where the decision fusion method of Senaras and Ozay [100] is employed with a modification in the pre-processing and the segmentation steps. We are planning to employ DS-MRF on other remote sensing problems such as urban detection, coastline detection, airport detection etc.

The DS-MRF method is very suitable for domain specific datasets, such as medical images, botanical or zoological datasets, criminal datasets etc., where well-defined set of DSIs is usually available. Prior information related to the shape, appearance and locality can all be unified by means of the MRF energy function in these domain specific image analysis problems.

Automatic extraction of domain specific information for a given dataset is also an interesting research area related to the domain specific segmentation problem. In this study, it is assumed that this information is provided a-priori. It is formulated mathematically and integrated into the segmentation process by means of the proposed domain specific segmentation method. If domain information can be extracted automatically, domain specific segmentation will probably attract the attention of more researchers in the future.

On the other hand, the application of the MRF model on segmentation fusion can be extended in the Boosted-MRF method to include different types of segmentation

114

methods, including the MRF method itself, in the baseline segmentation. Moreover, the fusion function can be formulated such that different reliabilities are assigned to the baseline segmentation outputs, or a new fusion function can be defined.

Application of the DS-MRF method on a video processing problem would arise interesting forms of domain specific information since deriving higher level information in the video processing problem is easier. Due to the availability of more domain specific information, more remarkable results can be obtained in an video processing application.

REFERENCES

[1] Preprocessing transformations and their effects on multispectral recognition. In *the Sixth International Symposium on Remote Sensing of Environment, Volume II*, 1969.

[2] A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1):62–66, 1979.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012.

[4] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC Superpixels. Technical report, EPFL, 2010.

[5] R. Adams and L. Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6):641–647, 1994.

[6] H.G. Akcay and S. Aksoy. Building detection using directional spatial constraints. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 1932–1935, 2010.

[7] Yagiz Aksoy and A. Aydin Alatan. Utilization of false color images in shadow detection. In *Proceedings of the 12th International Conference on Computer Vision - Volume 2*, ECCV'12, pages 472–481, Berlin, Heidelberg, 2012. Springer-Verlag.

[8] Luigi Ambrosio and Vincenzo Maria Tortorelli. Approximation of functional depending on jumps by elliptic functional via t-convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990.

[9] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916, 2011.

[10] T. Athanasiadis, P. Mylonas, Y. Avrithis, and S. Kollias. Semantic image segmentation and object labeling. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(3):298–312, 2007.

[11] Thanos Athanasiadis, Phivos Mylonas, and Yannis Avrithis. A contextbased region labeling approach for semantic image segmentation. In *In Semantic*

Multimedia, First International Conference on Semantics and Digital Media Technologies, Lecture Notes in Computer Science, pages 212–225. Springer, 2006.

[12] Rodney J. Baxter. *Exactly solved models in statistical mechanics*. Academic Press, London, 1982.

[13] Petr Berka, Thanos Athanasiadis, and Yannis Avrithis. Rule-based reasoning for semantic image segmentation and interpretation.

[14] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 1974.

[15] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.

[16] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[17] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

[18] E. Borenstein and S. Ullman. Combined top-down/bottom-up segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(12):2109–2125, 2008.

[19] Eran Borenstein and Shimon Ullman. Class-specific, top-down segmentation. In *Proceedings of the 7th European Conference on Computer Vision-Part II*, ECCV '02, pages 109–124, London, UK, UK, 2002. Springer-Verlag.

[20] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, 2004.

[21] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[22] Bo cai Gao. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257 – 266, 1996.

[23] Liangliang Cao and Fei-Fei Li. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. In *ICCV*, pages 1–8. IEEE, 2007.

[24] T.F. Chan and L.A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, 2001.

[25] Kyujin Cho and Peter Meer. Image segmentation from consensus information. *Computer Vision and Image Understanding*, 68(1):72 – 89, 1997.

[26] C.M. Christoudias, B. Georgescu, and P. Meer. Synergism in low level vision. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 4, pages 150–155 vol.4, 2002.

[27] Kuo-Liang Chung, Yi-Ru Lin, and Yong-Huai Huang. Efficient shadow detection of color aerial images based on successive thresholding scheme. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(2):671–682, Feb 2009.

[28] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, May 2002.

[29] T. Cour, F. Benezit, and Jianbo Shi. Spectral segmentation with multiscale graph decomposition. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1124–1131 vol. 2, 2005.

[30] Huawu Deng and D.A. Clausi. Unsupervised image segmentation using a simple mrf model with a new implementation scheme. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 691–694 Vol.2, 2004.

[31] Yining Deng and B.S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(8):800–810, 2001.

[32] H. Derin and Howard Elliott. Modeling and segmentation of noisy and textured images using gibbs random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(1):39–55, 1987.

[33] Ringo Doe. Multiscale normalized cut code., June 2009.

[34] Su Dongcai. Efficient graph based segmentation code @mathworks, November 2010.

[35] M. Donoser and H. Bischof. Roi-seg: Unsupervised color segmentation by combining differently focused sub results. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.

[36] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley, New York, 2 edition, 2001.

[37] Ayman El-Baz and G. Gimel'farb. Image segmentation with a parametric deformable model using shape and appearance priors. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.

[38] M. El-Melegy and H. Mokhtar. Incorporating prior information in the fuzzy c-mean algorithm with application to brain tissues segmentation in mri. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3393–3396, 2009.

[39] E. Erdem, S. Tari, and L. Vese. Segmentation using the edge strength function as a shape prior within a local deformation model. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2989–2992, 2009.

[40] Erkut Erdem, Aykut Erdem, and Sibel Tari. Edge strength functions as shape priors in image segmentation. In Anand Rangarajan, Baba Vemuri, and AlanL. Yuille, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 3757 of *Lecture Notes in Computer Science*, pages 490–502. Springer Berlin Heidelberg, 2005.

[41] Francisco J. Estrada and Allan D. Jepson. Benchmarking image segmentation algorithms. *Int. J. Comput. Vision*, 85(2):167–181, November 2009.

[42] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, September 2004.

[43] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient matching of pictorial structures. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 66–73 vol.2, 2000.

[44] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient belief propagation for early vision. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–261–I–268 Vol.1, 2004.

[45] Boris Flach and Dmitrij Schlesinger. Combining shape priors and mrf-segmentation. In Niels Vitoria Lobo, Takis Kasparis, Fabio Roli, JamesT. Kwok, Michael Georgiopoulos, GeorgiosC. Anagnostopoulos, and Marco Loog, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 5342 of *Lecture Notes in Computer Science*, pages 177–186. Springer Berlin Heidelberg, 2008.

[46] Lucas Franek, DanielDuarte Abdala, Sandro Vega-Pons, and Xiaoyi Jiang. Image segmentation fusion using general ensemble clustering methods. In Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, editors, *Computer Vision – ACCV 2010*, volume 6495 of *Lecture Notes in Computer Science*, pages 373–384. Springer Berlin Heidelberg, 2011.

[47] A.L.N. Fred and A.K. Jain. Combining multiple clusterings using evidence accumulation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):835–850, 2005.

[48] R. Gaetano, G. Masi, G. Scarpa, and G. Poggi. A marker-controlled watershed segmentation: Edge, mark and fill. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 4315–4318, 2012.

[49] Stuart Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(6):721–741, 1984.

[50] Aristides Gionis, Heikki Mannila, and Panayiotis Tsaparas. Clustering aggregation. In *Proceedings of the 21st International Conference on Data Engineering*, ICDE '05, pages 341–352, Washington, DC, USA, 2005. IEEE Computer Society.

[51] M. Goktepe, V. Atalay, N. Yalabik, and C. Yalabik. Unsupervised texture based image segmentation by simulated annealing using markov random field and potts models. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, volume 1, pages 820–822 vol.1, 1998.

[52] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[53] S. Gould, R. Fulton, and D. Koller. Decomposing a scene into geometric and semantically consistent regions. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1–8, 2009.

[54] Stephen Gould, Jim Rodgers, David Cohen, Gal Elidan, and Daphne Koller. Multi-class segmentation with relative location prior. *Int. J. Comput. Vision*, 80(3):300–316, December 2008.

[55] Leo Grady. Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(11):1768–1783, November 2006.

[56] V. Grau, A. U J Mewes, M. Alcaniz, R. Kikinis, and S.K. Warfield. Improved watershed transform for medical image segmentation using prior information. *Medical Imaging, IEEE Transactions on*, 23(4):447–458, 2004.

[57] Ghassan Hamarneh and Xiaoxing Li. Watershed segmentation using prior shape and appearance knowledge. *Image and Vision Computing*, 27(1–2):59 – 68, 2009. Canadian Robotic Vision 2005 and 2006.

[58] A. Hanbury and J. Stottinger. On segmentation evaluation metrics and region counts. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, 2008.

[59] John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition, 1975.

[60] Xuming He, R.S. Zemel, and M.A. Carreira-Perpindn. Multiscale conditional random fields for image labeling. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–695–II–702 Vol.2, 2004.

[61] Xiaoying Jin and Curt H. Davis. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP J. Appl. Signal Process.*, 2005:2196–2206, January 2005.

[62] Miyoun Jung, X. Bresson, T.F. Chan, and L.A. Vese. Nonlocal mumford-shah regularizers for color image restoration. *Image Processing, IEEE Transactions on*, 20(6):1583–1598, 2011.

[63] Kittipat Kampa, José C. Príncipe, Duangmanee Putthividhya, and Anand Rangarajan. Data-driven tree-structured bayesian network for image segmentation. In *ICASSP*, pages 2213–2216. IEEE, 2012.

[64] G. Karypis, R. Aggarwal, V. Kumar, and Shashi Shekhar. Multilevel hypergraph partitioning: applications in vlsi domain. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 7(1):69–79, 1999.

[65] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.*, 20(1):359–392, December 1998.

[66] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[67] Zoltan Kato, Ting chuen Pong, and Song Guo Qiang. Unsupervised segmentation of color textures images using a multi-layer mrf model," presented at the. In *10th IEEE Int. Conf. Image Processing*, pages 961–964, 2003.

[68] Zoltan Kato and Ting-Chuen Pong. A markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24(10):1103 – 1114, 2006.

[69] Zoltan Kato, Ting-Chuen Pong, and John Chung-Mong Lee. Color image segmentation and parameter estimation in a markovian framework. *Pattern Recognition Letters*, 22(3–4):309 – 321, 2001.

[70] P. Kohli, L. Ladický, and P. H S Torr. Robust higher order potentials for enforcing label consistency. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.

[71] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[72] V. Kolmogorov and M. J. Wainwright. On the Optimality of Tree-reweighted Max-product Message-passing. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI-2005)*. Morgan Kaufmann Publishers, 2005.

[73] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159, 2004.

[74] M. Pawan Kumar, P.H.S. Torr, and A. Zisserman. Objcut: Efficient segmentation using top-down and bottom-up cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):530–545, 2010.

[75] Lubor Ladicky, Chris Russell, Pushmeet Kohli, and Philip H. S. Torr. Graph cut based inference with co-occurrence statistics. In *Proceedings of the 11th European conference on Computer vision: Part V*, ECCV'10, pages 239–253, Berlin, Heidelberg, 2010. Springer-Verlag.

[76] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[77] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(8):799–813, 1989.

[78] D. Larlus and F. Jurie. Combining appearance models and markov random fields for category level object segmentation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7, 2008.

[79] Anat Levin and Yair Weiss. Learning to combine bottom-up and top-down segmentation. In Ales Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision - ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV*, volume 3954 of *Lecture Notes in Computer Science*, pages 581–594. Springer, 2006.

[80] Fei-Fei Li and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, CVPR '05, pages 524–531, Washington, DC, USA, 2005. IEEE Computer Society.

[81] S. Z. Li. Markov random field models in computer vision, 1994.

[82] Wei-Ying Ma and B.S. Manjunath. Edgeflow: a technique for boundary detection and image segmentation. *Image Processing, IEEE Transactions on*, 9(8):1375–1388, 2000.

[83] J. Malik, L. Bourdev, Chunhui Gu, S. Gupta, B. Hariharan, and P. Arbelaez. Semantic segmentation using regions and parts. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 0:3378–3385, 2012.

[84] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[85] Takeshi Motohka, Kenlo Nishida Nasahara, Hiroyuki Oguma, and Satoshi Tsuchida. Applicability of green-red vegetation index for remote sensing of vegetation phenology. *Remote Sensing*, 2(10):2369–2387, 2010.

[86] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. on Pure and Applied Mathematics*, (5):577–685.

[87] H. P. Ng, S. H. Ong, K. W C Foong, P. S. Goh, and W.L. Nowinski. Medical image segmentation using k-means clustering and improved watershed algorithm. In *Image Analysis and Interpretation, 2006 IEEE Southwest Symposium on*, pages 61–65, 2006.

[88] Hieu Tat Nguyen, M. Worring, and R. Van den Boomgaard. Watersnakes: energy-driven watershed segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(3):330–342, 2003.

[89] A.O. Ok, C. Senaras, and B. Yuksel. Automated detection of arbitrarily shaped buildings in complex environments from monocular vhr optical satellite imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(3):1701–1717, 2013.

[90] Yarman Vural F.T. Oztimur Karadag O., Senaras C. Segmentation and Decision Fusion For Building Detection. In *submitted to the Geoscience and Remote Sensing IEEE International Symposium*, 20104.

[91] David Martin Pablo Arbelaez, Charkes Fowlkes. The berkeley segmentation dataset and benchmark. `www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/`, 2001.

[92] D.K. Panjwani and G. Healey. Markov random field models for unsupervised segmentation of textured color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(10):939–954, 1995.

[93] Caroline Pantofaru and Martial Hebert. A comparison of image segmentation algorithms. Technical report, 2005.

[94] Airton Marco Polidório, Franklin César Flores, Nilton Nobuhiro Imai, Antonio M. G. Tommaselli, and Clélia Franco. In *SIBGRAPI*, pages 270–277. IEEE Computer Society.

[95] Renfrey B. Potts. Some Generalized Order-Disorder Transformation. In *Transformations, Proceedings of the Cambridge Philosophical Society*, volume 48, pages 106–109, 1952.

[96] Mukta Prasad, Andrew Zisserman, Andrew Fitzgibbon, M. Pawan Kumar, and P. H. S. Torr. Learning class-specific edges for object detection and segmentation. In *Proceedings of the 5th Indian Conference on Computer Vision, Graphics and Image Processing*, ICVGIP'06, pages 94–105, Berlin, Heidelberg, 2006. Springer-Verlag.

[97] A.J. Pérez, F. López, J.V. Benlloch, and S. Christensen. Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture*, 25(3):197 – 212, 2000.

[98] P. Saeedi and H. Zwick. Automatic building detection in aerial and satellite images. In *Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on*, pages 623–629, 2008.

[99] P. Sarabandi, F. Yamazaki, M. Matsuoka, and A. Kiremidjian. Shadow detection and radiometric restoration in satellite high resolution images. In *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, volume 6, pages 3744–3747 vol.6, Sept 2004.

[100] C. Senaras, M. Ozay, and F.T. Yarman Vural. Building detection with decision fusion. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(3):1295–1304, 2013.

[101] Caglar Senaras. *Self-Supervised Decision Fusion For Building Detection*. PhD thesis, Middle East Technical University, Ankara, Turkey, September 2013.

[102] Eitan Sharon, Meirav Galun, Dahlia Sharon, Ronen Basri, and Achi Brandt. Hierarchy and adaptivity in segmenting visual scenes. *Nature*, 442(7104):810–3, 2006.

[103] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 1997.

[104] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.

[105] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *In ECCV*, pages 1–15, 2006.

[106] George Stockman and Linda G. Shapiro. *Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 2001.

[107] Alexander Strehl and Joydeep Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3:583–617, March 2003.

[108] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, Aseem Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6):1068–1080, 2008.

[109] Mustafa Teke, Emre Başeski, Ali Özgün Ok, Baríş Yüksel, and Çağlar Şenaras. Multi-spectral false color shadow detection. In *Proceedings of the 2011 ISPRS conference on Photogrammetric image analysis*, PIA'11, pages 109–119, Berlin, Heidelberg, 2011. Springer-Verlag.

[110] A. Topchy, A.K. Jain, and W. Punch. Clustering ensembles: models of consensus and weak partitions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1866–1881, 2005.

[111] V.J.D. Tsai. A comparative study on shadow compensation of color aerial images in invariant color models. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(6):1661–1671, June 2006.

[112] Compton J. Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127 – 150, 1979.

[113] R. Unnikrishnan, C. Pantofaru, and M. Hebert. A measure for objective evaluation of image segmentation algorithms. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 34–34, 2005.

[114] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):929–944, June 2007.

[115] R.G. von Gioi, J. Jakubowicz, J. M Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(4):722–732, 2010.

[116] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. *IEEE Transactions on Information Theory*, 51:3697–3717, 2002.

[117] Xi Wang, Chunyu Yang, and Jie Zhou. Clustering aggregation by probability accumulation. *Pattern Recogn.*, 42(5):668–675, May 2009.

[118] P. Wattuya, K. Rothaus, J.-S. Prassni, and X. Jiang. A random walker based approach to combining multiple segmentations. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, 2008.

[119] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*, ICCV '05, pages 1800–1807, Washington, DC, USA, 2005. IEEE Computer Society.

[120] Chee Won and H. Derin. Segmentation of noisy textured images using simulated annealing. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87.*, volume 12, pages 563–566, 1987.

[121] Lin Yang, P. Meer, and D.J. Foran. Multiple class segmentation using a unified framework over mean-shift patches. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.

[122] Hui Zhang, Jason E. Fritts, and Sally A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260 – 280, 2008.

[123] Bin Zhao, Li Fei-Fei, and Eric P. Xing. Image segmentation with topic random field. In *Proceedings of the 11th European conference on Computer vision: Part V*, ECCV'10, pages 785–798, Berlin, Heidelberg, 2010. Springer-Verlag.

[124] Haipeng Zheng, S.R. Kulkarni, and H.V. Poor. Consensus clustering: The filtered stochastic best-one-element-move algorithm. In *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pages 1–6, 2011.

[125] S.C. Zhu, T. S. Lee, and A.L. Yuille. Region competition: unifying snakes, region growing, energy/bayes/mdl for multi-band image segmentation. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 416–423, 1995.

[126] Örsan Aytekın, Arzu Erener, İlkay Ulusoy, and Şebnem Düzgün. Unsupervised building detection in complex urban environments from multispectral satellite imagery. *International Journal of Remote Sensing*, 33:2152–2177, 2012.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Öztimur Karadağ, Özge
Nationality: Turkish (TC)
Date and Place of Birth: 20.05.1982, Ankara
Marital Status: Married
Phone: 90 533 345 42 32
Fax: 90 242 518 20 51

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Department of Computer Engineering, Middle East Technical University	2008
B.S.	Department of Computer Engineering, Middle East Technical University	2005
High School	Alanya A.M.E Anatolian High School	2000

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2005-2006	HAVELSAN A.Ş.	Software Engineer
2006-2008	Department of Computer Engineering Middle East Technical University	Research Assistant
2008-	Alanya Faculty of Business, Akdeniz University	Instructor

PUBLICATIONS

Journal Articles

1. Oztimur Karadag O., Yarman Vural F.T. "Image Segmentation by Fusion of Low Level and Domain Specific Information via Markov Random Fields" being revised for Pattern Recognition Letters.

2. Oztimur Karadag, O., Yarman Vural F.T. "Domain Specific Segmentation For Building Detection", being revised for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

International Conference Publications

1. Oztimur Karadag O., Senaras Caglar, Yarman Vural F.T. "Segmentation and Decision Fusion For Building Detection", submitted to the IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2014.

2. Oztimur Karadag O., Yarman Vural F.T.,"Fusion of Image Segmentations Under Markov Random Fields", accepted to International Conference on Pattern Recognition, ICPR 2014.

3. Oztimur Karadag O., Yarman Vural F.T., "MRF Based Image Segmentation Augmented with Domain Specific Information", Image Analysis and Processing - ICIAP 2013, Lecture Notes in Computer Science Volume 8157, 2013, pp 61-70, Springer Berlin Heidelberg.

4. Oztimur Karadag O., Yarman Vural F.T., "Integration of Domain Specific Information in the Form of Color Homogeneity into MRF Based Image Segmentation", The 2013 World Congress in Computer Science Computer Engineering and Applied Computing, International Conference, The 2013 International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV'13), Las Vegas, Nevada, USA, 22-25 July 2013, http://world-comp.org/p2013/IPC.html.

5. O.O. Karadag, F.T.Y. Vural, "HANOLISTIC: A Hierarchical Automatic Image Annotation System Using Holistic Approach," cvprw, pp.16-21, 2009 IEEE

Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009.

6. Oztimur, O., Vural, F., "HANOLISTIC: A Hierarchical automatic Image Annotation System Using Holistic Approach," Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on , vol., no., pp.1-6, 27-29 Oct. 2008.

National Conference Publications

1. Karadag, O.O.; Vural, F.T.Y., "Semantic image segmentation with Markov Random Fields," Signal Processing and Communications Applications Conference (SIU), 2012 20th , vol., no., pp.1,4, 18-20 April 2012.