PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO


**Jhonatan Contreras Duarte**

# A Comparison Between Classical Object Based Methods and Conditional Random Fields.

**Dissertação de Mestrado**

Dissertation presented to the Programa de Pós-Graduação em Engenharia Elétrica of the Departamento de Engenharia Elétrica, PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica.

Advisor: Prof. Raul Queiroz Feitosa

Rio de Janeiro

April 2016


**Jhonatan Contreras Duarte**

# A Comparison Between Classical Object Based Methods and Conditional Random Fields

Dissertation presented to the Programa de Pós-Graduação em Engenharia Elétrica of the Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requeriments for the degree of Mestre.

**Prof. Raul Queiroz Feitosa**
Advisor
Departamento de Engenharia Elétrica – PUC-Rio

**Profa. Marley Maria Bernardes Rebuzzi Vellasco**
Departamento de Engenharia Elétrica – PUC-Rio

**Dr. Dário Augusto Borges Oliveira**
G**E Centro Brasileiro de Pesquisa**

**Dr. Peter Hofmann**
Universität Salzburg

**Prof. Márcio da Silveira Carvalho**
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, April 27th, de 2016

**Jhonatan Contreras Duarte**

Graduated in Electronic Engineering from the Universidad Industrial de Santander –UIS, Bucaramanga, Colombia, in 2013, and is currently a graduate student in the Electrical Engineering program at PUC-Rio. He has experience in the area of image processing, whit a focus in the detection of patterns, segmentation and classification. Additionally, he has knowledge in building models with applications mainly in remote sensing.

Bibliographic Data

Contreras Duarte, Jhonatan.

A Comparison Between Classical Object Based Methods and Conditional Random Fields/ Jhonatan Contreras Duarte; advisor: Raul Queiroz Feitosa. – 2016.

87 f. : il. (color.) ; 30 cm

Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2016.

Incluí bibliografia.

1. Engenharia elétrica – Teses. 2. OBA. 3. Campos aleatórios condicionais. 4. Segmentação. I. Feitosa, Raul Queiroz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

For the orchestrators, God and Maria Auxiliadora.
For my parents, Rosendo and Marina.
For my sisters, Yesika and Katherine.
For my nieces, Sarita and Camila.

## Acknowledgments



I am truly thankful to my advisor, Prof. Raul Queiroz Feitosa, for the encouragement, inspiring suggestions, advices and generous support throughout the development of my MSc research.

I thank my parents, Rosendo and Marina, my sisters, Yesika and Katherine, for their support.

I would like to express my gratitude to all the colleagues from LVC, for the companionship and valuable scientific discussion.

I also gratefully acknowledge the financial support of CAPES.

## Abstract

Contreras Duarte, Jhonatan; Feitosa, Raul Queiroz (Advisor). **A Comparison Between Classical Object Based Methods and Conditional Random Fields.** Rio de Janeiro, 2016. 87p. Master Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation investigates semantic segmentation techniques for the analysis of Earth observation data. This study has two main task. The first one is to assess the potential of semantic segmentation techniques as an option to traditional image segmentation methods that typically ignore the semantic information. The second objective is to compare the semantic segmentation with the typical object-based approach (OBIA). The study is based on an implementation of semantic segmentation based on Conditional Random Fields. The object-based approach is represented in this study by the segmentation algorithm known as Multiresolution. The Random Forests classifier is used to generate the association potentials for the conditional random fields and to perform the classification task in a representative implementation of the typical object-based approach. Experiments carried out on two high spatial resolution images (8 cm) indicated a clear superiority of semantic segmentation, both in terms of spatial accuracy and thematic accuracy. Although a more extensive analysis is required for the generalization of the aforementioned conclusions, the results of this study provide enough evidence to encourage a future research on the use of semantic segmentation to compose sophisticated image classification models, in particular being part of models inspired in the OBIA approach.



## Keywords

OBIA; Conditional Random Fields; Segmentation.

## Resumo

Contreras Duarte, Jhonatan; Feitosa, Raul Queiroz. **Uma comparação entre Métodos Clássicos Baseados em Objeto e Campos Aleatórios Convencionais.** Rio de Janeiro, 2016. 87p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta dissertação visa investigar técnicas de segmentação semântica para a análise de dados de observação da Terra. Dois são os objetivos perseguidos neste estudo. O primeiro é avaliar o potencial de técnicas de segmentação semântica como opção aos métodos tradicionais de segmentação de imagens que tipicamente ignoram a informação semântica. O segundo objetivo consiste em comparar a segmentação semântica com a abordagem típica baseada em objeto (OBIA). O estudo apoia-se numa implementação de segmentação semântica baseada em Campos Aleatórios Condicionais. A estratégia baseada em objeto é representada neste estudo pelo algoritmo de segmentação conhecido como Multiresolução. O classificador Florestas Aleatórias (Random Forests) é utilizado para gerar os chamados potenciais de associação dos campos aleatórios condicionais, bem como para realizar a tarefa de classificação na cadeia de processamento típico da abordagem baseada em objeto. Experimentos realizados sobre duas imagens de altíssima resolução espacial (8 cm) indicaram uma clara superioridade da segmentação semântica, tanto em termos de acurácia espacial quanto de acurácia temática. Embora carentes de uma análise mais aprofundada que permita a generalização de suas conclusões, os resultados obtidos no presente estudo provêm elementos suficientes para encorajar a pesquisa futura sobre a aplicação da segmentação semântica na composição de estratégias sofisticadas de classificação de imagens, em particular sendo parte de modelos baseadas em objeto.

## Palavras-chave

OBIA; Campos aleatórios condicionais; Segmentação.

## Contents


1 INTRODUCTION 16
1.1. Motivation 17
1.2. Objectives 19
1.3. Organization of the following chapters 20

2 Background 21
2.1. Segmentation Approaches 21
2.1.1. Bottom Up Methods 21
2.1.2. Superpixels 24
2.1.3. Interactive methods 25
2.1.4. Object Proposals 25
2.1.5. Semantic Segmentation 25
2.2. Object-Based Image Analysis - OBIA 28
2.3. Multiresolution Segmentation 29
2.4. Segmentation Parameter Tuning 31
2.5. Simple Linear Iterative Clustering (SLIC) 32
2.6. Conditional Random Fields – CRF 33
2.6.1. Association Potential 35
2.6.2. The Interaction Potential 35
2.6.3. Inference 36

3 Methodology 37
3.1. Thematic and Spatial accuracy metrics 37
3.1.1. Spatial accuracy - **F**-measure 37
3.1.2. Thematic Accuracy 39
3.2. Supervised Segmentation Parameter Tuning Methodology 40
3.3. Semantic Segmentation Methodology 40
3.3.1. SSeg Processing Steps 41
3.3.2. Tuning the Interaction Potential 42
3.4. OBIA Methodology 42

4 Experimental Analysis 44
4.1. Dataset description. 44
4.2. Features 46
*4.3.* Training and test procedure for *segmentation parameter tuning* 46
*4.4.* Selecting training and test data for *SSeg* 47
4.5. Spatial accuracy of SSeg and MRS. 49
4.5.1. Sensitivity of CRF to superpixel size 55
4.5.2. Sensitivity of CRF to parameter $\boldsymbol{\beta}$ 57
4.6. Thematic accuracy of SSeg and OBIA 58
4.6.1. Thematic accuracy of semantic segmentation 58
4.6.2. Thematic accuracy of basic OBIA results. 67
4.6.3. Comparing thematic accuracies 69

5 Conclusions 72

6 Reference 74

## List of Figures

Figure 2-1 Optimization methodology taken from (Achancccaray, et al., 2015) 32

Figure 2-2 Pixel connectivity, four and eight connected. 34

Figure 3-1 Left segmentation outcome. Right, spatial accuracy result, reference segment in green, $tp =$ yellow, $fn =$ blue and $fp =$ red. 38

Figure 3-2 Confusion Matrix 39

Figure 3-3 Image sites generated by SLIC for few (large) and many (smaller) superpixels. 41

Figure 4-1: (left) Image 1, Vaihingen Area 13; (right) Ground Truth: 'Building' (blue), 'Low vegetation' (Cian), 'Tree' (Green), 'Car' (yellow) and 'Street' (white). 45

Figure 4-2: (left) Image 2, Vaihingen Area 17; ; (right) Ground Truth: 'Building' (blue), 'Low vegetation' (Cian), 'Tree' (Green), 'Car' (yellow) and 'Street' (white). 45

Figure 4-3 Reference Segment of Image 1 for SPT 47

Figure 4-4 Reference Segment of Image 2 for SPT 47

Figure 4-5 labeled training data for Image 1 (left) and Image 2 (right): 'Building' (blue), 'Low vegetation' (Cian), 'Tree' (Green), 'Car' (yellow) and 'Street' (white). 48

Figure 4-6: Segmentation outcome for Image 1 (left); positive and negatives (yellow=TP, red=FP, blue=FN) (right). 49

Figure 4-7 Zoom over the region with red circles first image. 50

Figure 4-8 CRF using 140.000 superpixels with different values of β, (a) Small value, $\beta = 0.1$, (b) Large value, $\beta = 1.4$, (c) medium value, $\beta = 0.7$. (d) Supervised segmentation parameter tuning for Image 1. 51

Figure 4-9 Positive and negatives for Image 1(yellow=TP, red=FP, blue=FN) produced by SSeg (left) and SSPT (right). 52

Figure 4-10 Segmentation outcome for Image 2 positive and negatives (yellow=TP, red=FP, blue=FN) (down). 53

Figure 4-11 Results for CRF using 140.000 superpixels for different values of β: (a) small $\beta = 0.1$; (b) Large $\beta = 2$; (c) medium $\beta = 1.45$; (d) results for supervised segmentation parameter tuning for Image 2. 54

Figure 4-12: Positive and negatives for Image 2 (yellow=TP, red=FP, blue=FN) (left) for SSeg; (right) for supervised segmentation PT. 54

Figure 4-13 CRF spatial accuracy vs. number of superpixels 55

Figure 4-14 Image 1 results for reference segments (yellow=TP, red=FP, blue=FN). Upper left, CRF using 4000 SP. Upper right, CRF using 140000 SP. Bottom left, CRF using 50000 SP. Upper right, CRF using 70000 SP. 56

Figure 4-15 Image 2 results for reference segments (yellow=TP, red=FP, blue=FN). Upper left, CRF using 4000 SP. Upper right, CRF using 140000 SP. Bottom left, CRF using 50000 SP. Upper right, CRF using 70000 SP. 56

Figure 4-16 Image 1, spatial accuracy F- measure vs. $\beta$ 57

Figure 4-17 Image 2, spatial accuracy F-measure vs. β 58

Figure 4-18 optimum $\beta$ for classification vs. number of superpixels 59

Figure 4-19 Classification results Image 1 of the SSeg model for β below (a), above (b) and equal (c) to the optimum as well as the ground truth (d). In all cases the number of superpixels was set to 140,000. 60

Figure 4-20 Classification results Image 2 of the SSeg model for β below (a), above (b) and equal (c) to the optimum as well as the ground truth (d). In all cases the number of superpixels was set to 140,000. 60

Figure 4-21. Average Accuracy and Overall Accuracy for different values of number of superpixels for the Image 1 and Image 2. 61

Figure 4-22 samples of cars in Image 2. 64

Figure 4-23 Overall Accuracy vs. $\beta$ for Image 1 66

Figure 4-24 Overall Accuracy vs. β for Image 2 66

Figure 4-25 Segmentation of Image 1 for scale parameter equal to 5 (a) and to 50 (b) 67

Figure 4-26 Segmentation of Image 2 for scale parameter equal to 5 (a) and to 50 (b) 67

Figure 4-27 Average Accuracy and Overall Accuracy for different values of Scale parameter, OBIA. 68

Figure 4-28 Left, classification result of the Image 2 using Scale 5 (left), classification results of the Image 2 using Scale 20 (right) 69

Figure 4-29 Best classification results for Image 1. (a) SSeg using superpixels. OBIA using over-segmented input image using (b) MRS. (c) Ground truth. 69

Figure 4-30 Best classification results for Image 2. (a) SSeg using superpixels. OBIA using over-segmented input image using (b) MRS. (c) Ground truth 70

## List of Tables


Table 4-1 percentage of pixels of Image 1 used for training and test 48
Table 4-2 Percentage of pixels of Image 2 used for training and test 48
Table 4-3 Parameters tuned for Image 1 49
Table 4-4 Parameters tuned for Image 2 52
Table 4-5 Confusion matrix for Image 1 with 140,000 SP 62
Table 4-6 Confusion matrix for Image 1 with 4,000 SP 62
Table 4-7 Confusion matrix for Image 1 with 30,000 SP 63
Table 4-8 Confusion matrix for Image 1 with 100,000 SP 63
Table 4-9 Confusion matrix for Image 1 with 140,000 SP 63
Table 4-10 Confusion matrix for Image 2 with 140,000 SP 64
Table 4-11 Confusion matrix for Image 2 with 7,000 SP 65
Table 4-12 Confusion matrix for Image 2 with 40,000 SP 65
Table 4-13 Confusion matrix for Image 2 with 140,000 SP 65
Table 4-14 Highest values for OA and AA for Image 1 70
Table 4-15 Highest values for OA and AA for Image 2 70

## List of Symbols and Abbreviations

| | |
|---|---|
| $MS$ | Multispectral. |
| $Pan$ | Panchromatic. |
| $OBIA$ | Object Based Image Analysis. |
| $GEOBIA$ | Geographic Object Based Image Analysis. |
| $VHR$ | Very High Resolution. |
| SSeg | Semantic Segmentation. |
| SSPT | Supervised Segmentation Parameter Tuning. |
| $CRF$ | Conditional Random Fields. |
| $\beta$ | Beta parameter. |
| $CNN$ | Convolutional Neural Networks. |
| $FCN$ | Fully Convolutional Networks. |
| $DN$ | Deconvolutional Networks. |
| $MRF$ | Markov Random Fields. |
| $RS$ | Remote Sensing. |
| $GIS$ | Geographic Information Systems |
| $SLIC$ | Simple Linear Iterative Clustering. |
| $MRS$ | Multiresolution Segmentation. |
| $LMBF$ | Local Mutual Best Fitting |
| $C_n$ | Segment $n$ |
| $f$ | Merging cost or degree of fitting. |
| $h_{color}, h_{shape}$ | Spectral and shape components. |
| $\omega_{color}, \omega_L$ | Spectral and band weights. |
| $A$ | Area of a region or segment in pixels. |
| $L$ | Spectral band. |
| $C_1 \cup C_2$ | Resulting region after merging $C_1$ and $C_2$. |
| $Sol$ | Solidity. |
| $Comp$ | Compactness. |
| $SPT$ | Segmentation Parameter Tuning. |
| $GM$ | Graphical Model. |
| $P(x, y)$ | Probability distribution. |

| | |
|---|---|
| $P(y\|x)$ | Conditional distribution. |
| $x_i$ | Random variable $i$. |
| $G(V,E)$ | Graph. |
| $V,E$ | Nodes and edge of a Graph $G$ |
| $S$ | Subset of the graph $G$. |
| $\Psi_S$ | Subset of factors. |
| $Z$ | Partition function. |
| $A_i(x,y_i)$ | Association Potential. |
| $I_{ij}\ (y_i,y_j,x)$ | Interaction Potential. |
| $LBP$ | Loopy Belief Propagation. |
| $RF$ | Random Forest Classifier. |
| $SP$ | Superpixels. |
| $SPT$ | segmentation parameter tuning |
| $MG$ | Mixture of Gaussians |
| $MS$ | Mean Shift |
| $VOC$ | Visual Object Classes |

# 1
# INTRODUCTION

The task of segmentation in computer vision consists of dividing an image into regions or objects, which are composed of subgroups of pixels (Gonzales & Woods, 2008). The segmentation task is one of the most important processing steps in the field of image analysis. Its quality is a determining factor for the success of the following steps in computer vision such as, recognition and object extraction.

Remote sensing is defined as the acquisition of information about objects or phenomena on the Earth's surface without physical contact through the placement of devices on aircrafts or satellites (Rocha, 2007). In remote sensing, image classification aims to categorize all pixels of a digital image into one predefined land cover classes (Lillesand, et al., 2004). An example of land cover could be a forest, a building, farmland or a road among other types of classes. Image classification can be divided into two methods: unsupervised and supervised classification. The unsupervised classification consists of the generation of clusters and the subsequent manual assignment of a type of class to each of them. On the other hand, supervised classification involves the uses of a training set, which contains samples of the classes of interest, in order to develop a statistical characterization of the data and later predict a class for each pixel in an image.

The traditional techniques of pixel based image analysis for high and very high resolution remote sensing were limited, inadequate and insufficient to handle the high interclass variation of complex scenes (Schiewe, 2002; Carleer, et al., 2005; Blaschke, 2010). This insight was the main reason for the emergence of two news areas of scientific research in image analysis called Object Based Image Analysis (OBIA) and Geographic Object Based Image Analysis (GEOBIA) for very high-resolution images (Hay & Castilla., 2006).

OBIA and GEOBIA have been regarded by many researchers as a trend, or even, a paradigm shift (Blaschke, et al., 2014) in the analysis of remotely sensed images. Although OBIA concepts have been stablished in the eighties and nineties

(McKeown, et al., 1985; Matsuyama, 1987; Matsuyama & Hwang, 1990; Clément, et al., 1993; Liedtke, et al., 1997), it was only after the first commercial OBIA oriented software came into the marked that this methodology started being more extensively exploited by the community of environmental sciences.

The superiority of object-based over traditional pixel-based approaches for the analysis of very high resolution images (VHR) has been attested by many publications (Platt & Rapoza, 2008; Im, et al., 2009; Moran, 2010; Myint, et al., 2011; Vieira, et al., 2012; Pinho, et al., 2012).

The basic OBIA processing chain comprises two sequential steps: the segmentation that partitions the image into homogeneous regions, followed by the classification that assigns a class label to each segment produced in the segmentation step. Segmentation is the critical step in this scheme, since often its outcome is not fully consistent with the object borders (Lübker & Schaab, 2009; Smith, et al., 2010). This is due to the fact that segmentation relies solely on low-level image features, such as color or texture, and fully ignores semantic, which is highly subjective, and depends both on the application and on the user. In the basic OBIA processing chain, segmentation errors propagate to the classification step, which does not fix them.

This work is focused in this context and compares two strategies to partially alleviate the segmentation problem mentioned above into the basic OBIA. Additionally, this work assesses an alternative to the basic OBIA called Semantic Segmentation.



## 1.1. Motivation

In the last several decades, many segmentation algorithms have been proposed (Vantaram & Saber, 2012; Dey, et al., 2010; Neubert, et al., 2008; Haralick & Shapiro, 1985), which underscore the importance of this process in computer vision and remote sensing. Additionally, it is a confirmation of the growing interest in this topic, which is far from being fully developed. Three strategies to partially improve the segmentation outcomes which are not fully consistent with the object borders are described in the following paragraphs.

The first strategy, is what we call hereafter supervised segmentation parameter tuning (SSPT) (Feitosa, et al., 2006; Pignalberi, et al., 2003; Fourier & Shoepfer, 2014). In this strategy segmentation is guided by the semantics embedded in manually delineated segment samples. For instance, an optimization algorithm searches the parameter space for the set of values that leads to the optimum match between samples provided by the user and the segmentation outcome, as expressed by an empirical discrepancy metric.

A second strategy to overcome the aforementioned limitation of OBIA basic procedure consist of an iterative segmentation-classification loops (Tiede, et al., 2010). This strategy first over-segments the image into small segments, which are preliminary classified. The typically small homogeneous segments produced this way rarely extend over more than one object. From then on segments are aggregated through multiple iterative segmentation and classification steps. At each cycle, errors in the segmentation step might be fixed by the subsequent classification step.

A third approach, called semantic segmentation (SSeg), aims to partition an image into semantically homogeneous regions. Instead of performing segmentation and classification as independent steps, possibly in an iterative way, semantic segmentation does both simultaneously.

This dissertation investigates SSeg as an alternative to supervised segmentation parameter tuning into the basic OBIA approach. SSPT supposed to be the best segmentation result obtained for a particular algorithm. Although different SSeg techniques have been proposed so far, in this study we focus on a technique based on Conditional Random Fields (CRF), which represents the approaches most widely investigated in the recent years (Lafferty, et al., 2001; Ladicky, et al., 2009; Yang, et al., 2010; Csurka & Perronnin, 2011; Zhu, et al., 2016).

More than merely classifying pixels as isolated entities, CRF allows modeling the interaction among neighboring pixels in a class-by-class basis. Driven by computational constraints, this approach is often preceded by image partitioning into spectrally homogenous image sites called “superpixels”. The major difference between superpixels and “small segments”, is that superpixels have a nearly regular geometry and their sizes do not vary much over the image, for reasons that will be later clarified in this manuscript.

The study is organized in two parts. Firstly, SSeg is evaluated as an alternative to conventional segmentation from the perspective of the accuracy of object delineation. Specifically, the spatial accuracies of SSeg and SSPT approaches are experimentally compared. Secondly, the study evaluates SSeg as alternative to the basic OBIA approach, i.e., segmentation followed by classification, in terms of final thematic accuracy.

## 1.2. Objectives

The general objective of this dissertation is twofold. First, to compare the semantic segmentation (SSeg) and the supervised segmentation parameter tuning (SSPT) in terms of spatial accuracy. Second, to compare SSeg and the basic OBIA strategy in terms of thematic accuracy.

Each approach investigated in the present study admits a number of variants. An exhaustive analysis of those alternatives would not be possible within the scope of this dissertation. Thus, a particular configuration was chosen for each analyzed approach. The choices were mostly determined by what has been more widely used in the community in recent times. In some cases, a simple solution was taken as representative for a given approach.

In this path, an extension of the *Simple Linear Iterative Clustering* (SLIC) (Achanta, et al., 2012) algorithm was selected for superpixel generation, By far the most popular superpixel method is the SLIC algorithm (Achanta, et al., 2012). For segmentation the *Multiresolution Segmentation* (MRS) algorithm (Baatz & Schäpe, 2000) was used, MRS is based on *region growing* methods, which have been widely employed especially in the area of remote sensing (Tilton & Lawrence, 2000). Segmentation parameters were tuned using the *Segmentation Parameter Tuner* (SPT) tool (Achanccaray, et al., 2015). Random Forest (RF) (Breiman, 2001) was elected as the basic classifier both for implementing association potentials in CRF as well as for composing the solution that represents the OBIA approach in the present dissertation, since for both tasks any local classifier with a probabilistic output can be used.

## 1.3. Organization of the following chapters

The next chapter presents a brief survey of the different image segmentation approaches.

Chapter 2 describes succinctly some techniques this study is based upon: Object Based Image Analysis, Segmentation Parameter Tuning as well as the Multiresolution Segmentation and Conditional Random fields.

Chapter 3 describes the methodology proposed in this thesis and the set of metrics used to assess the thematic and spatial accuracy.

Chapter 4 presents the dataset used in the experiments as well as the results obtained in this study.

Chapter 5 presents some concluding remarks and points to future extensions of this study.

# 2
# Background

This chapter discusses some important theoretical foundations for the understanding of this work. . Section 2.1 reviews image segmentation techniques. Section 2.2 reviews the Object-Based Image Analysis (OBIA). Section 2.3 describes the Multiresolution Segmentation (MRS) algorithm. Section 2.4 presents a supervised segmentation parameter tuning method as well as a tool that implements it, which was used in our experimental analysis. Section 2.5 defines a superpixel method used for image site generation for the CRF called Simple Linear Iterative Clustering (SLIC). Finally, Section 2.6 addresses the basic concepts of the conditional random field (CRF).



## 2.1.
## Segmentation Approaches

The task of segmentation in computer vision consists of the division of an image into subgroups of pixels called segments. The grouping procedure is guided by some properties that the pixels belonging to the same segment are expected to share (Gonzalez & Woods, 2008).

Many techniques of image segmentation have been proposed for about four decades. Some of the most widely used algorithms can be divided into five methods according to (Zhu, et al., 2016) and (Thoma, 2016). These methods are: Bottom up Methods, Superpixels, Interactive methods, Object Proposals and Semantic Segmentation or Image Parsing.

### 2.1.1.
### Bottom Up Methods

Bottom-up segmentation methods rely entirely on image data and do not consider semantic. This class of segmentation methods aims at grouping nearby pixels, which share some local characteristics in the feature space, e.g., color, texture or curvature.

Zhu divides the bottom-up methods into two sub-categories: discrete bottom-up and continuous bottom-up (Zhu, et al., 2016). Discrete bottom-up methods regard an image as a fixed discrete grid, whereas continuous methods consider an image as a continuous surface (Mumford, et al., 1989; Kass, et al., 1988).

Discrete bottom-up approaches are by far the most widely used segmentation methods in remote sensing image analysis. In the following we briefly describe the most important subgroups of bottom-up segmentation algorithms.

*K-means*

K- Means (Agarwal, et al., 2002) is perhaps the simplest among all methods listed in this short survey. Given $k$ initial centers in the feature space, each pixel represented by its feature vector is assigned to one of the centers according to their distance to $k$ centers, where by $k$ denotes the number of cluster expected to exist in the feature space. Subsequently the centers are updated. These two steps iterate until a stopping criterion is satisfied. The segments are then formed by agglomerates of pixels belonging to the same cluster.

*Mixture of Gaussians*

Another clustering method called mixture of Gaussians –MG (Gupta & Sortrakul, 1998) bases on Gaussian Mixture Models. Each image pixel is assumed to belong to a class that can be described by a single multivariate Gaussian distribution. Thus, feature vectors representing all image pixels are modeled as a mixture of Gaussians, whose parameters are determined by the expectation maximization (EM) algorithm (Shi & Malik, 1997). A byproduct of EM is the assignment of each pixel to one of the Gaussians, what ultimately determines the cluster each pixel belongs to. Each spatial cluster of pixels belonging to the same Gaussian forms a segment.

*Mean-Shift*

The Mean Shift- MS (Comaniciu & Meer, 2002) segmentation is a non-parametric clustering method, which applies what is known as *kernel density estimation*. The basic Mean Shift algorithm finds the modes, i.e., local maxima,

of multivariate functions, MS automatically decide the cluster number and modes in the feature space. In order for MS to segment images, the feature vector of each pixel is extended by the incorporation of its spatial coordinates. This makes the clusters to consist of similar pixels both in terms of their features as well as in terms of their location in the image.

The MS procedure (usually) starts at each pixel, the centroid of its neighbors around a fixed window in the extended feature space is computed and the procedure moves from the initial pixel to the centroid position (mean shift). This procedure is repeated through many iterations making the centroid to move toward a mode. The procedure stops when the centroid shift between two consecutive iterations is small indicating that a mode has been reached.

*Watershed*

Whatershed (Beucher & Meyer, 1993) is a segmentation method that considers images as topographic surfaces composed by valleys and mountains, where the gradient magnitude on the pixels intensity corresponds to the altitude of the topographic surface (Beucher & Meyer, 1993). The watershed process simulates the flooding of the surface from the local minima, forming pools. A containment line is created when the water of two neighboring basins are about to make contact to turn into a single basin. The containment lines obtained this way define the final segments borders (Pedrini & Schwartz, 2008).

*Graph Based*

Graph Based methods (Felzenszwalb, et al., 2004) map an image into a graph with four or eight connectivity nodes. The graph $G = (V, E)$, is composed by nodes $(V)$ and edges $(E)$. Nodes correspond to pixels, while edges reflect the adjacency among them. Furthermore, each edge is associated to a weight that represents the color dissimilarity between nodes connected by that edge. A segmentation of a graph is the division of all nodes into segments. Nodes in the same segment should be similar to each other and adjacent nodes of different segments should be different. So, the sum of the weights related to edges connecting pixels within a segment should be low, whereas the sum of weights of the edges connecting nodes in different segments should be high.

The segmentation process consists in separating groups of neighboring pixels (nodes) by eliminating the edges that connect pixels inside each group to pixels outside it in a way that a weighted sum of the remaining edges is below a given limit.

*Region Growing*

The segmentation algorithm most widely used by the OBIA community falls in the category called Region Growing. Such methods start from pixels or super-pixels (see later) and merge adjacent regions based on some homogeneity criterion that may take spectral, morphologic and topological features into account. The algorithms of this group vary mostly on the adopted homogeneity metric. A better insight on how region growing segmentation works is illustrated in the next chapter, where the Multiresolution Segmentation algorithm is described with some details.



### 2.1.2. Superpixels

The objective of superpixel methods is to over-segment an image into homogeneous regions, which are smaller than an object and have a nearly regular geometry. According to (Ren, et al., 2003), superpixels are more natural and efficient representations than pixels, because local features extracted from a pixel can be ambiguous and more sensitive to noise.

Pixel-based classification methods imply the use of a large volume of data, which hinder the training and the inference procedures. In this research, superpixels serve as a basis for a more sophisticated algorithm called conditional random fields (CRF) (Lafferty, et al., 2001). The use of superpixels reduces the model's complexity and the associated computational cost, improving the algorithm's efficiency.

Most approaches grow superpixels from an initial set of regions determined by a regular grid. Then the region boundaries are adjusted iteratively to adhere to salient object contours. By far the most popular superpixel method is the SLIC algorithm (Achanta, et al., 2012), which is explained in the next chapter.

### 2.1.3. Interactive methods

Interactive methods allow for the user to assist the segmentation. They aim at capturing human perception, prior knowledge or constraints provided by the user as input to or during the segmentation procedure. Such methods are useful in applications where accuracy is of key importance. Examples are medical image analysis and image editing. Interactive methods are not commonly used in remote sensing image analysis. Surveys of interactive segmentation methods can be found in (McKeown, et al., 1985; Yi & Moon, 2012; He, et al., 2013).

### 2.1.4. Object Proposals

Object proposals segmentation methods aim to divide an image between "objects" and "stuffs", where an "object" has a particular size and shape (e.g. car, house) and the homogenous background or non-delineated objects are considered as "stuff" (e.g. sky, river). Object proposals can be divided in two groups, class-specific and class-independent object proposals, according to (Zhu, et al., 2016).

Class-specific object proposals are tailored for a limited and well defined object class (Larlus & Jurie, 2008; Shi & Malik, 1997). In contrast, class-independent object proposals aim at finding general, non-specific objects that emerge from background (Borji, et al., 2014). The underlying idea is that objects-of-interest differ from background in certain appearance or geometry cues, no matter what they are.

To the knowledge of the author, object proposal segmentation has not been used in remote sensing image analysis.

### 2.1.5. Semantic Segmentation

Semantic segmentation or image parsing aims to divide an image into non-overlapped segments which correspond to predefined semantic classes (e.g. car, grass, building, etc.). Semantic image parsing requires a given information to use high-level learned representation and complete its task. The learned models can

be used to predict regions that belong to the training semantic classes in new images.

A large number of semantic segmentation methods are formulated as the problem to find the most likely labeling on a Markov Random Field (MRF) (Li, et al., 2004; Zhang, et al., 2010) or a Conditional Random Field (CRF) (Lafferty, et al., 2001; Ladicky, et al., 2009; Yang, et al., 2010; Csurka & Perronnin, 2011; Zhu, et al., 2016).

This study exploits a CRF based approach as representative of semantic segmentation. The theoretical formulation of this approach will be presented in chapter 3.

Besides methods based on CRF or MRF, some new semantic segmentation approaches have been proposed recently. In order to assess the relative performance of some approaches mentioned in the following paragraphs, a benchmark of visual object category recognition and detection called the PASCAL Visual Object Classes (VOC) was used.

Mostajabi introduced in 2015 an algorithm of semantic segmentation based on feed-forward architecture. This algorithm extracts local features as color, texture, and location from different levels of spatial context around a superpixel. That means that, it extracts information from a superpixel, from a small region around it, from a larger region around it and from the entire image. Later, this algorithm combines all the local features previously found. Thus, the algorithm classifies the superpixels in the image by a feedforward multilayer network. This algorithm presents accuracy of 64.2% in PASCAL VOC 2012 dataset (Mostajabi, et al., 2015).

Convolutional Neural Networks (CNN) are a variation of multi-layer neural networks, trained with a version of the back-propagation algorithm. CNN are composed of a set of layers, each of which with a different purpose. In the firsts layers occur the features of the extraction of the images, which, consists of convolutional neurons and down sampling. On the other hand, at the final layers of the network, simple perceptron neurons are responsible for the final classification of the extracted features. Many algorithms based on CNN are presenting satisfactory results in visual recognition problems, such as face recognition (Lawrence, et al., 1997) and semantic segmentation problems (Petersen, et al., 2002; Gondra & Xu, 2010; Csurka & Perronnin, 2011;

Krizhevsky, et al., 2012; Pinheiro & Collobert, 2014; Simonyan & Zisserman, 2014).

Chen (Chen, et al., 2014) presents a semantic segmentation algorithm based on CNN. His algorithm combines deep convolutional networks and Fully Connected Conditional Random Fields. It shows that it is able to produce semantically accurate predictions and detailed segmentation maps. This algorithm presents accuracy of 71.6% in PASCAL VOC 2012 dataset.

An important variant of CNN is called Deconvolutional Networks (DN). DN is a framework that allows the unsupervised construction of hierarchical image representations providing features for object recognition and semantic segmentation approaches (Zeiler, et al., 2010).

Separately, Noh (Noh, et al., 2015) proposed a pixel-wise semantic segmentation algorithm composed by a linked Convolutional and Deconvolutional Networks. The convolutional network learns in the same manner as CNN (convolutions and down sampling). Moreover, the Deconvolutional networks are composed of deconvolutions and up sampling layers. This algorithm presents accuracy of 72.5% in PASCAL VOC 2012

Recently, Long (Long, et al., 2015) proposed a variation of CNN called Fully Convolutional Networks (FCN). This algorithm solves a problem that limits CNN. The CNN training process is pixels-to-pixels. Thus, the size of the input and output image is predefined. In contrast, a FCN takes any size of input images and produces outputs with the respective size of the input. This algorithm presents accuracy of 62% in PASCAL VOC 2012 dataset.

Finally, Liu (Liu, et al., 2015) combines different methods mentioned above as CNN and CRF, in which a pre-trained deep CNN generates features to train a CRF. The CRF is trained by the deep convolutional features, extracted from superpixels, using a structured support vector machine (SSVM). Additionally, this approach includes spatial information related to objects that appear side by side in the scenes. It influences the labelling of objects frequently co-occurred in the training data set. Thus, the objects with certain spatial relationship are labeled during the inference (e.g. cars and street) unlike to the objects without relationships (computers and trees).

## 2.2. Object-Based Image Analysis - OBIA

The classical remote sensing image analysis relies on pixel-wise classification, whereby spectral features and derivatives, such a texture descriptors are used as pixel descriptors. In low spatial resolution imagery, the objects of interest are often similar or smaller than the size of the pixels and the spectral information might be enough to discriminate the targets (Hay & Castilla, 2006).

The classical pixel-wise approach includes parameters of the image in addition to the spectral information such as tone, texture, shape, context, etc. At higher spatial resolutions, objects of interest are composed of many pixels. Thus, a paradigm shifts from the pixel-wise to object-based methods, whereby the last ones consider the characteristics of an object through spatial, spectral and temporal scales. These latest methods became to be known as the object-based image analysis (OBIA) (Blaschke, 2010) .

In OBIA, image-objects are expected to correspond to 'meaningful' entities that are internally consistent and different from their surroundings (e.g., a building, tree or vehicle) (Castilla, et al., 2007).

In its initial step, OBIA applies some bottom-up segmentation algorithm, top-down or even mixed algorithms, in order to obtain the segments or image-objects. In the next step, segments are classified based on features describing the segments' color, shape, texture and spatial context. Though many different approaches can be used in this second step, knowledge base approaches are visibly more often applied than in other remote sensing domains. OBIA has been successfully applied in many fields such as the biological, habitat mapping, urban mapping, medical, mineral exploration, transportation, and security, among others.

This methodology has been referred by many authors as Geographic Object-Based Image Analysis (GEOBIA) (Hay & Castilla, 2008) instead of OBIA, in order to emphasize the objective of generating geographical information. In this sense GEOBIA has been defined as a "sub-discipline of the Geographic Information Science (GIScience) devoted to developing automated methods to partition remote sensing imagery into meaningful image-objects, and to assessing

their characteristics through spatial spectral and temporal scales, so as to generate new geographic information in GIS-ready format" (Hay & Castilla, 2008).

How the GEOBIA uses RS data and generates Geographic Information Systems (GIS) as output represents a bridge between two ways to represent the location component of geographic information: the raster (grid-based) domain of RS, and the vector (point-based) domain of GIS. The linking of both domains is the generation of polygons (i.e., classified or segmented image-objects) representing geographic objects (Castilla, et al., 2007). Finally, the generation and use of geographic information (GI) and RS in Computer Vision distinguish GEOBIA from OBIA (object-based image analysis).

## 2.3. Multiresolution Segmentation

The first and crucial step of OBIA/GEOBIA processing chain is image segmentation. Many segmentation algorithms have been used for that task. Among them, the one proposed by Baatz and Schäpe (Baatz & Schäpe, 2000), so called Multiresolution Segmentation (*MRS*), is beyond a doubt the most widely used one within the OBIA community. In our experiments we used an MRS implementation developed by Happ and co-workers (Happ, et al., 2013), specifically the variant named Local Mutual Best Fitting (LMBF).

In *MRS,* firstly, each pixel is considered as a segment. In the later steps any two adjoining segments are considered for being merged into one larger segment. The merging decision is based on a *local homogeneity criterion* involving both segments. Basically, a *merging cost* represents the increase of heterogeneity resulting from merging two segments. The *merging cost* can be viewed as a degree of fitting between the segments being considered to fuse into a single one. Only if the *merging cost* is inferior to a user selected threshold, called *scale* parameter, the merge is admissible. The segmentation procedure ends when no additional merging can be executed.

In LMBF variant, a merge only occurs if the *best fitting* condition is mutual between both segments, i.e. if $C_1$ is the *best fit* of $C_2$ among all of the adjacent segments of $C_2$ and simultaneously $C_2$ is the *best fit* of $C_1$among all of the adjacent segments of $C_1$.

The merging cost ($f$) or degree of fitting is composed by a spectral ($h_{color}$) and a morphological component ($h_{shape}$). The merging cost is expressed by equation 2-1.

$$f = \omega_{color} \cdot h_{color} + (1 - \omega_{color}\ ) \cdot h_{shape} \quad 2\text{-}1$$

where $\omega_{color}$ takes values in the range [0 1] and represents the relative importance of color ($h_{color}$) and morphologic ($h_{shape}$) features.

The spectral component $h_{color}$ is defined by Equation 2-2, where $L$ is a spectral band and $\omega_L$ its respective weight, determined by the user as input parameter. $A$ is the area in pixels of a given region; $\sigma_L^{C_1}$ , $\sigma_L^{C_2}$ and $\sigma_L^{C_1 \cup C_2}$ are the standard deviations of pixels in regions $C_1, C_2$ and $C_1 \cup C_2$ respectively, where $C_1 \cup C_2$ represents the resulting region after the merge.

$$h_{color} = \sum_L \omega_L \left( A_{C_1 \cup C_2} \times \sigma_L^{C_1 \cup C_2} - \left( A_{C_1} \times \sigma_L^{C_1} + A_{C_2} \times \sigma_L^{C_2} \right) \right) \quad 2\text{-}2$$

The morphological component (see equation 2-3) has two morphological features, *Smoothness* and *Compactness*. The compactness weight ($\omega_{comp}$) is defined to control the relative importance of each morphological feature.

$$h_{shape} = \sum_L \omega_L \left( \omega_{comp} \cdot h_{comp} + \left( 1 - \omega_{comp} \right) \cdot h_{sol} \right) \quad 2\text{-}3$$

Smoothness computation requires measuring the border length of the resulting segment after the merging of two adjacent segments, an operation that might be computationally expensive when performed in GPUs. Happ and co-authors (Happ, et al., 2013) proposed to take two other morphological features, *Solidity* and *Compactness*, as alternative to *Compactness* and *Smoothness.*

*Solidity* is defined by Equation 2-4, where $A$ is the area of the segment; and $Abox$ is the area of its bounding box. This feature is sensitive to the convexity of the segment, taking its minimum value for rectangular segments.

$$Sol = \frac{Abox}{A} \quad 2\text{-}4$$

*Compactness* is defined in equation 2-5, where $dmax$ is the length of the major axis of the ellipse with identical second order moment. *Compactness* is minimum for circular shapes.

$$Comp = \frac{dmax}{\sqrt{\frac{4A}{\pi}}} \quad 2\text{-}5$$

The implementation of Happ's algorithm is available on the website of the Computer Vision Lab (LVC): http://www.lvc.ele.puc-rio.br/wp/?p=1092#more-1092. The parameter selection for this algorithm was done automatically using the Segmentation Parameter Tuning (SPT), which is presented in section 2.4. Three parameters need tuning in this algorithm, they are: the scale parameter, the color weight, and the compactness weight. In order to reduce the computational cost and taking into account that the information of each band is equally important, the weights assigned to each band ($\omega_L$) were set to 0.33 (Diaz, 2014).



## 2.4. Segmentation Parameter Tuning

All segmentation algorithms have parameters that must be tuned so as to obtain *good quality* segment delineations. Segmentation quality can be assessed by *empirical methods* that compare the segmentation outcome with a set of reference (Zhang, 2001). Dragut proposed a tool to tuned scale parameter for multiresolution image segmentation of remotely sensed data (Drăgut, et al., 2010).

Segmentation Parameter Tuner (SPT) is a tool that finds the local optimal segmentation parameters values according to a specified set of segment references (Achanccaray, et al., 2015). The optimum set of parameter values maximizes the agreement (similarity) between segmentation output and the reference. Different metrics can be used to express similarity (Zhang, 2001).

An optimization procedure searches the parameter space for the optimal set of parameter values. Figure 2-1 shows the methodology followed by SPT. First, the input image is segmented using an initial set of parameter values. Later, the selected fitness function is calculated by comparing the segmentation result with the references provided by the user. This process is repeated iteratively, taking

different segmentation parameters, until the minimum value is found or the convergence criterion is satisfied.

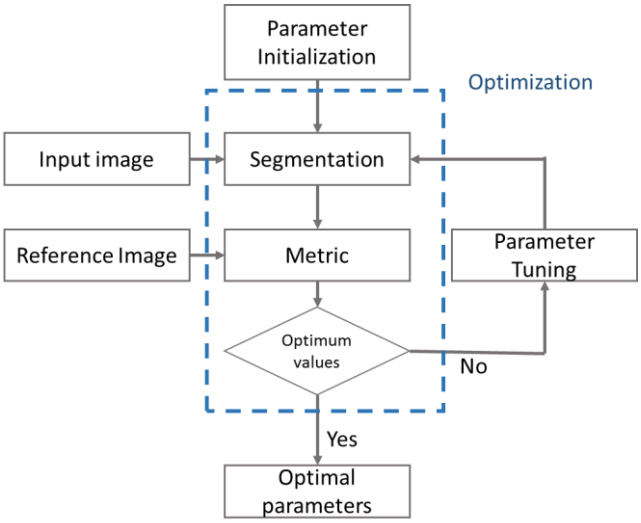



Figure 2-1 Optimization methodology taken from (Achancccaray, et al., 2015)

The SPT tool includes five segmentation algorithms, four optimization algorithms and seven discrepancy metrics. In this research, the MRS algorithm was used as the segmentation algorithm (Happ, et al., 2013), the Nelder-Mead optimization algorithm as the parameter tuning (Nelder & Mead, 1965), according to (Achanccaray Diaz, et al., 2014) Nelder-Mead demonstrated a good performance at this task, and F-measure was used as the similarity metric since is a combination of precision and recall.

## 2.5. Simple Linear Iterative Clustering (SLIC)

Simple Linear Iterative Clustering (SLIC) (Achanta, et al., 2012) performs K-means in the 5D space $[labxy]$. It combines color information (in CIELAB color space with pixel color vector $[lab]$) and image location (with pixel position vector $[xy]$ ) in order to produce superpixels.

Achanta, et al. ( 2012) introduces a new 5D distance function or metric, that allows the generation of approximately superpixels' sizes. SLIC has two parameters: compactness and number of superpixels $(K)$. The compactness parameter defines a balance between color-similarity and spatial proximity.

The number of superpixels indicates the number of centers for a k-means procedure, whose basic steps were described in chapter 2. Then, for an image with $N$ pixels, the initial approximate size of each superpixels is $N/K$ pixels.

Euclidean distances related with CIELAB color space and pixel position are showed in equation 2-6 and equation 2-7 respectively. The sum of these distances is denoted $D_s$ in equation 2-8, where cluster centers are $C_k = [l_k, a_k, b_k, x_k, y_k]$, with $k$ being an integer values, $0 < k < K$.

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \quad 2\text{-}6$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \quad 2\text{-}7$$

$$D_s = d_{lab} + \frac{m}{S} d_{xy} \quad 2\text{-}8$$

where, $S$ represents the distance between centers of adjacent superpixels, $S = \sqrt{(N/K)}$ . A parameter denoted by $m$ controls the compactness of a superpixel.

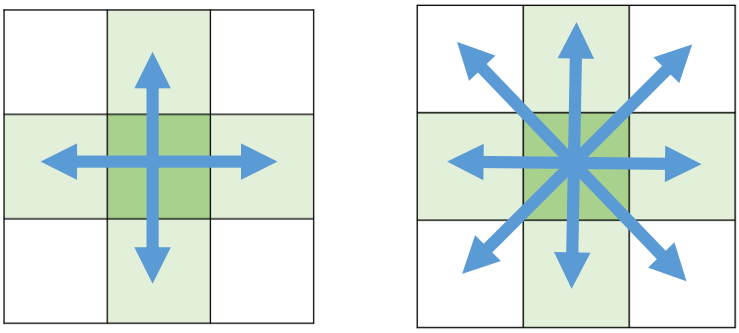
SLIC starts from an initial regular grid of superpixels separated by $S$. The initial superpixels deform through a number of iteration as the membership of each pixel to nearby superpixels are tested based on its distance to superpixels' centers, in procedure quite similar to k-means clustering. In this way, SLIC updates superpixels delineation and cluster centers repeatedly until convergence (Achanta, et al., 2012).

## 2.6. Conditional Random Fields – CRF

This section describes basic concepts underlying the Conditional Random Fields. Conditional Random Fields (CRF), proposed by (Lafferty, et al., 2001), is a popular undirected graphical model that describe conditional probability distributions to predict a label image in computer vision.

In many computer vision application, CRF is used to model a scene by a graph $G = (V, E)$, where $V$ represents a set of nodes and $E$ a set of edges. Each node $V_i \epsilon V$ corresponds to an image site, which might be a pixel, a superpixel, or a block of pixels in a regular grid, or segment. Each edge $E_{ij} \epsilon E$ connecting nodes $V_i$ and $V_j$ indicates a conditional dependence between them.

In computer vision a graph has the form of a lattice, where each node has four neighbors. Alternatively an eight-neighbor lattices can be used (see Figure 3.2).



4- Connected 8- Connected

Figure 2-2 Pixel connectivity, four and eight connected.

The set of class labels $\boldsymbol{y} = \{y_i\}$ denotes a particular class assignment over $V$, where $y_i$ refers to site $V_i$ and may take values within a finite set of classes. Similarly, $\boldsymbol{x} = \{\boldsymbol{x}_i\}$ denotes the set of observed feature vectors, where $\boldsymbol{x}_i$ refers to site $V_i$. The set of nodes connected to a site $V_i$ in $G$ is represented by $N_i$.

CRF models the posterior distribution $P(\boldsymbol{y}|\boldsymbol{x})$ of a class assignment $\boldsymbol{y}$ conditioned to the set of observations $\boldsymbol{x}$ as

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} exp\left(\sum_{V_i \epsilon V} A_i(y_i, \boldsymbol{x}) + \sum_{V_i \epsilon V} \sum_{V_j \epsilon N_i} I_{ij}(y_i, y_j, \boldsymbol{x})\right) \quad 2\text{-}9$$

where $Z(x)$ is defined as:

$$Z(\boldsymbol{x}) = \sum_{\boldsymbol{y}} exp\left(\sum_{V_i \epsilon V} A_i(y_i, \boldsymbol{x}) + \sum_{V_i \epsilon V} \sum_{V_j \epsilon N_i} I_{ij}(\boldsymbol{x}, y_i, y_j)\right) \quad 2\text{-}10$$

In equation 2-9 $Z$ is called partition function, and it is defined in equation 2-10. $Z$ is a normalizing constant that guaranties that $P(\boldsymbol{y}|\boldsymbol{x})$ add up to one. $A_i$ and $I_{ij}$ are called association potential for image site $V_i$ and the interaction potential relative to edge $E_{ij}$ that connects nodes $V_i$ and $V_i$ , respectively. These terms are described in the subsequent sections.

### 2.6.1. Association Potential

The Association potential links the data to the class labels, and determines the most likely label $y_i$ for a single image site $V_i$ given an observation $x$. The Association Potential is modeled to be proportional to the logarithm of this posterior probability (see equation 2-11). Therefore, any local classifier with a probabilistic output can be used. In this work, the Random Forest classifier (RF) (Breiman, 2001) was used.

$$A_i(\boldsymbol{x}, y_i) \leftrightarrow log\ P(y_i|\boldsymbol{x}) \quad 2\text{-}11$$

It is common practice to model the association potential by a function whose arguments are the observed value ($\boldsymbol{x}_i$) only at node $V_i$ instead of at all sites ($\boldsymbol{x}$), and the class label ($y_i$) at node $V_i$.

### 2.6.2. The Interaction Potential

The Interaction Potential represents the dependencies of a site $V_i$ on its adjacent image sites $V_j \in N_i$, which are connected to $V_i$ by edge $E_{ij}$. There are different methods to obtain the interaction potential, the simplest method is the Simple Potts model, and it was selected to model the spatial interaction potential in this work. It is defined as follow.

$$I_{ij}(y_i, y_j, \boldsymbol{x}) = I_{ij}(y_i, y_j) = \begin{cases} 0, if\ y_i = y_j \\ -\beta, if\ y_i \neq y_j \end{cases} \quad 2\text{-}12$$

This model only depends on the labels: different labels are penalized, whereas similar labels are not penalized. The degree of penalization depends on the value of the parameter $\beta$. This interaction potential has a smoothing effect on the labels since it favors neighboring sites with the same class label.

Cross validation is the standard way to estimate the value of $\beta$. However, would imply in long processing time. For this reason we decided to estimate the optimum values of β only upon the training data, we used a metric as an objective function, and we compare different outputs generated by different values of $\beta$. In section (see section 3.3.2) we come back to this issue.

### 2.6.3. Inference

Inference in CRFs corresponds to determining the optimal label configuration $\hat{\boldsymbol{y}}$, the one that maximizes $P(y|\boldsymbol{x})$, formally

$$\hat{\boldsymbol{y}} = \underset{\boldsymbol{y}}{\operatorname{argmax}} \left( \sum_{V_i \epsilon V} log\ P(y_i|\boldsymbol{x}_i) + \sum_{V_i \epsilon V} \sum_{j \epsilon N_i} I_{ij}(y_i, y_j) \right) \qquad 2\text{-}13$$

This graph structure is complex and usually has cycles, no explicit computation by message passing algorithms is possible. Thus, exact inference is intractable for 2D lattices. According to (Vishwanathan, et al., 2006) approximate methods are used for inference, in this study was used an algorithm called Loopy Belief Propagation (Frey, et al., 1998), which is a standard iterative message passing algorithm used for inference.

# 3
# Methodology

This chapter presents a general description of the methodology to compare semantic segmentation as an alternative to segmentation and as an alternative to the typical OBIA approach. Section 3.1 exhibits the metrics used to assess the segmentation and the classification task. Section 3.2 exposes the methodology used for the Supervised Segmentation Parameter Tuning. Section 3.3 defines the methodology used for Conditional Random Field the selected Semantic Segmentation method. Section 3.4 explains the methodology adopted for the selected Object Based method.



## 3.1.
## Thematic and Spatial accuracy metrics

According to (Gao, et al., 2011), the average size of the image sites (segments or superpixels) has a significant impact on the classification accuracy. For this reason, we evaluated the sensitivity of OBIA and CRF based approaches to the parameter most related to the site size, specifically, the *number of superpixels* for SLIC and the *Scale* parameter for MRS, both in terms of spatial and thematic accuracy. The spatial accuracy metric is used to assess segmentation outcomes and the thematic accuracy metrics are used to assess classification outcomes.

The next section explains the quality metrics we used to assess the spatial and thematic accuracies. The Thematic accuracy was evaluated using average accuracy (AA) and overall accuracy (OA), where a value of 1 would be a perfect classification and a value of 0 would be an unsatisfactory classification.

### 3.1.1.
### Spatial accuracy - F-measure

The Spatial accuracy was quantitatively evaluated using the $F$-measure (Van Rijsbergen, 1979). A value equal to 1 means a perfect match between segmentation result and references, whereas a value of 0 represents a complete

mismatch. The $F$-measure quantifies a trade-off between Precision ($P$) and Recall ($R$). Given a reference object $RO$ and the segment $S$ from the segmentation outcome with the largest overlap with $RO$, the F1 score is defined by equation 3-1

$$F1 = \frac{P \cdot R}{R + P} \quad 3\text{-}1$$

where Precision ($P$) and Recall ($R$) are defined, respectively (see Equation 3-2.) as

$$P = \frac{tp}{tp + fp} \qquad R = \frac{tp}{tp + fn} \quad 3\text{-}2.$$

where $tp$, is the true positives and represents the pixels from the reference segment ($RO$) that are also in the segment $S$. $fp$, so called false positives, represents the pixels from the segment $S$ that do not belong to the reference ($RO$). $fn$, the false negatives represent the pixels from the reference segment $RO$ that do not belong to the segment $S$. For an appropriate segmentation, the objective is to obtain a segment high related to the reference. It means that it is preferable more $tp$ and less $fn$ and $fp$.

Figure 3-1, on the left, shows a segmentation output, the letter $S$ in orange over the Figure 3-1 on the left, indicates the segment of interest of the segmentation outcome. The image on the right shows a reference segment ($RO$) in green. The yellowish intersection between the reference and the segment $S$ is the true positive $tp$. The false negatives $fn$ is the blueish region, whereas the false positive the $fp$ corresponds to the red region

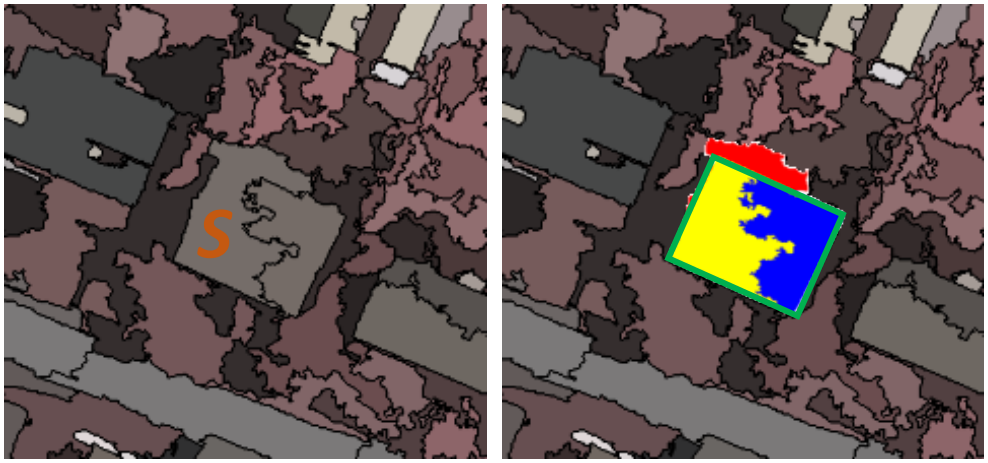



Figure 3-1 Left segmentation outcome. Right, spatial accuracy result, reference segment in green, $tp$ = yellow, $fn$ = blue and $fp$ = red.

### 3.1.2. Thematic Accuracy

There are many metrics for thematic accuracy. The research *Comparative assessment of the measures of thematic classification accuracy* (Liu, et al., 2007), summarizes different methods to evaluate thematic accuracy. In this work we express thematic accuracy by two of the mostly widely used metrics: Average Accuracy (AA) and Overall Accuracy (OA).

Both metrics derive from the confusion matrix. This is generally a $mxm$ square array where $m$ denotes the number of classes in the problem. Each element of the confusion matrix expresses the number of samples assigned by the classifier to a particular class relative to the actual class. In the confusion matrix (see Figure 3-2), the position $p_{ij}$, represents the proportion of pixels classified as $i$ in the classification outcome and the reference categorized as $j$.



| | | Reference Data | | | |
|---|---|---|---|---|---|
| **classified Data** | **1** | **2** | **...** | **m** | **total** |
| **1** | *P11* | *P12* | *...* | *P1m* | *P1+* |
| **2** | *P21* | *P22* | *...* | *P2m* | *P2+* |
| **...** | *...* | *...* | *...* | *...* | *...* |
| **m** | *Pm1* | *Pm2* | *...* | *Pmm* | *Pm+* |
| **total** | *P+1* | *P+2* | *...* | *P+M* | *N* |

Figure 3-2 Confusion Matrix

OA is computed by dividing the total number of pixel correctly classified (sum of elements along the diagonal of the *confusion matrix*) by $N$, the total number of pixels (Congalton, 1991). OA is defined by equation 3-3.

AA is defined as the normalized sum of the relation between the numbers of pixels classified correctly in each class and the total number of pixels in that respective class. It is defined by equation 3-4.

$$OA = \frac{1}{N}\sum_{i=1}^{m} p_{ii} \quad \text{3-3.}$$

$$AA = \frac{1}{m}\sum_{i-m}^{m} \frac{p_{ii}}{pi_{+}} \quad \text{3-4.}$$

## 3.2. Supervised Segmentation Parameter Tuning Methodology

The Supervised Segmentation Parameter Tuning approach finds the appropriate segmentation parameter values using a tool to tuning the set of parameters. In this study the *MRS* algorithm was used as a representative of most used bottom-up segmentation approaches within OBIA.

Parameter tuning was accomplished by SPT tool. The SPT underlying procedure requires that the user provides a set of reference segments that represent what should be regarded as a "good segmentation outcome", the reference segments are described in section 4.3 . SPT searches the parameter space so as to maximize the similarity between outcome and references (see Section 2.4). Three parameters were tuned, the *scale*, the *color* weight, and the *compactness* weight. For simplicity, the weights assigned to the bands ($\omega_L$) were all set to 0.33 (Diaz, 2014).

## 3.3. Semantic Segmentation Methodology

This study is focused on a specific SSeg model based on CRF as described in chapter 3. The pixel-wise classification through CRF can be intractable for medium to large images. For this reason, instead of pixels our SSeg implementation classified superpixels generated by the SLIC algorithm.

Figure 3-3 shows two examples of sites generated from a scene for different superpixel sizes, the superpixel size in the SLIC algorithm is controlled through a parameter *K* described in section 2.5. Note that SLIC produces nearly regular sites in terms of size and shape compared to Bottom Up methods, this uniformity increases according to the number of superpixels in the scene (see Figure 3-3). A further characteristic of superpixels that distinguish them from segments produced by bottom-up algorithms is that, due to their nearly regular shapes, the number of adjacent superpixels is almost constant for all sites. This is convenient because the Interaction Potential of CRF considers for each site all its neighbors.

Figure 3-3 Image sites generated by SLIC for few (large) and many (smaller) superpixels.

### 3.3.1. SSeg Processing Steps

The Semantic Segmentation methodology consists of six main steps: image sites generation, features extraction, training, association potential estimation, interaction potential estimation, and CRF inference.

1. Site generation: the sites required for the CRF were generated using SLIC.
2. Features extraction: the site descriptors were computed as explained in section 4.2.
3. Training: the Random Forest (RF), which will provide the association potentials for CRF, is trained in this step (see section 4.4).
4. Computation of association potential: with the RF trained in the previous step, the association potential of all sites being classified are estimated.
5. Computation of interaction potential: the optimum value of parameter $\beta$ is determined for all sites being classified (see section 2.6.2).
6. Inference: the labels of the test sites are determined via CRF using Loopy Believe Propagation.

The Conditional Random Field models were implemented using the Undirected Graphical Model (UGM) library available on the website: \\www.cs.ubc.ca/~schmidtm/Software/UGM.html, (Schmidt, 2007).

### 3.3.2. Tuning the Interaction Potential

The central objective of this dissertation is to evaluate semantic segmentation (SSeg) under two perspectives, firstly, as an alternative to segmentation into the object based image analysis and secondly as an alternative to the typical OBIA approach, which involves both segmentation and classification, SSeg does both simultaneously.

The interaction potential must be properly tuned in each case. In the present context it involves setting up the parameter $\beta$ (see equation 2-12).

In some of the experiments reported in Chapter 4 the optimum $\beta$ value was computed by searching for the maximum of a given objective function. In the comparison of SSeg with bottom-up segmentation, the F1-Score (see section 4.6.1) was the objective function over the training samples.

In the comparison of SSeg with OBIA the thematic accuracy was the focus. The overall and the average class accuracies were the objective function used for the computation of the optimum value for $\beta$, (see section 4.6.1).

Cross validation is the standard approach to estimate $\beta$. However, it can involve large processing time. For this reason in this work the optimum values of $\beta$ only upon the training data.

For the first task, we used a variant of Harmony Search algorithm ( (Geem, et al., 2001), (Contreras, et al., 2014) ) available in MATLAB, as the optimization procedure and for the second task, we used an algorithm based on golden section search and parabolic interpolation available in MATLAB.

## 3.4. OBIA Methodology

In this work, we do not apply potential further improvement steps, which in OBIA would follow the initial segmentation and classification. The basic OBIA approach can also be divided into four steps: segmentation, feature extraction, training and classification.

In the first step, initial objects are generated by some bottom-up segmentation algorithm, in this work was used MRS. The importance of segmentation has been emphasized by many authors in the last ten years or even

longer (Vantaram & Saber, 2012; Dey, et al., 2010; Neubert, et al., 2008). The problem of choosing a segmentation algorithm, and, once it has been selected, tuning its parameters so that the image is partitioned in a convenient way, has been acknowledged as the critical step of OBIA processing chain.

A thorough analysis of the alternatives addressing this issue could not be accommodated in a dissertation. So, we decided to use in our experiments the MRS algorithm, briefly explained in section 2.3, because it is knowingly the most widely used algorithm within the OBIA community (Tilton & Lawrence, 2000).

Among its input parameters, the *scale parameter* is the critical one, followed by the *color* and *compactness* weights. In order to render the analysis tractable under the dissertation's constraints, we fixed *color* and *compactness* to 0.5 and set the band weights to the same value for all bands. The impact of segmentation quality over OBIA's thematic accuracy was assessed by varying the *scale* parameter.

In the second step of OBIA processing chain, features are extracted from each segment to form the so called segments descriptors. Each segment was described by a feature vector containing the average feature values of all pixels enclosed by that segment.

Next, a classifier is trained based on a set of labeled sites (supervised classification). To select the training segments we used the same strategy adopted for the SSeg implementation (see section 4.4). Segments having more than 70% overlap with the regions shown Figure 4-5 were taken for training and the remaining ones for testing purpose.

Finally, in the fourth step, segments are classified based on their feature values. Even though OBIA allows for very sophisticated classification strategies, we decided to use a Random Forest for the classification task, so as to provide a common basis for comparison between OBIA and SSeg.

# 4
# Experimental Analysis

This chapter reports the experiments carried out with the purpose of assessing semantic segmentation under different scenarios. The analysis had two main objectives. First to compare SSeg with MRS segmentation in terms of spatial accuracy, and second, to compare SSeg with a basic OBIA from the perspective of thematic accuracy. The sensitivity of CRF to its parameter was also addressed.

Section 4.1 describes the dataset used for all the experiments. Section 4.2 presents the Feature set used for classification. Section 4.3 defines the training and test procedure for segmentation parameter tuning. Section 4.4 describes the training and test data for the classification task. Section 4.5 reports the experiments carried out to compare semantic segmentation and the bottom-up segmentation in terms of spatial accuracy. Section 4.6 reports the experiments carried out to compare semantic segmentation and a basic Object based Image analysis strategy in terms of thematic accuracy.



## 4.1.
## Dataset description.

The dataset used in these experiments comprises 2 high-resolution remote sensing images, with corresponding ground truth. The selected images have heterogeneous objects like buildings, streets, trees and cars in very high-resolution data, which carry high intra-class variance and, in some cases, low inter-class variance.

The dataset covers about 7.4 × 4.7 km² of Vaihingen, a neighborhood 25km north-west of Stuttgart, Germany. The dataset was provided by the *German Association of Photogrammetry, Remote Sensing and Geoformation* (DGPF) (Cramer, 2010): http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html.

The ground truth was produced by visual interpretation and comprises five land cover classes: ‘Building, ‘Low vegetation’, ‘Tree’, ‘Car, and ‘Street’.

The images are referred henceforth as Image 1 and Image 2. They correspond to Area 13 (Figure 4-1) and the Area 17 (see Figure 4-2) of Vaihingen dataset, respectively. Both images cover residential areas mostly characterized by small separated houses.

The images have a spatial resolution of 8 cm and comprise three bands: red, blue and near infrared. The Digital Surface Model (DSM) represents the earth's surface including all objects on it. A DSM with spatial resolution of 8 cm is also available for each area. Image 1 is an array of 2818×2558 pixels, whereas Image 2 is 2336×1281 pixels large.

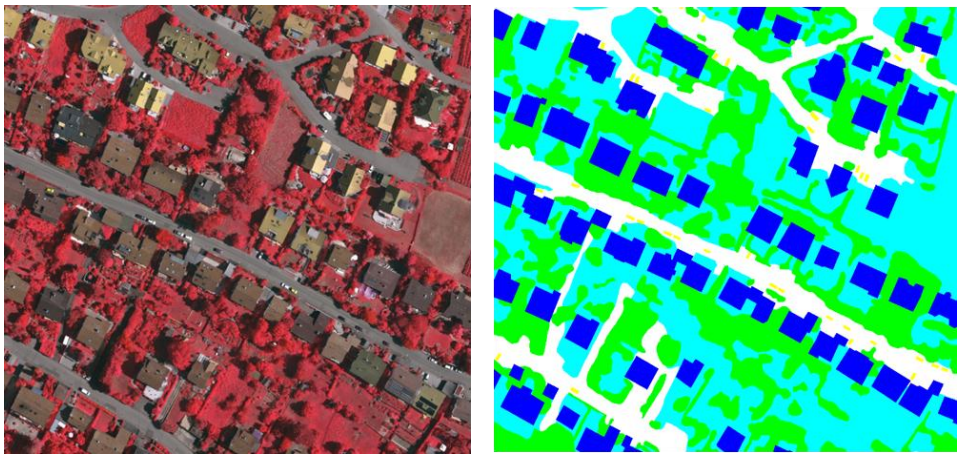


Figure 4-1: (left) Image 1, Vaihingen Area 13; (right) Ground Truth: 'Building' (blue), 'Low vegetation' (Cian), 'Tree' (Green), 'Car' (yellow) and 'Street' (white).

The false color composition (Red-Blue-NIR) and the corresponding ground truth (GT) for both images are presented in Figure 4-1 and Figure 4-2.

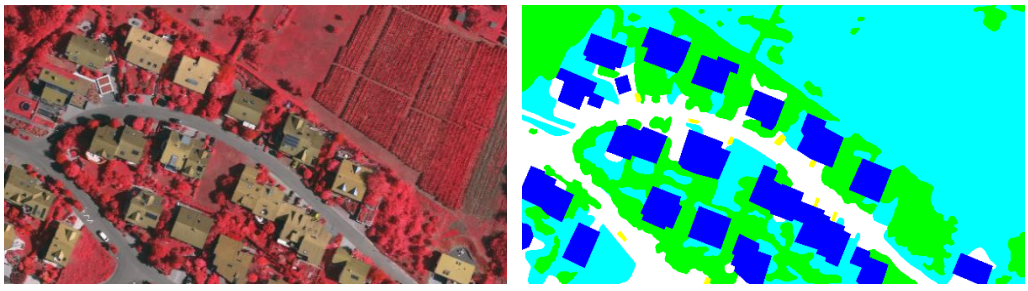


Figure 4-2: (left) Image 2, Vaihingen Area 17; ; (right) Ground Truth: 'Building' (blue), 'Low vegetation' (Cian), 'Tree' (Green), 'Car' (yellow) and 'Street' (white).

## 4.2. Features

Four groups of features were available for the experiments. They were:

- Three spectral bands corresponding to the *near infrared, red and green*, attribute with dimension 3.
- Normalized Difference Vegetation Index (NDVI), an index of the photosynthetic activity, attribute with dimension 1.
- Digital surface model (DSM), representing the height data, attribute with dimension 1.
- The outcomes of a set of Gabor filter banks at five scales and eight orientations, which represent texture. Attribute with dimension 40.

For classification, each image site was described by a feature vector containing the average feature values of all pixels enclosed by that site (segment or superpixel). For classification all aforementioned features were exploited, building up a 45 dimensional descriptor for each image site. Generally, for segmentation using MRS and for superpixels generation using SLIC algorithm only the spectral bands are considered.



## *4.3.* Training and test procedure for *segmentation parameter tuning*

As mentioned before, the Segmentation Parameter Tuning (SPT) tool was used to tune the parameters of the MRS segmentation algorithm.

For parameter tuning only segments of the class ‘Building’ were taken as reference. The objects of classes ‘Road’, ‘Tree’ and ‘Low vegetation’ can hardly be embraced by a single segment, and are consequently improper references for the SPT approach. Objects of class ‘Car’, on the other hand, are much smaller than ‘Buildings’. Consequently, a good *scale* value for ‘Car’ is normally too small for ‘Buildings’, or vice-versa.

Thus, only ‘Building’ samples have been used as references for SPT. In Image 1, nine samples were considered for training and forty for testing. Figure 4-3 shows the references selected for training (blue) and for testing (green).

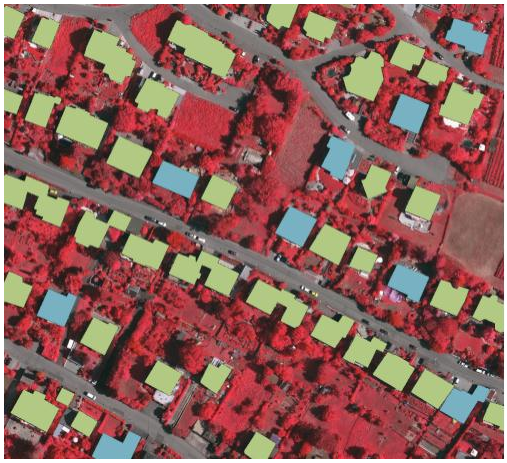


Figure 4-3 Reference Segment of Image 1 for SPT

In Image 2, nine building samples were selected for training (blue polygons in Figure 4-4) and ten for testing (green polygons in Figure 4-4).

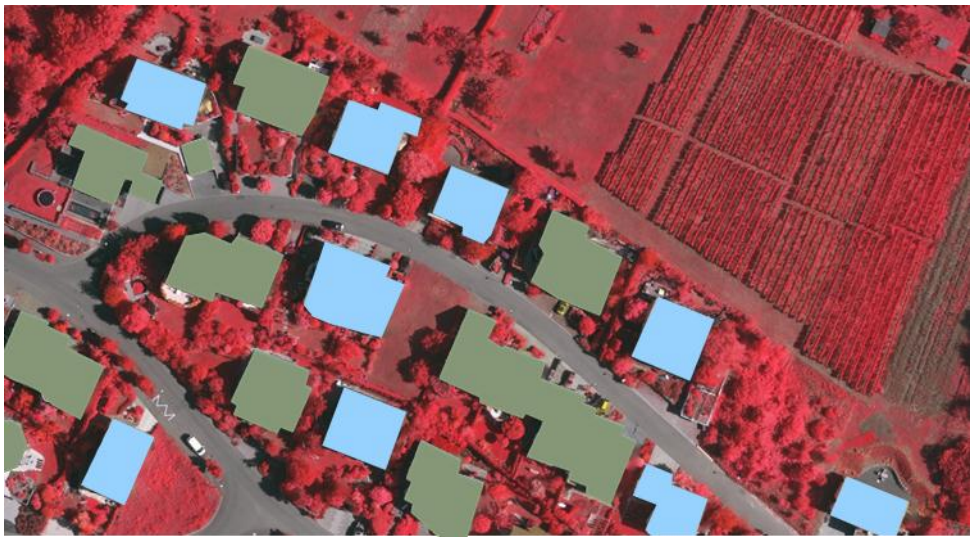


Figure 4-4 Reference Segment of Image 2 for SPT



## *4.4.* Selecting training and test data for *SSeg*

The training data was chosen randomly, it is shown in Figure 4-5. Superpixels having at least 70% of its area inside a training region were used for training. The remaining superpixels were used for test.

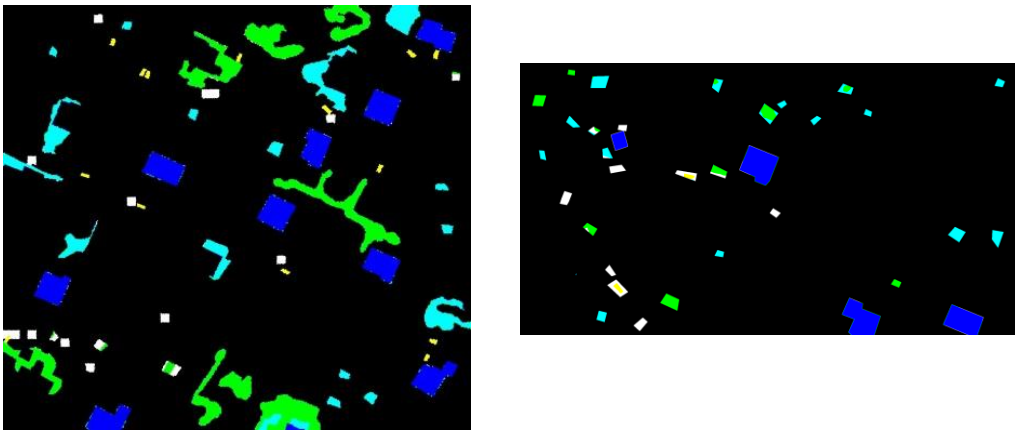


Figure 4-5 labeled training data for Image 1 (left) and Image 2 (right): 'Building' (blue), 'Low vegetation' (Cian), 'Tree' (Green), 'Car' (yellow) and 'Street' (white).

Table 4-1 and Table 4-2 present for Image 1 and Image 2, respectively, the approximate percentage of pixels of each class used for training and for test.

| Class | Train | Test |
|---|---|---|
| Building | 3.6 | 14.5 |
| tree | 4.3 | 25.8 |
| Low vegetation | 2.8 | 32.4 |
| Car | 0.2 | 0.2 |
| Road | 0.7 | 15.6 |

Table 4-1 percentage of pixels of Image 1 used for training and test

| Class | Train | Test |
|---|---|---|
| Building | 2.1 | 14.6 |
| tree | 0.6 | 25.1 |
| Low vegetation | 0.9 | 43.5 |
| Car | 0.1 | 0.3 |
| Road | 0.6 | 12.2 |

Table 4-2 Percentage of pixels of Image 2 used for training and test

## 4.5. Spatial accuracy of SSeg and MRS.

The experiments reported in this section were designed to compare semantic segmentation (SSeg) and supervised segmentation parameter tuning (SSPT) in terms of spatial accuracy.

*Testing supervised segmentation parameter tuning*

Image 1 and Image 2 were segmented using the MRS algorithm, whose parameter values were estimated by the SPT tool based on the bluish references shown in Figure 4-3 and Figure 4-4, respectively.

Table 4-3 shows the optimal parameter values found by SPT for Image 1. Figure 4-6 on the left shows the segmentation outcome produced with these parameter values. Some buildings' segments matched perfectly their references. It should be noted that some buildings having half-dark and half bright halves were split in two segments (see red circles in Figure 5-5 on the right).

| Parameters tuned | Value |
|---|---|
| Scale. | 80 |
| Color weight. | 0.2 |
| Compactness weight. | 0.74 |

Table 4-3 Parameters tuned for Image 1

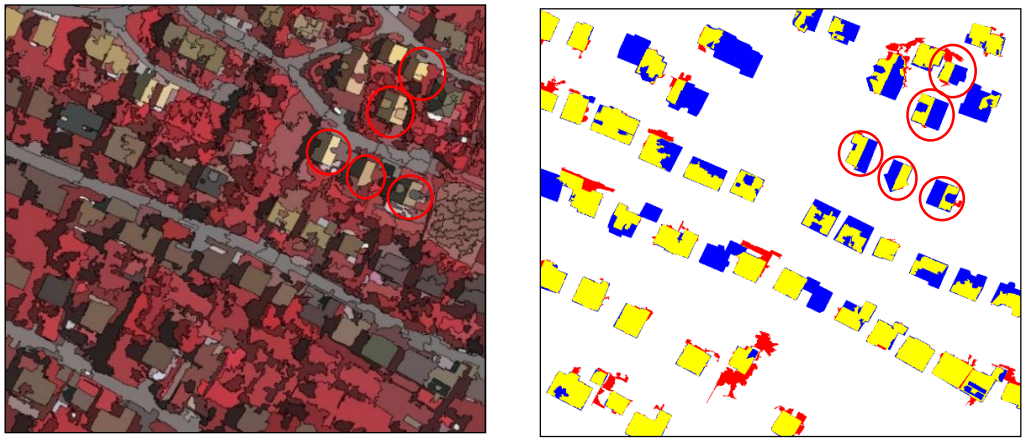


Figure 4-6: Segmentation outcome for Image 1 (left); positive and negatives (yellow=TP, red=FP, blue=FN) (right).

Figure 4-7 shows a zoom over the region containing the aforementioned red circles. Clearly, roofs with a non-uniform surface were divided into multiple segments, even two of these buildings were used as reference segments. In this experiment the $F$- measure was 0.7004. Figure 4-7 (right) also shows many false negatives in the segmentation output (blue), mainly due to non-uniform illumination, causing an over-segmented outcome.

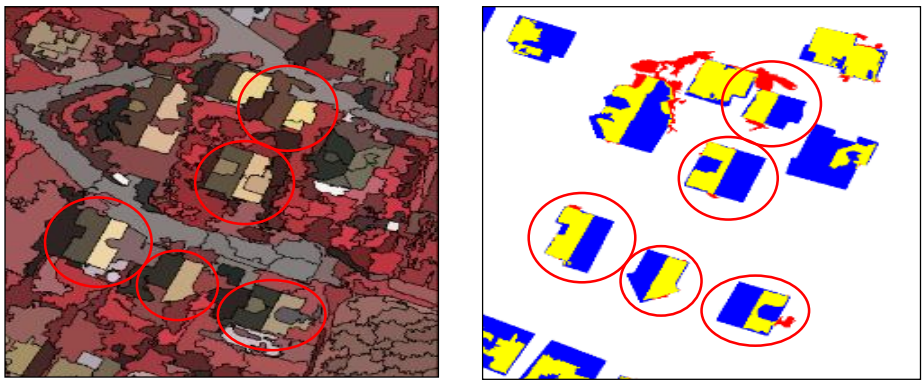


Figure 4-7 Zoom over the region with red circles first image.

*Testing semantic segmentation and comparison with parameter tuning*

The selected SSeg model is based on CRF. Two alternative approaches have been tested to generate image sites: superpixels (SP) and small segments produced by the MRS algorithm. For the CRF using superpixel ($CRF + SP$), different superpixels' sizes in a range of 4000-140000 were tested.

We have confirmed experimentally that small values of $\beta$ makes CRF permissive regarding changes of classes. These class changes appear as a Salt & Pepper effect. On the other hand, a high values of $\beta$ produce a smoothing effect. Large $\beta$ values may induce over-smoothing as it can be seen in Figure 4-8 for Image 1 and Figure 4-11 for Image 2. The objects of classes 'Road', 'Tree' and 'Low vegetation' can hardly be embraced by a single segment, and are consequently improper for segmentation evaluation. Objects of class 'Car', on the other hand, are much smaller than 'Buildings'. Therefore, to make a fair comparison between SSPT and SSeg, for segmentation assessment only segments of the class 'Building' were considered.

Figure 4-8 (a), (b) and (c) show the segmentation of Image 1 produced by the SSeg model based on CRF working upon 140.000 superpixels for different $\beta$ values. In Figure 4-8 (a), a short value of $\beta$ ( $\beta = 0.1$), produces a large quantity of segments (3149), most of them formed by few pixels. Figure 4-8 (b) shows

that a large value of $\beta$ ($\beta = 1.4$) may imply in few segments (261), due to the fusion of segments that should be separated. With the best value of $\beta$ ($\beta = 0.7$) found for this configuration, SSeg produced 551 segments. This was the best segmentation result obtained with SSeg and using superpixels as images sites. In this case the obtained F-measure was $F1 = 0.8123$. The best result obtained with SSPT for Image 1 is shown in Figure 4-8 (d) corresponding to $F1 = 0.7004$. This is clearly inferior to the result achieved by SSeg for the best $\beta$.

.

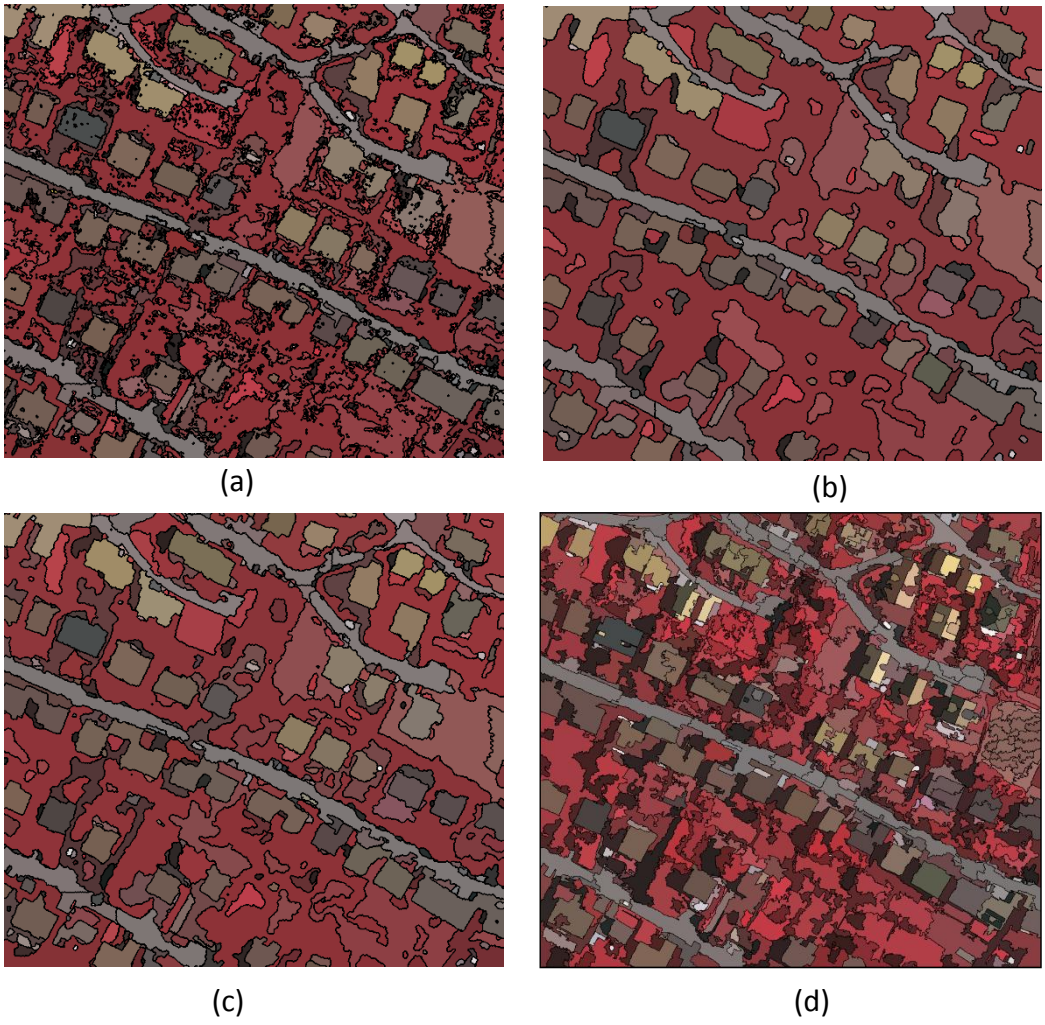



Figure 4-8 CRF using 140.000 superpixels with different values of β, (a) Small value, $\beta = 0.1$, (b) Large value, $\beta = 1.4$, (c) medium value, $\beta = 0.7$. (d) Supervised segmentation parameter tuning for Image 1.

Figure 4-9 shows the segmentation results for the given references (see Figure 4-3). The figure on the left shows the optimum results for the SSeg and on the right for supervised segmentation SSPT, both for Image 1. SSeg produced

much less false negatives than SSPT, because the semantic information helped to merge spectrally inhomogeneous parts of the roofs. SSeg managed to delineate most building almost perfectly.

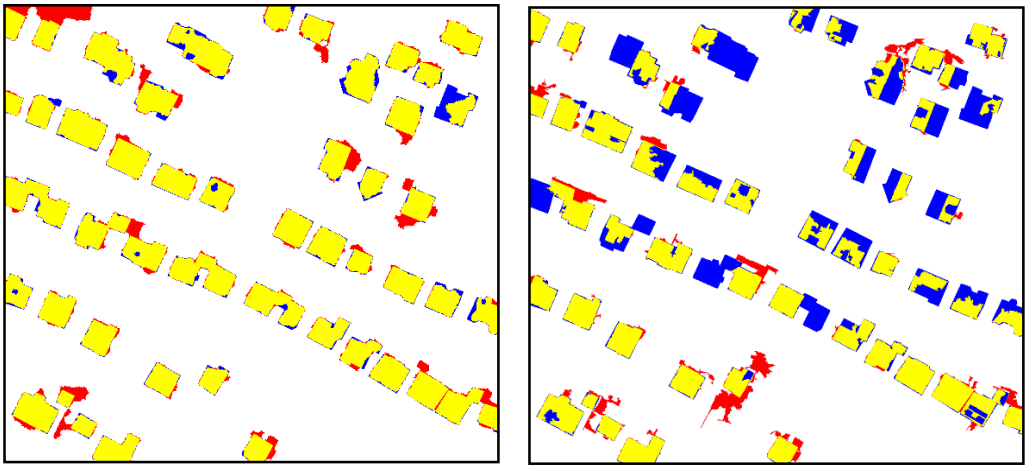


Figure 4-9 Positive and negatives for Image 1(yellow=TP, red=FP, blue=FN) produced by SSeg (left) and SSPT (right).

Table 4-4 shows the optimal parameter values found by STP for Image 2. Figure 4-10 shows the corresponding segmentation outcome.

The SSPT method did not produce good results, as many roofs were divided into multiple segments. In some cases, both the half-dark and half bright roof parts were split in several parts. Even the buildings inside the yellow circles in Figure 4-10 were used as references to train the SPT and were divided into multiple segments. These results show clearly the limitations of this method to produce a single segment out of a spectrally inhomogeneous object. In this experiment the SSPT approach achieved the spatial accuracy $F1 = 0.788$.

| Parameters tuned | Value |
|---|---|
| Scale. | 245 |
| Color weight. | 0.604 |
| Compactness weight. | 0.473 |

Table 4-4 Parameters tuned for Image 2

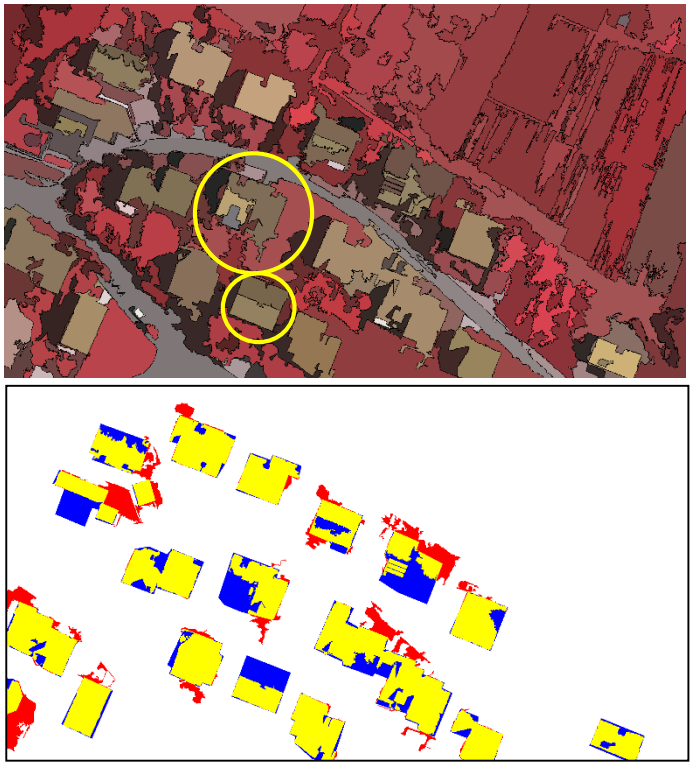


Figure 4-10 Segmentation outcome for Image 2 positive and negatives (yellow=TP, red=FP, blue=FN) (down).

Figure 4-11 (a), (b) and (c) show the segmentation SSeg results for Image 2. Once again 140.000 superpixels have been used as image sites. Figure 4-11 (a), shows the results obtained with a short value of $\beta$ ( $\beta = 0.1$), which produced many segments (1732). Figure 4-11 (b) corresponds to a large value of $\beta$ ($\beta = 2$), where few segments (84) were generated. Figure 4-11 (c) shows the best results obtained for this configuration using a value of $\beta = 1.45$ that led to 117 segments. In this case, the spatial accuracy F-measure obtained was $F1 = 0.92$. Figure 4-11 (d) shows the segmentation result for the supervised segmentation parameter tuning for Image 2. It obtained a spatial accuracy of $F1 = 0.788$ discussed above, again a result substantially inferior to SSeg.

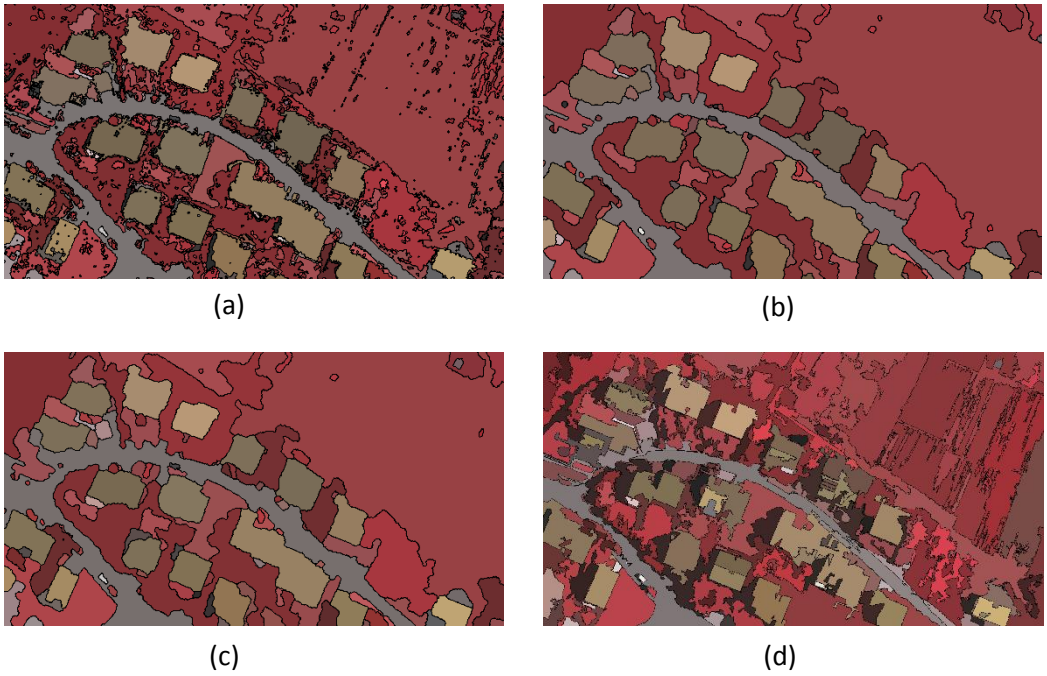



Figure 4-11 Results for CRF using 140.000 superpixels for different values of β: (a) small $\beta = 0.1$; (b) Large $\beta = 2$; (c) medium $\beta = 1.45$; (d) results for supervised segmentation parameter tuning for Image 2.

Figure 4-12 shows the segmentation results for the given references (see Figure 4-4). The figure on the left shows the results for the SSeg and the figure on the right shows the results for SSPT for Image 2. SSeg segmentation has left more false negatives than SSPT for the same reason than in Imagen 1. However, it can be seen large false positives pixels in the boundaries of the buildings.

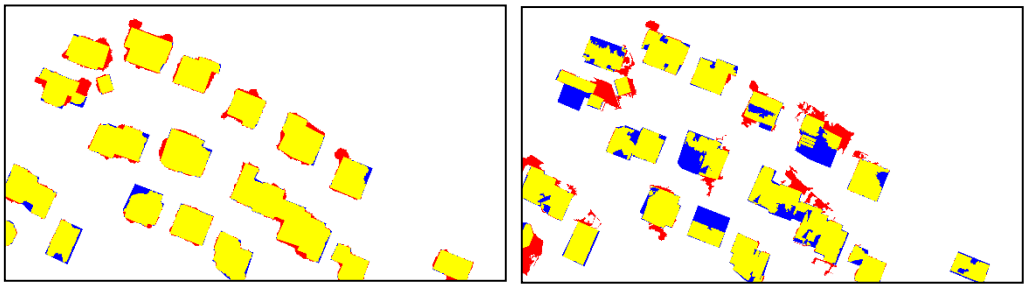


Figure 4-12: Positive and negatives for Image 2 (yellow=TP, red=FP, blue=FN) (left) for SSeg; (right) for supervised segmentation PT.

### 4.5.1. Sensitivity of CRF to superpixel size

The objective of this experiment was to assess the sensitivity of the CRF approach to the number of superpixels, or equivalently, to the average superpixel size. The experiment with CRF reported in the previous section was repeated for different number of superpixels.

Figure 4-13 shows the recorded results in terms of spatial accuracy ($F$-measure) as a function of the number of superpixels. Recall that the measurement was carried out only on segments of class "Buildings". For Image 1, all the values were close to each other, in a range between 0.7 to 0.82. The best spatial accuracy ($F = 0.8123$) occurred with tested number of superpixels (140.000), while the worst result occurred for 5000 superpixels.

For Image 2 we observed a similar behavior, although the $F$-measure dropped for larger (few) superpixels. This behavior is explained with more details in the subsequent paragraphs.

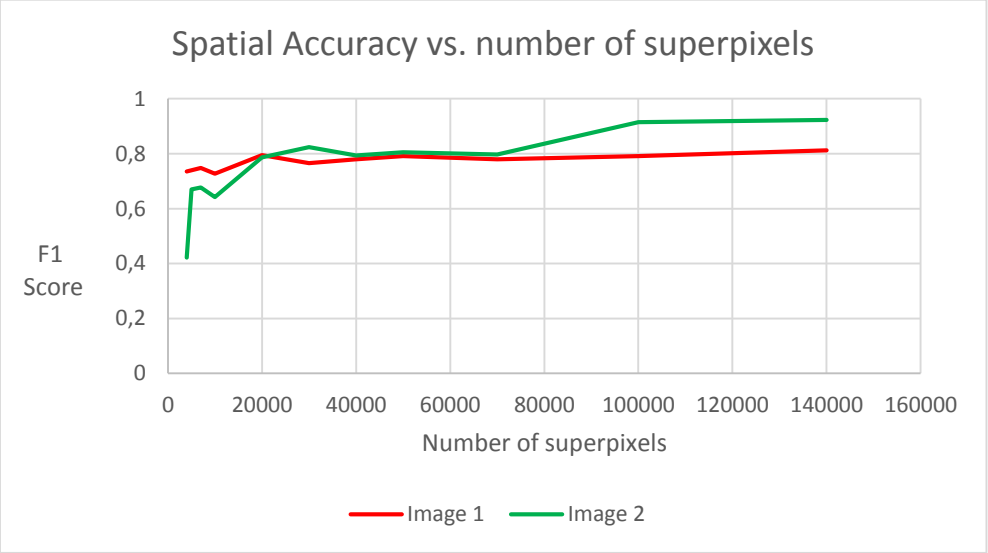



Figure 4-13 CRF spatial accuracy vs. number of superpixels

Figure 4-14 and Figure 4-15 depict the evaluation of the segmentation for the SSeg model based on CRF for Image 1 and 2. Figure 4-14 and Figure 4-15 show the results for different experiments using different values of the parameter *number of superpixels (SP)*. For Figure 4-14 and Figure 4-15, the images on the right present less false positives (red color) in almost all segments than in the images on the left.

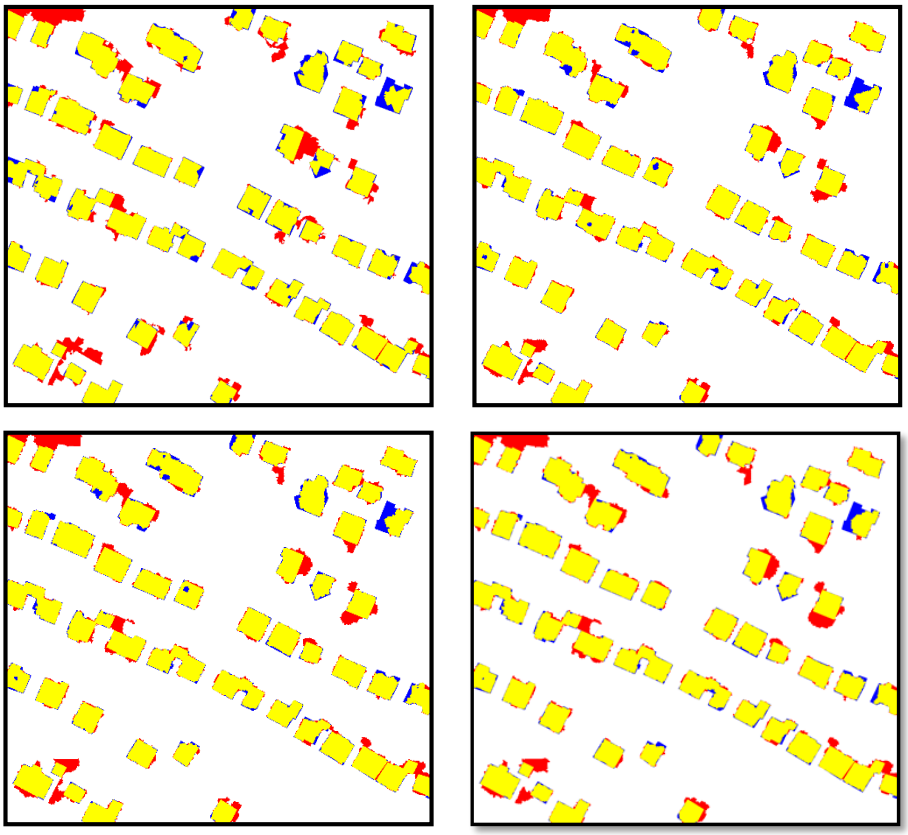


Figure 4-14 Image 1 results for reference segments (yellow=TP, red=FP, blue=FN). Upper left, CRF using 4000 SP. Upper right, CRF using 140000 SP. Bottom left, CRF using 50000 SP. Upper right, CRF using 70000 SP.

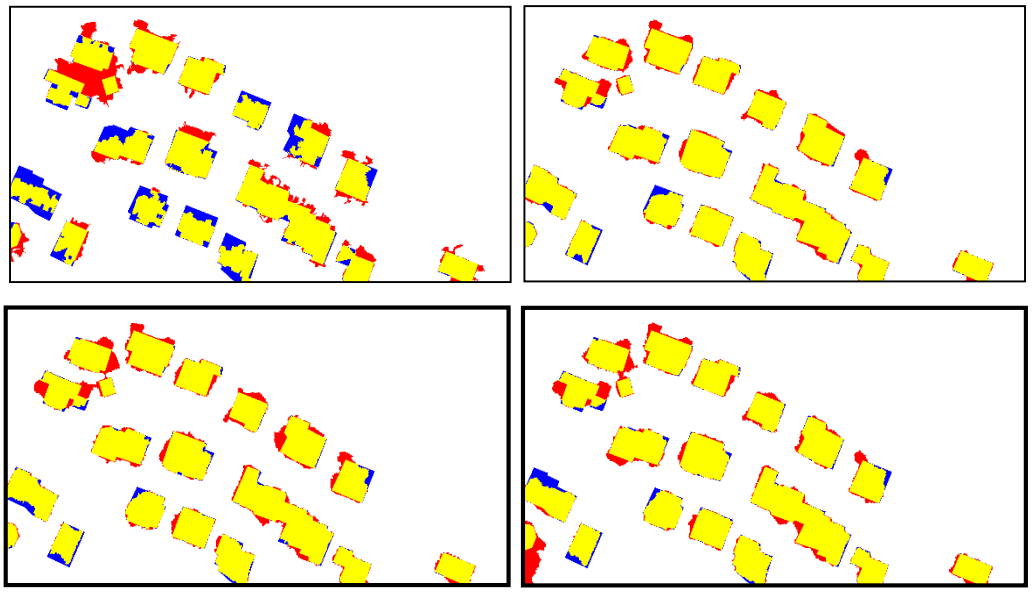


Figure 4-15 Image 2 results for reference segments (yellow=TP, red=FP, blue=FN). Upper left, CRF using 4000 SP. Upper right, CRF using 140000 SP. Bottom left, CRF using 50000 SP. Upper right, CRF using 70000 SP.

### 4.5.2. Sensitivity of CRF to parameter $\beta$

Figure 4-16 and Figure 4-17 summarize the results of an experiment that aimed at assessing the sensitivity of spatial accuracy to parameter $\beta$ that balances the association and the interaction potential in CRF approach for Image 1 and Image 2. The number of superpixels was set to three different values. The plots lead to the conclusion that in terms of spatial accuracy, CRF might perform worse than classifier used by CRF to produce the association potentials. When the parameter $\beta$ corresponds to $\beta = 0$, the Interaction potential term is not considered, therefore, only the classifier is taking into account. These results show that the proper estimate is critical in the CRF approach, at least in what refers to spatial accuracy.

The curves are different for both images given the distributions of the buildings. The gaps between the buildings are smaller for Image 1 than for Image 2. Consequently, Image 2 requires higher values of $\beta$ than Image 1, on the other hand, in some cases the buildings in Image 1 were so close that some values of $\beta$ induces the fusion of buildings that should be separated.

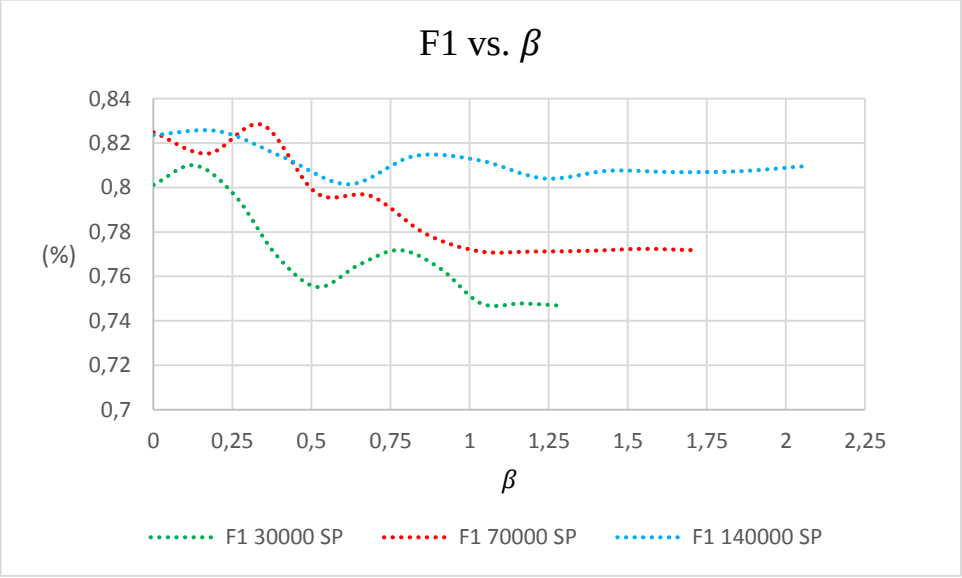



Figure 4-16 Image 1, spatial accuracy F- measure vs. $\beta$

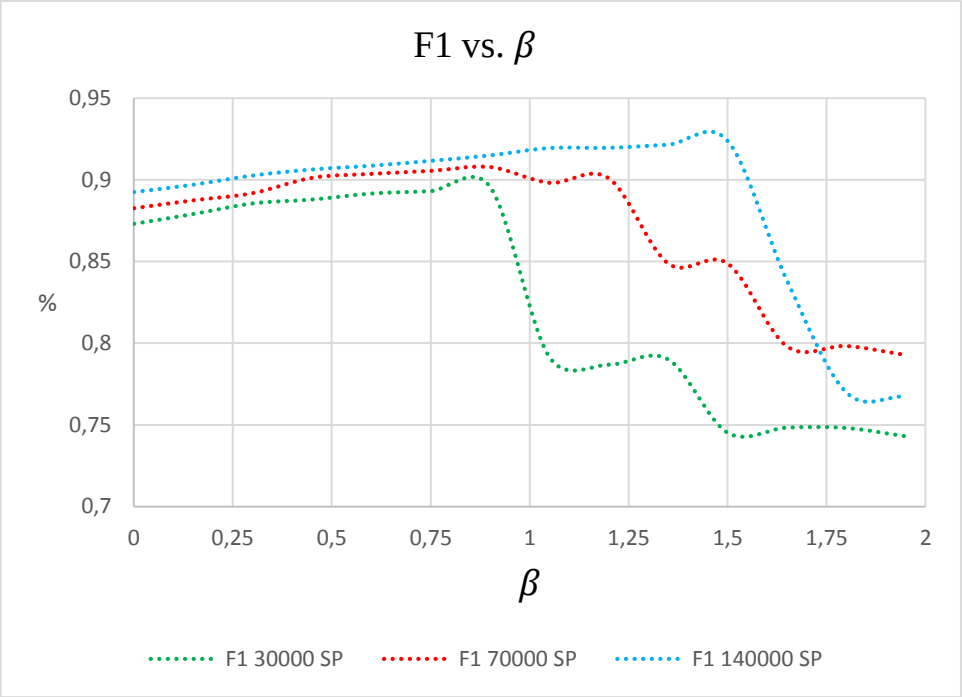



Figure 4-17 Image 2, spatial accuracy F-measure vs. β



## 4.6. Thematic accuracy of SSeg and OBIA

The experiments described in this section aimed at comparing Semantic Segmentation with the typical OBIA strategy in terms of thematic accuracy.

### 4.6.1. Thematic accuracy of semantic segmentation

$CRF$ was tested for the number of superpixels varying in a range 4000 to 140000. For each of these experiments the best value of $\beta$ was determined using the same procedure adopted in the previous experiments.

Figure 4-18 shows how the optimum $\beta$ varies with the number of superpixels. As discussed in the preceding section, $\beta$ represents the penalty for class change. In other words, when $\beta$ has larges values the smoothing effect increases. The curves for Image 1 and Image 2 in Figure 4-18 show that the optimum $\beta$ tends to increase with the number of superpixels. In other words, the smaller the superpixels size the higher is the optimum $\beta$. This can be explained by

the following rationale. The number of superpixels that cover a meaningful image object increases as the superpixels become smaller. So, $\beta$ must increase so that the smoothing effect propagates over mover superpixels to avoid false class changes inside the region comprised by said object.

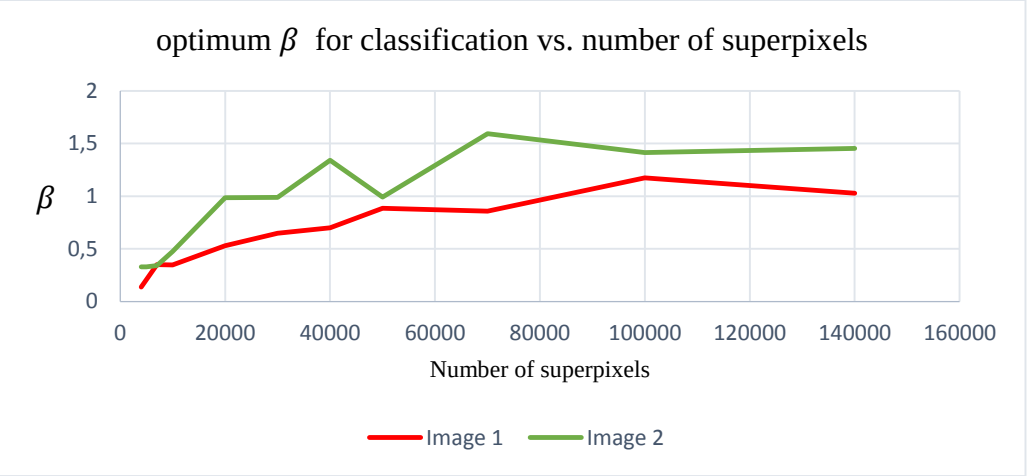



Figure 4-18 optimum $\beta$ for classification vs. number of superpixels

Figure 4-19 for Image 1 and Figure 4-20 for Image 2, show the classification results of the SSeg model for $\beta$ below (a), above (b) and equal(c) to the optimum as well as the ground truth (d). In all cases the number of superpixels was set to 140,000.

Figure 4-19 (a) and Figure 4-20 (a) show the results for a $\beta$ lower than the optimum. The Salt & Pepper effect is visible. Figure 4-19 (b) and Figure 4-20 (b) show how a large values of $\beta$ induce an over smooth effect in the outcome, these values of $\beta$ produce the merging of regions that should be separated. Good examples are the small 'Grass' regions of Figure 4-19 (a), which were merged to larger regions classified as 'Tree' in Figure 4-19 (b). It also occurs in the opposite direction: regions classified as 'Tree' for small $\beta$ become larger 'Grass' by increasing $\beta$.

Figure 4-19 (c) and Figure 4-20 (c) show the results obtained with the optimum value of $\beta$.

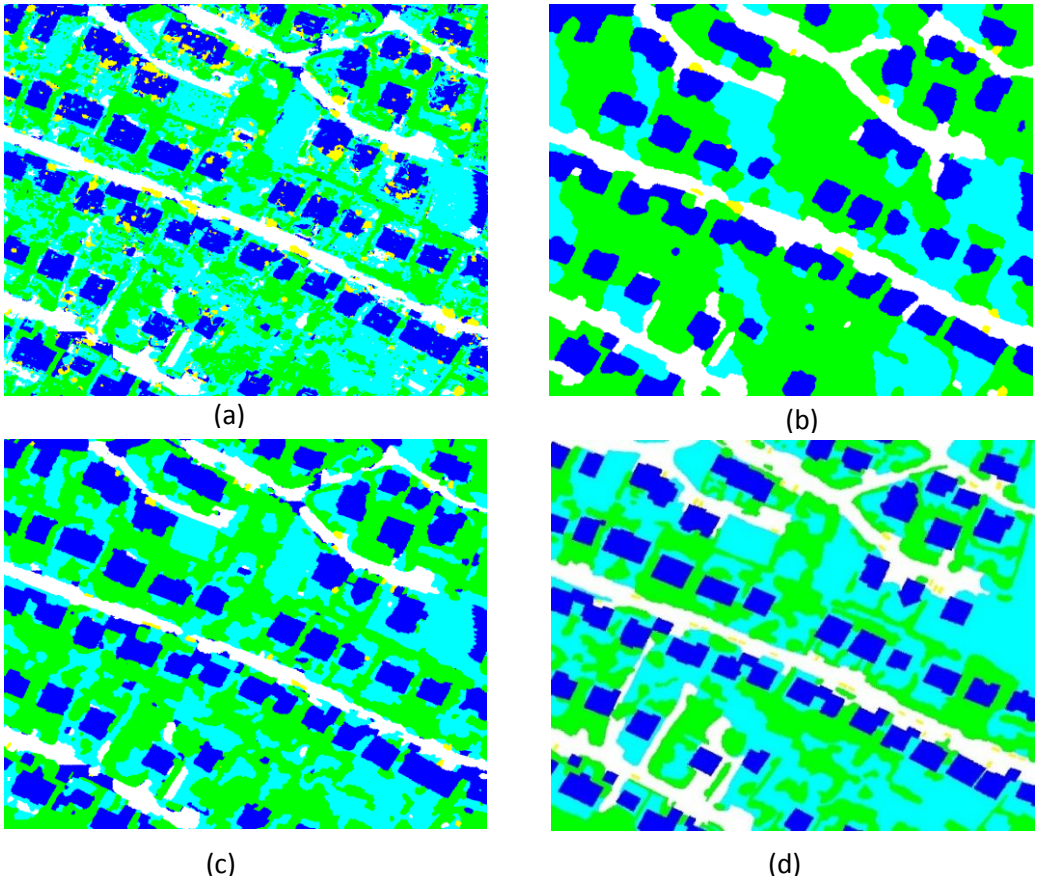



Figure 4-19 Classification results Image 1 of the SSeg model for β below (a), above (b) and equal (c) to the optimum as well as the ground truth (d). In all cases the number of superpixels was set to 140,000.

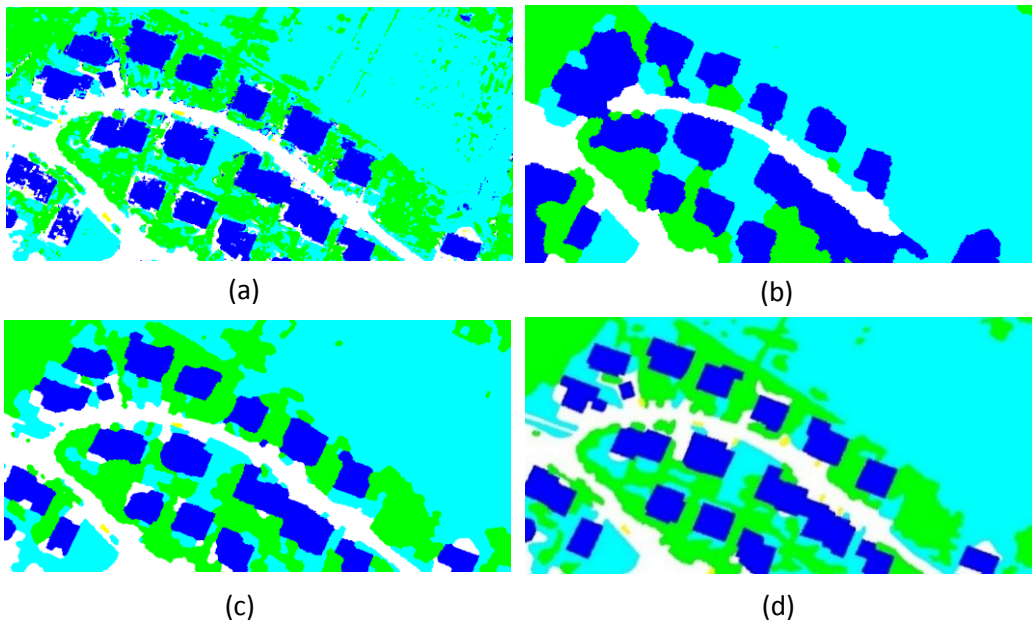



Figure 4-20 Classification results Image 2 of the SSeg model for β below (a), above (b) and equal (c) to the optimum as well as the ground truth (d). In all cases the number of superpixels was set to 140,000.

Figure 4-21 shows the average accuracy (AA) and overall accuracy (OA) obtained using as inputs the Image 1 and 2 for the optimum $\beta$. The OA and AA curves are similar. OA initially increases with the number of superpixels until it reaches a saturation value.

For Image 1, the highest overall accuracy was $OA = 0.7614$, using 30000 superpixels. For Image 2, the highest value was $OA = 0.849$, using 140000 superpixels. For more superpixels, OA was almost constant around $OA = 0.75$ and $OA = 0.84$ for the Image 1 and 2, respectively.

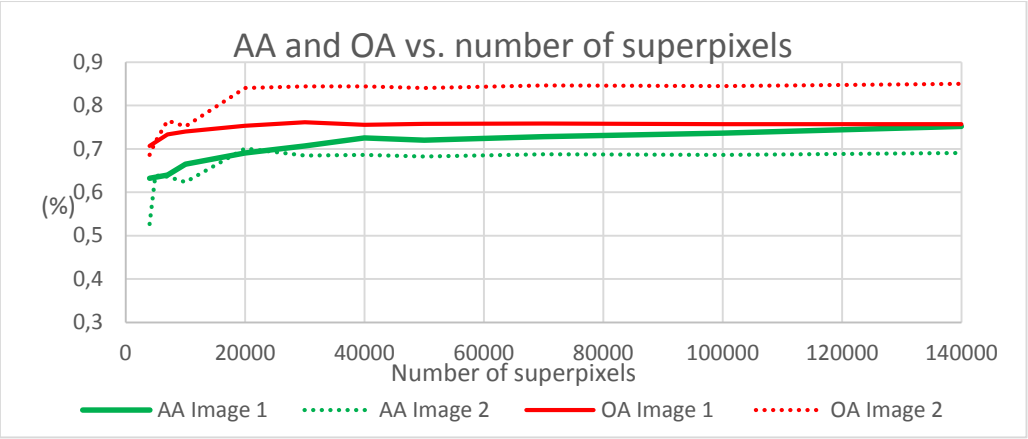



Figure 4-21. Average Accuracy and Overall Accuracy for different values of number of superpixels for the Image 1 and Image 2.

Similarly, AA increases with the number of superpixels, achieving the highest values $AA = 0.75$ for Image 1 using 30.000 superpixels and $AA = 0.69$ for Image 2, for about 40000 superpixels. For more/smaller superpixels the accuracy does not change considerably.

Table 4-5 presents the confusion matrix used to calculate the thematic accuracy for Image 1 and 140.000 superpixels. In this experiment, the classes 'Building' and 'Tree' were the two classes better classified with 93.7% and 81.9% accuracy respectively, the classes 'Grass' and 'Car' achieved 70% and for the class 'Street' 61% was obtained.

| | | Reference Data | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | B | T | G | C | S | Total | Land Cover Categories |
| Classified Data | B | 21910 | 221 | 852 | 389 | 13 | 23385 | B = Building |
| | T | 393 | 34097 | 6624 | 320 | 210 | 41644 | T = Trees |
| | G | 2790 | 9948 | 36750 | 1391 | 1689 | 52568 | G = Grass |
| | C | 9 | 0 | 34 | 228 | 57 | 328 | C = Car |
| | S | 3403 | 934 | 4565 | 973 | 15413 | 25288 | S = Street |
| Total | | 28505 | 45200 | 48825 | 3301 | 17382 | 143213 | |

**Average Accuracy** AA = 75.19%

**Overall Accuracy** OA = 75.69%

Table 4-5 Confusion matrix for Image 1 with 140,000 SP

It can be seen that Building' was detected accurately. Classification of Grass' was less accurate, mainly, caused by the confusion with Tree'. Being both vegetation this is understandable.

Table 4-6, Table 4-7, Table 4-8 and Table 4-9 show the confusion matrix for Image 1 using 4.000, 30.000, 100.000 and 140.000 superpixels respectively. As mentioned above the average and overall accuracy increase according to the number of superpixels. It can be seen that the classes "building" and "tree" were not very affected by the superpixels' size, because the objects belonging to these classes are large and were composed by many superpixels in all the cases. On the other hand, the results obtained by classes "car", "grass" and "Street" improved according to the number of superpixels mainly the class which has smaller objects ("car").



| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | B | T | G | C | S | Land Cover Categories |
| Classified Data | B | 0.93 | 0.02 | 0.05 | 0 | 0 | B = Building |
| | T | 0.01 | 0.89 | 0.1 | 0 | 0 | T = Trees |
| | G | 0.06 | 0.34 | 0.58 | 0.01 | 0.01 | G = Grass |
| | C | 0.43 | 0 | .29 | 0.29 | 0 | C = Car |
| | S | 0.17 | 0.05 | 0.29 | 0.02 | 0.48 | S = Street |

**Average Accuracy** AA =63.4%

**Overall Accuracy** OA=70.6%

Table 4-6 Confusion matrix for Image 1 with 4,000 SP

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | **B** | **T** | **G** | **C** | **S** | **Land Cover Categories** |
| **Classified Data** | **B** | 0.96 | 0.01 | 0.03 | 0 | 0 | **B = Building** |
| | **T** | 0.01 | 0.81 | 0.18 | 0 | 0 | **T = Trees** |
| | **G** | 0.06 | 0.18 | 0.75 | 0 | 0.01 | **G = Grass** |
| | **C** | 0.10 | 0.01 | 0.34 | 0.48 | 0.06 | **C = Car** |
| | **S** | 0.15 | 0.04 | 0.25 | 0.02 | 0.54 | **S = Street** |

**Average Accuracy** AA =70.6%

**Overall Accuracy** OA=76.1%

Table 4-7 Confusion matrix for Image 1 with 30,000 SP

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | **B** | **T** | **G** | **C** | **S** | **Land Cover Categories** |
| **Classified Data** | **B** | 0.94 | 0.01 | 0.04 | 0.01 | 0 | **B = Building** |
| | **T** | 0.01 | 0.82 | 0.16 | 0 | 0.01 | **T = Trees** |
| | **G** | 0.06 | 0.21 | 0.69 | 0.02 | 0.02 | **G = Grass** |
| | **C** | 0.05 | 0 | 0.10 | 0.60 | 0.25 | **C = Car** |
| | **S** | 0.13 | 0.04 | 0.17 | 0.03 | 0.63 | **S = Street** |

**Average Accuracy** AA =73.6%

**Overall Accuracy** OA=75.7%

Table 4-8 Confusion matrix for Image 1 with 100,000 SP

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | **B** | **T** | **G** | **C** | **S** | **Land Cover Categories** |
| **Classified Data** | **B** | 0.94 | 0.01 | 0.04 | 0.02 | 0 | **B = Building** |
| | **T** | 0.01 | 0.82 | 0.16 | 0 | 0.01 | **T = Trees** |
| | **G** | 0.05 | 0.19 | 0.70 | 0.02 | 0.03 | **G = Grass** |
| | **C** | 0.03 | 0 | 0.10 | 0.70 | 0.17 | **C = Car** |
| | **S** | 0.13 | 0.04 | 0.18 | 0.04 | 0.61 | **S = Street** |

**Average Accuracy** AA =75.2%

**Overall Accuracy** OA=75.7%

Table 4-9 Confusion matrix for Image 1 with 140,000 SP

Table 4-10 presents the confusion matrix used to calculate the thematic accuracy for Image 2 and 140.000 superpixels. In this experiment, the classes ‘Building’, ‘Tree’, “grass” and “street” were classified with 95%, 80%, 84% and 85% of accuracy respectively, on the other hand the class ‘Car’ achieved 0%.

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | B | T | G | C | S | Total |
| Classified Data | B | 16740 | 120 | 311 | 0 | 396 | 17567 |
| | T | 526 | 24153 | 4843 | 0 | 534 | 30056 |
| | G | 1068 | 5885 | 43985 | 0 | 1124 | 52062 |
| | C | 52 | 36 | 28 | 0 | 139 | 255 |
| | S | 489 | 686 | 991 | 0 | 12485 | 14651 |
| Total | | 18875 | 30880 | 50158 | 0 | 14678 | 114591 |

**Land Cover Categories**
**B = Building**
**T = Trees**
**G = Grass**
**C = Car**
**S = Street**

**Average Accuracy** AA = 79.35%

**Overall Accuracy** OA = 84.97%

Table 4-10 Confusion matrix for Image 2 with 140,000 SP

Figure 4-21 shows the remarkable difference between AA and OA for Image 2 due to the misbehavior of the RF classifying car class. Therefore, for Image 2 using 140.000 SP was obtained $AA = 0.79$ compared with $OA = 0.84$. Figure 4-22 shows some examples of cars in Image 2. In Image 2 there are few cars and high intraclass variance in this class, almost all the cars are different and consequently the samples of the car class used for training are not enough to discriminate this class. However, the same training data was used for both, SSeg and OBIA, affecting both methods in the same manner.

Figure 4-22 samples of cars in Image 2.

Table 4-11, Table 4-12 and Table 4-13show the confusion matrix for Image 2 using 7.000, 40.000 and 140.000 superpixels respectively. As mentioned above the average and overall accuracy increase according to the number of superpixels. It can be seen that the classes "building" and "grass" were not very affected by the superpixels' size. On the other hand, the results obtained by classes "tree" and "Street" improved according to the number of superpixels until 40.000 superpixels. The confusion matrices for 40.000 and 140.000 superpixels do not change considerably.

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | **B** | **T** | **G** | **C** | **S** | **Land Cover Categories** |
| **Classified Data** | **B** | 0.98 | 0 | 0.01 | 0 | 0.01 | **B = Building** |
| | **T** | 0.06 | 0.60 | 0.31 | 0 | 0.03 | **T = Trees** |
| | **G** | 0.09 | 0.06 | 0.83 | 0 | 0.02 | **G = Grass** |
| | **C** | 0.65 | 0 | 0.10 | 0 | 0.25 | **C = Car** |
| | **S** | 0.21 | 0.01 | 0.04 | 0 | 0.74 | **S = Street** |

**Average Accuracy** AA =63.5%

**Overall Accuracy** OA=76.5%

Table 4-11 Confusion matrix for Image 2 with 7,000 SP

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | **B** | **T** | **G** | **C** | **S** | **Land Cover Categories** |
| **Classified Data** | **B** | 0.95 | 0.01 | 0.02 | 0 | 0.02 | **B = Building** |
| | **T** | 0.02 | 0.80 | 0.16 | 0 | 0.02 | **T = Trees** |
| | **G** | 0.02 | 0.12 | 0.84 | 0 | 0.02 | **G = Grass** |
| | **C** | 0.05 | 0.25 | 0.10 | 0 | 0.61 | **C = Car** |
| | **S** | 0.05 | 0.05 | 0.06 | 0 | 0.84 | **S = Street** |

**Average Accuracy** AA =68.5%

**Overall Accuracy** OA=84.37%

Table 4-12 Confusion matrix for Image 2 with 40,000 SP

| | | Reference Data | | | | | |
|---|---|---|---|---|---|---|---|
| | | **B** | **T** | **G** | **C** | **S** | **Land Cover Categories** |
| **Classified Data** | **B** | 0.95 | 0.01 | 0.02 | 0 | 0.02 | **B = Building** |
| | **T** | 0.02 | 0.80 | 0.16 | 0 | 0.02 | **T = Trees** |
| | **G** | 0.02 | 0.11 | 0.84 | 0 | 0.02 | **G = Grass** |
| | **C** | 0.20 | 0.14 | 0.11 | 0 | 0.55 | **C = Car** |
| | **S** | 0.03 | 0.05 | 0.07 | 0 | 0.85 | **S = Street** |

**Average Accuracy** AA =69.0%

**Overall Accuracy** OA=84.9%

Table 4-13 Confusion matrix for Image 2 with 140,000 SP

Figure 4-23 and Figure 4-24 present the results from a different perspective. It shows how the overall accuracy for the Image 1 and Image 2 varies with $\beta$ for different numbers of superpixels. The curves are mostly concave, reaching the maximum at different values of $\beta$ depending on the number of superpixels.

To the left of the maximum the accuracy decreases due to the Salt & Pepper effect. To the right, the accuracy decrease due to the over-smoothing effect. The maximum OA is obtained in an intermediate value of $\beta$, we call this $\beta$ as the $\beta$ optimum.

Figure 4-23 and Figure 4-24 show further that the curve becomes flat as superpixels become smaller. This means that the thematic accuracy is more sensitive to the proper estimate of $\beta$ when working with fewer/larger superpixels.

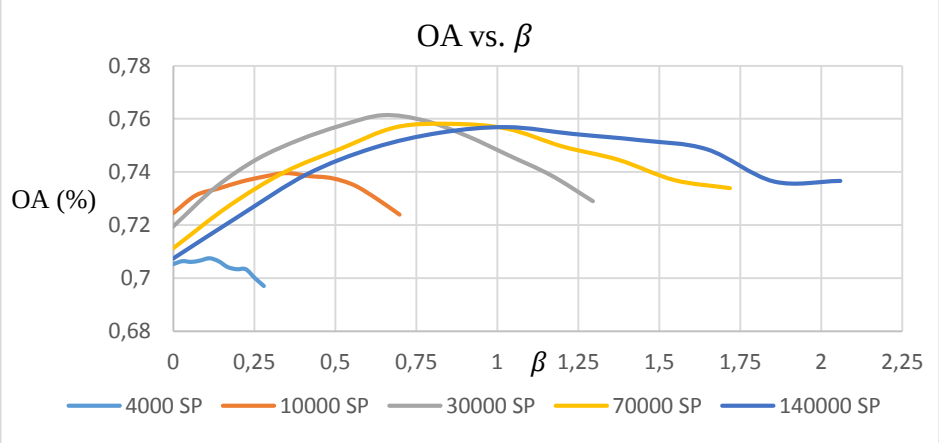



Figure 4-23 Overall Accuracy vs. $\beta$ for Image 1

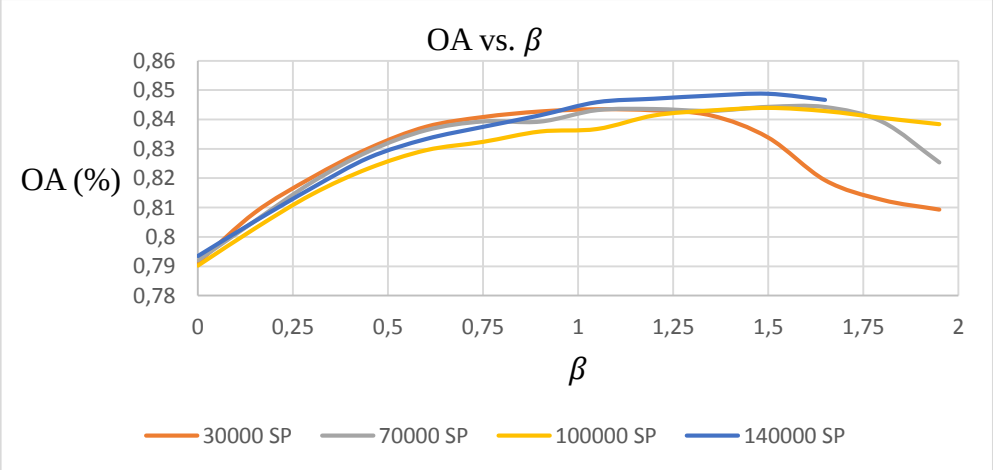



Figure 4-24 Overall Accuracy vs. β for Image 2

### 4.6.2. Thematic accuracy of basic OBIA results.

The experiments described in this section had as objective to assess the thematic accuracy of a basic OBIA processing chain for the same dataset used in the analysis of CRF.

The input images were segmented using MRS algorithm. Color and compactness parameters were set to 0.5 while the *scale* parameter took values in the range of 5-50. Again in these experiments the band weights were kept constant and equal for all bands.

The segmentation outcome can be seen in Figure 4-25 and Figure 4-26 for Image 1 and Image 2, respectively for the *scale* parameter set to 5 and 50.


Figure 4-25 Segmentation of Image 1 for scale parameter equal to 5 (a) and to 50 (b)

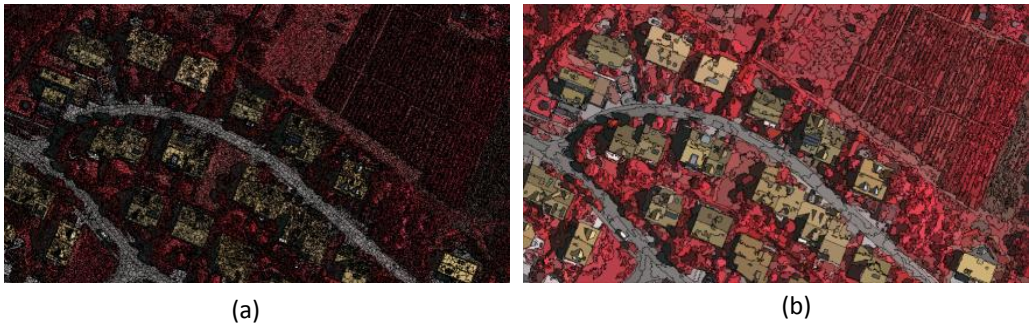



Figure 4-26 Segmentation of Image 2 for scale parameter equal to 5 (a) and to 50 (b)

Each segments was described by a vector comprising the average values of all features (see section 4.2).

For each scale, segments having more than 50% of its area inside the regions selected for training (see Figure 4-5 ) were taken for training. The remaining segments were separated for test. The classification step was carried out by a Random Forest classifier.

Figure 4-27 shows the average accuracy and the overall accuracy for Image 1 and Image 2 as a function of *scale*.

For Image 1, AA and OA were nearly constant. The highest value for OA was $OA = 0.63$ for *scale* = 5, and the highest value for AA was $AA = 0.62$ for *scale*= 30. Looking at Figure 4-25 we observe that even for the highest *scale* the segments were mostly smaller than the objects of interest. Although we didn't test it, the OA and AA curves for Image 1 are expected to go down for larger *scales*, due to single segments that spill over object borders. This effect is observed in the curves of Image 2, which achieve the highest values for *scale* 5, the lowest one tested in this experiment.



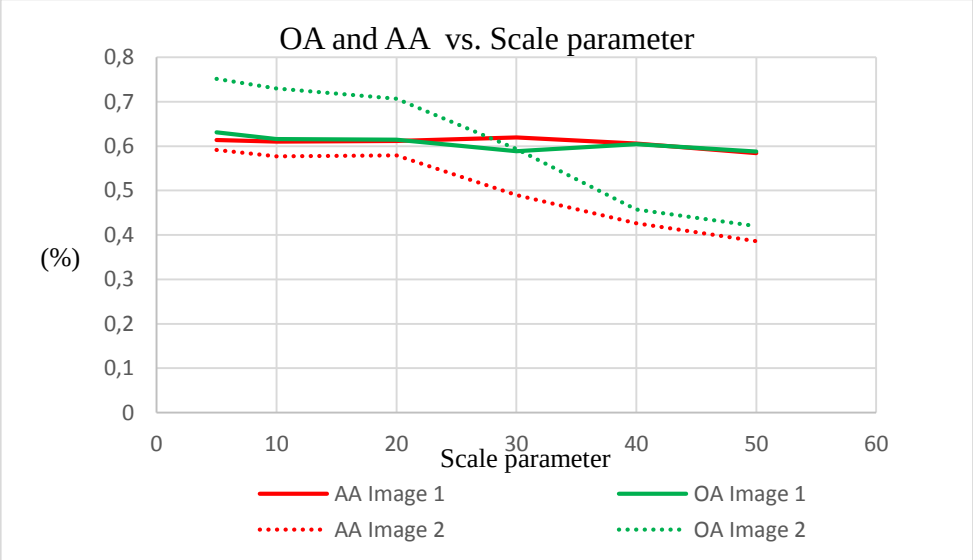



Figure 4-27 Average Accuracy and Overall Accuracy for different values of Scale parameter, OBIA.

Figure 4-28 on the left, presents the classification result for Image 2 using *scale* 5, the best results obtained in our experiments in this experiment. Figure 4-28 on the right, shows the classification results for the Image 2 using *scale* 20, some of the smoothing effect resulting from increasing the scale can be seen by comparing both parts of the Figure.

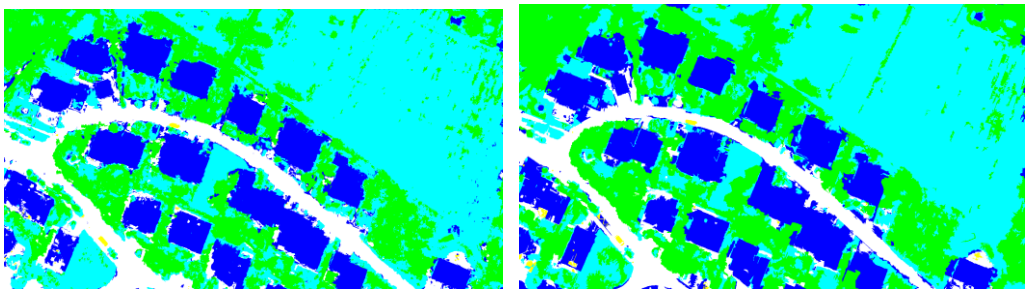


Figure 4-28 Left, classification result of the Image 2 using Scale 5 (left), classification results of the Image 2 using Scale 20 (right)

### 4.6.3. Comparing thematic accuracies

In this section the results for SSeg and OBIA reported in the two preceding sections are put side-by-side for comparison.

Figure 4-29 shows the best classification results delivered by SSeg and by OBIA for Image 1. SSeg was able to produce softer object contours and is much less affect by the Salt & Pepper effect than OBIA, Figure 4 27 shows the results for Image 2. The same behavior is observed.

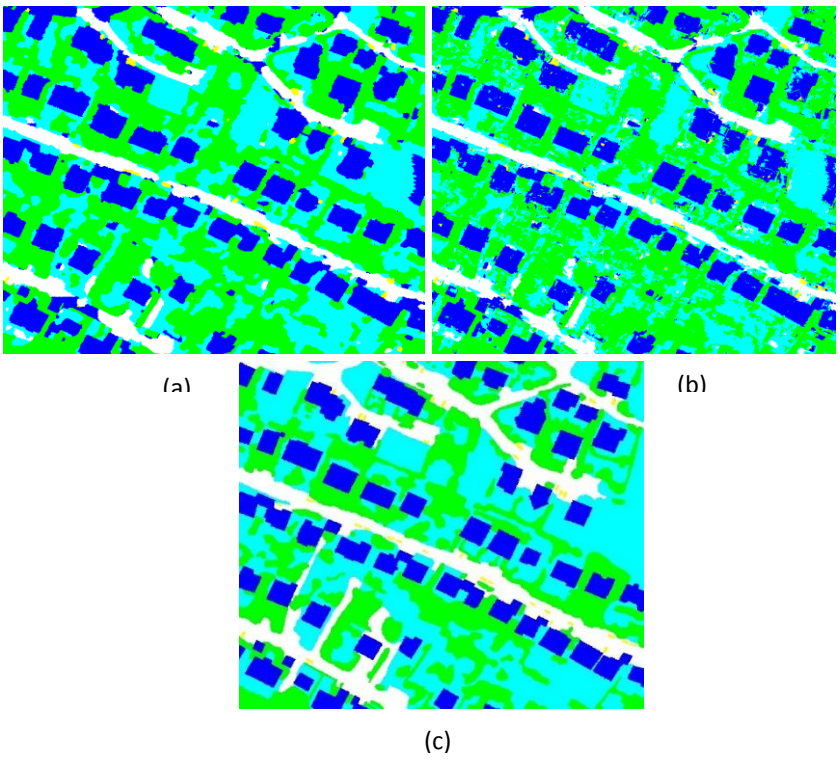



Figure 4-29 Best classification results for Image 1. (a) SSeg using superpixels. OBIA using over-segmented input image using (b) MRS. (c) Ground truth.

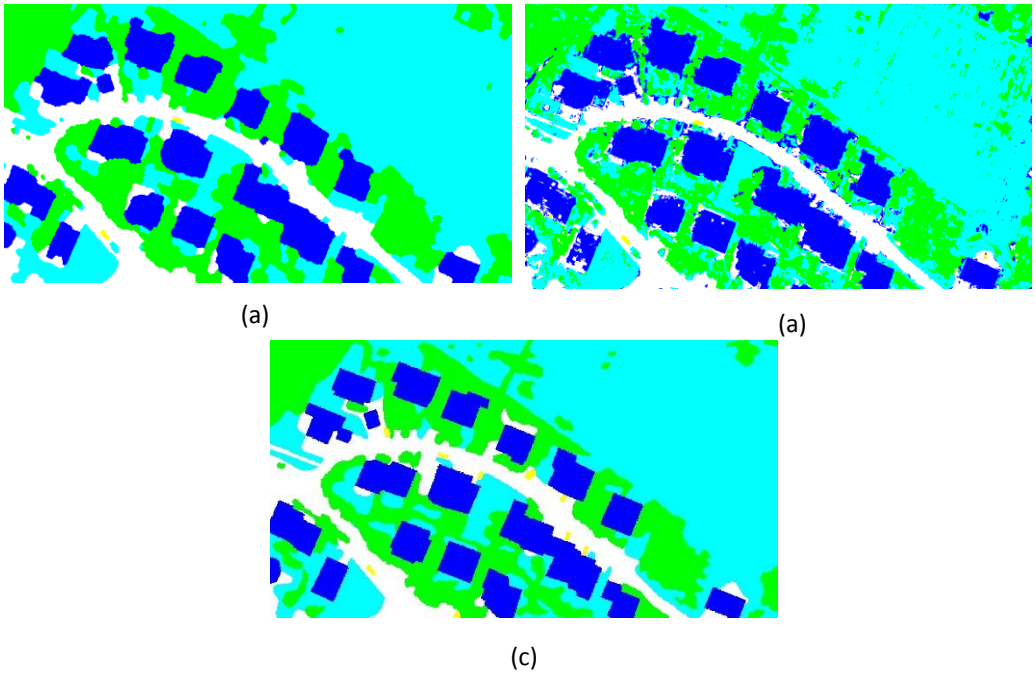



Figure 4-30 Best classification results for Image 2. (a) SSeg using superpixels. OBIA using over-segmented input image using (b) MRS. (c) Ground truth



The visual superiority of SSeg over OBIA observed in these Figures is corroborated by the measured accuracies, as reported in Tables 5-6 and 5-7.

| Method | OA | AA |
|---|---|---|
| OBIA (Scale 5) | 0.63 | 0.61 |
| SSeg (30,000 SP) | 0.76 | 0.71 |
| SSeg (140,000 SP) | 0.76 | 0.75 |

Table 4-14 Highest values for OA and AA for Image 1

| Method | OA | AA |
|---|---|---|
| OBIA (Scale 5) | 0.75 | 0.59 |
| SSeg (40,000 SP) | 0.842 | 0.684 |

Table 4-15 Highest values for OA and AA for Image 2

These figures clearly favor SSeg in detriment of OBIA. In all cases, SSeg outperformed OBIA in about 0.10, both in terms of overall and average accuracy.

It should be noted that the present study did not exploit the full OBIA potential. Classification strategies more sophisticated than a simple Random

Forest can be designed within the OBIA framework. Indeed, RF makes little or no use of explicit prior knowledge, as it is commonly done in OBIA based solutions. Besides, as mentioned in chapter 1, OBIA also admits schemes involving iterative segmentation + classification circles, a possibility not investigated in this dissertation.

Nevertheless, the results achieved in this work indicated that SSeg is an approach worth being considered as an alternative to OBIA for image classification or, at least as a building block of more elaborated OBIA based solutions.

# 5
# Conclusions

This dissertation reports a study with the aim at comparing semantic segmentation with the basic workflow Object-Based Image Analysis. Is important to mention that SSeg combines two basic OBIA operational steps in just one, namely: segmentation and classification. This is an advantage of SSeg compared to the usual OBIA workflow steps, that is, an initial segmentation followed by an initial classification. The two steps in OBIA are then usually to do iteratively: knowledge based segmentation improvement and re-classification.

First, the study considered semantic segmentation as an alternative to bottom-up segmentation. Semantic segmentation was compared with supervised segmentation parameter tuning in terms of spatial accuracy.

Second, semantic segmentation was compared with the typical OBIA strategy from the perspective of thematic accuracy.

Each approach investigated in this study was represented by a particular implementation. Specifically, Conditional Random Fields were used to represent Semantic Segmentation. The Multiresolution algorithm was chosen to represent bottom-up segmentation methods. Random Forest was the basic classifier used to produce association potentials for the Conditional Random Fields, as well as to perform the classification task in the OBIA approach.

The experiments conducted upon two very high resolution images indicated the superiority of Semantic Segmentation under both criteria, namely spatial and thematic accuracy.

The study still does not allow generalizing the aforesaid conclusion, mainly due to two reason. Firstly, because the number of experiments and the data set they relied upon are limited. Secondly and more importantly, because the spectrum of alternative bottom-up segmentation methods and potential of OBIA

of building complex classification strategies were not fully explored in this analysis.

Nevertheless, the results in section 4.5 demonstrated convincingly that Semantic Segmentation is at least worth being considered as part of an OBIA based solution for many image analysis problems. This could be done by replacing bottom-up segmentation by semantic segmentation, or even by using the semantic segmentation outcome as a preliminary classification result to be later refined by some knowledge based approach implemented as a rule set or any other typical OBIA scheme.

So, we envisage the investigation towards testing this idea on real image interpretation applications as a natural extension of this study, therefore SSeg could be used as an alternative for existing segmentation methods.

# 6
# Reference

**Achanccaray Diaz, P. M., Feitosa, R. Q. and Janeiro, Pontifícia Universidade Católica do Rio de. 2014.** A Comparison of Segmentation Algorithms for Remote Sensing. *Departamento de Engenharia Elétrica.* 2014.

**Achanccaray, P, et al. 2015.** SPT 3.1:A free Software for Automatic Tuning of Segmentation Parameters in Optical, Hyperspectral and SAR Images. *IGARSS.* 2015.

**Achanta, Radhakrishna, et al. 2012.** SLIC superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions.* 34, 2012, Vol. 11, 2274-2282.

**Agarwal, Pankaj K. and Procopiuc, Cecilia Magdalena. 2002.** Exact and approximation algorithms for clustering. 2002, Vol. 33, 2, pp. 201-226.

**Aha, D.W., Kibler, D. and Albert, M.K. 1991.** Instance-based learning algorithms. *Mach. Learn.* 6, 1991, pp. 37-66.

**Arbeláez, P., et al. 2012.** Semantic Segmentation using Regions and Parts. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* 2012, pp. 3378-3385.

**Avery, T. and Berlin, G. 1985.** *Fundamentals of Remote Sensing and Airphoto Interpretation.* s.l. : Maxwell Macmillan International, 1985.

**Baatz, M. and Schäpe, A. 2000.** Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung.* 2000, Vol. XII, 58, pp. 12-23.

**Baig, A, Bouridane, A and Kurogollu, F. 2008.** A Corner Strength Based Fingerpint Segmentation Algorithm with Dynamic Thresholding. . [ed.] Pattern Recognition. *ICPR 2008. 19th International Conference.* Dec 8, 2008, pp. 1-4.

**Ball, G.H. and Hall, D.J. 1965.** *Isodata, a Novel Method of Data Analysis and Pattern Classification;*. Menlo Park, USA : Stanford Research Institute, 1965.

**Bandyopadhyay, S. 2005.** Satellite image classification using genetically guided fuzzy clustering with spatial information. 2005, Vol. 26, 3, pp. 579-593.

**Bergh, M. V. den, et al. 2012.** SEEDS: superpixels extracted via energy-driven sampling. *ECCV.* 2012, pp. 13-26.

**Beucher, S. and Meyer, F. 1993.** The morphological approach to segmentation: the watershed transformation in Mathematical Morphology in Image Processing. [ed.] E. R. Dougherty. s.l. : Marcel Dekker Inc., 1993, pp. 433-481.

**Bishop, C.M. 1995.** *Neural Networks for Pattern Recognition.* Oxford University Press: New York, NY, USA : s.n., 1995.

**Blaschke, T. 2010.** Object based image analysis for remote sensing. 2010, Vol. 65, 1, pp. 2-16.

**Blaschke, T., et al. 2014.** Geographic object-based image analysis:A new paradigm in remote sensing and geographic information science. *International Journal of Photogrammetry.* 87, 2014, Vol. 1, pp. 180-191.

**Borji, A., et al. 2014.** Salient object detection: A survey. *arXiv preprint arXiv:1411.5778.* 2014.

**Breiman, L. 1996.** Bagging predictors. *Mach. Learn.* 24, 1996, pp. 123-140.

**Breiman, L. 2001.** Random Forests. *Machine Learning.* 2001, Vol. 45, pp. 5-32.

**Breiman, L., et al. 1984.** Classification and Regression Trees. *Wadsworth Press: Monterey, CA, USA.* 1984.

**Carleer, A. P., Debeir, O. and wolff, E. 2005.** Assessment of very high spatial resolution satelllite image segmentations. *Photogrammetric Engineering & Remote Sensing.* 71, 2005, Vol. 11, pp. 1285-1294.

**Carreira, J. and C. Sminchisescu. 2012.** Cpmc: Automatic object segmentation using constrained parametric min-cuts,. *IEEE Trans. Pattern Analysis and Machine Intelligence.* 2012, Vol. 34, 7, pp. 1312-1328.

**Castilla, G, et al. 2007.** Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. [book auth.] G Castilla and GJ Hay. *Image-objects and geographic-objects.* Berlin : Springer-Verlag., 2007.

**Chan, T. F. and L. A. Vese. 2001.** Active contours without edges. *IEEE Transactions and Image Processing.* 2001, Vol. 10, 2, pp. 266-277.

**Chang, C.C. and Lin, C.J. 2012.** LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 2012, pp. 1-27.

**Chen, L. C., et al. 2014.** Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062.* 2014.

**Chen, Y. and Gong, P. 2013.** Clustering based on eigenspace transformation—CBEST for efficient classification. *ISPRS J. Photogramm. Remote Sens.* 83, 2013, pp. 64-80.

**Clément, V., et al. 1993.** Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture. *IEEE Transactions on Geoscience and Remote Sensing.* 1993, Vol. 31, pp. 779-791.

**Comaniciu, D. and Meer, P. 2002.** Mean Shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2002, Vol. 24, pp. 603-619.

**Congalton, R. G. 1991.** A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment.* 1991, Vol. 37, pp. 35-46.

**Contreras, Jhonatan, Amaya, Iván and Correa, Rodrigo. 2014.** An improved variant of the conventional harmony search algorithm. *Applied Mathematics and Computation.* 2014, Vol. 227, pp. 821-830.

**Cramer, M. 2010.** The DGPF-Test on Digital Airborne Camera Evaluation. *Photogrammetrie-Fernerkundung-.* 2010, pp. 77-82.

**Csurka, G. and Perronnin, F. 2011.** An Efficient Approach to Semantic Segmentation. *Int. Journal on Computer Vision.* April 2011, Vol. 95, 2, pp. 198-212.

**Cuevas, E., Zaldivar, D. and Perez-Cisneros, M. 2011.** Segmentation with learning automata, image segmentation. *P. -G. Ho (Ed.) Intech.* 2011, pp. 83-98.

**Dey, V., Zhang, Y. and Zhong, M. 2010.** *A review on image segmentation techniques with remote sensing perspective.* Vienam, Austria : ISPRS, 2010.

**Diaz, P. 2014.** A Comparison of Segmentation Algorithms for Remote Sensing. Rio de Janeiro : s.n., 2014.

**Drăgut, Lucian, Tiede, Dirk y Levick., Shaun R. 2010.** ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science.* 2010, Vol. 24, 6, págs. 859-871.

**Endres, I. and D. Hoiem. 2000.** Category-independent object proposals with diverse ranking. *IEEE Trans. Pattern Analysis and Machine Intelligence.* 2000, Vol. 22, 8, pp. 888-905.

**Feitosa, R. Q., et al. 2006.** A genetic approach for the automatic adaptation of segmentation parameters. *1st International Conference on Object Based Image Analysis.* May 2006.

**Felzenszwalb, P. F. and Huttenlocher, D. P. 2004.** Efficient graph-based image segmentation. 2004, Vol. 59, 2, pp. 167-181.

**Felzenszwalb, Pedro and Huttenlocher, Daniel. 2004.** Efficient graph-based image segmentation. *International Journal of Computer Vision (IJCV).* 59, 2004, Vol. 2, 167-181.

**Fourier, C. and Shoepfer, E. 2014.** Data Transformation Functions for Expanded Search Space in Geographic Sample Supervised Segment Generation. 2014, Vol. 6, pp. 3791-3821.

**Frey, Brendan J and MacKay, David J.C. 1998.** A revolution: Belief propagation in graphs with cycles. *Advances in neural information processing systems.* 1998, pp. 479-485.

**Friedman, J.H. 2002.** Stochastic gradient boosting. *Comput. Stat. Data Anal.* 38, 2002, pp. 367-378.

**Fung, T., y LeDrew, E. 1988.** The determination of optimal threshold levels. *Photogrammetric Engineering and Remote Sensing.* 1988, Vol. 54, 1449-1454.

**Gao, Y, et al. 2011.** Optimal region growing segmentation and its effect on classification accuracy. *Int.J .Remote sens.* 2011, Vol. 32, 13, pp. 3747-3763.

**Geem, Zong Woo, Kim, Joong Hoon y Loganathan, G. V. 2001.** A new heuristic optimization algorithm: harmony search. *Simulation.* 76, 2001, Vol. 2, págs. 60-68.

**Gondra, I. and Xu, T. 2010.** A multiple instance learning based framework for semantic image segmentation. 2010, Vol. 48, 2, pp. 339-365.

**Gonzales, R.G. and Woods, R. E. 2008.** *Digital Image Processing.* 3rd. s.l. : Prentice Hall, 2008.

**Gould, S., Fulton, R. and Koller, D. 2009.** Decomposing a scene into geometric and semantically consistent regions,. *ICCV.* 2009.

**Grady, L. 2006.** Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2006, Vol. 28, 11, pp. 1768-1783.

**Gu, C., et al. 2000.** Recognition using regions. *IEEE Trans Pattern Anal. Mach. Intell.* 2000, Vol. 22, 8, pp. 888-905.

**Gupta, L., and Sortrakul, T. 1998.** A Gaussian-mixture-based image segmentation algorithm. *Pattern Recognition.* 1998, Vol. 31, 3, pp. 315-325.

**Happ, P., et al. 2013.** A Region-Growing Segmentation Algorithm for GPUs. *IEEE Geoscience and Remote Sensing Letters.* November 2013, Vol. 10, 6, pp. 1612-1616.

**Haralick, R. and Shapiro, L. G. 1985.** Image Segmentation Techniques. *CVGIP*. 29, 1985, pp. 100-132.

**Hay, G.J and Castilla, G. 2006.** Object-Based Image Analysis: Strengths, Weaknesses,Opportunities and Threats (SWOT). *International Archives of Photogrammetry,Remote Sensing and Spatial Information Sciences.* 2006, Vols. Vol. No. XXXVI-4/C42.

**Hay, G.J. and Castilla, G. 2008.** Geographic Object-Based Image Analysis (GEOBIA): a new name for a new discipline. s.l. : Springer Berlin Heidelberg, 2008, pp. 75-89.

**He, Jia, Kim, Chang-Su and Kuo., C-C. Jay. 2013.** Interactive Segmentation Techniques: Algorithms and Performance Evaluation. *Springer Science & Business Media.* 2013.

**Horowitz, Steven L. and Pavlidis, Theodosios. 1974.** Picture segmentation by a directed split-and-merge procedure. *Proceedings of the Second International Joint Conference on Pattern Recognition.* 1974, Vol. 424.

**Humayun, A., Li, F. and Rehg., J. M. 2014.** RIGOR: reusing inference in graph cuts for generating object regions. *CVPR.* 2014, pp. 336-343.

**Im, J., et al. 2009.** Hyperspectral remote sensing analysis of short rotation woody crops grown with controlled nutrient and irrigation treatments. *Geocarto International.* 2009, Vol. 24, 4, pp. 293-312.

**Kass, M., A, Witkin and Terzopoulos, D. 1988.** Snakes: Active contour models. *International Journal of Computer Vision.* 1988, Vol. 1, 4, pp. 321-331.

**Kohli, P., Ladický, L. and Torr, P. H. S. 2009.** Robust higher order potentials for enforcing label consistency. *IJCV.* 1, 2009, Vol. 82, pp. 302-324.

**Koller, Daphne and Friedman, Nir. 2009.** *Probabilistic graphical models: principles and techniques.* s.l. : MIT press, 2009.

**Krähenbühl, P. and Koltun, V. 2012.** Efficient inference in fully connected crfs with gaussian edge potentials. *arXiv preprint arXiv:1210.5644.* 2012.

**Krähenbühl, P. and Koltun, V. 2014.** Geodesic object proposals. *Computer Vision ECCV.* 2014, pp. 725-739.

**Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2012.** Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems.* 2012, pp. 1097-1105.

**Kumar, N., et al. 2012.** Leafsnap: A Computer Vision System for Automatic Plant Species Identification. *Computer Vision–ECCV 2012.* 2012, pp. 502-516.

**Ladicky, L, et al. 2009.** Associative hierarchical crfs for object class image segmentation. *ICCV.* 2009.

**Lafferty, John, McCallum, Andrew and Pereira, Fernando CN. 2001.** Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

**Larlus, D. and Jurie., F. 2008.** Combining appearance models and markov random fields for category level object segmentation. *CVPR.* 2008.

**Lawrence, S., et al. 1997.** Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions.* 1, 1997, Vol. 8, pp. 98-113.

**Le Cessie, S. and van Houwelingen, J.C. 1992.** Ridge estimators in logistic regression. *Appl. Stat.* 41, 1992, pp. 191-201.

**Levinshtein, A., et al. 2009.** Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal.* 2009, Vol. 31, 12, pp. 2290-2297.

**Li, C, et al. 2014.** Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sensing.* 24, 2014, Vol. 6, 2, pp. 964-983.

**Li, F., Peng, J. and Zheng, X. 2004.** Object-based and Semantic Image Segmentation using MRF. 2004, Vol. 3, pp. 833-840.

**Liedtke, C.E., et al. 1997.** AIDA: a system for the knowledge based interpretation of remote sensing data. Third Intern. *Airborne Remote Sensing Conference.* July 7-17, 1997.

**Lillesand, Thomas, Kiefer, Ralph W. and Chipman, Jonathan. 2004.** *Remote sensing and image interpretation.* s.l. : John Wiley & Sons, 2004.

**Liu, C, Frazier, P and Kumar, L. 2007.** Comparative assessment of the measures of thematic classification acuracy. *Remote Sensing of En- vironment.* 2007, pp. 606-616.

**Liu, Desheng and Xia, Fan. 2010.** Assessing object-based classification: advantages and limitations. *Remote Sensing Letters.* 1, 2010, Vol. 4, pp. 187-194.

**Liu, Fayao, Lin, Guosheng and Shen, Chunhua. 2015.** CRF learning with CNN features for image segmentation. *Pattern Recognition.* 48, 2015, Vol. 10, pp. 2983-2992.

**Liu, M., et al. 2014.** Entropy-rate clustering: Cluster analysis via maximizing a submodular function subject to a matroid constraint. *IEEE Trans. Pattern Anal. Mach. Intell.* 2014, Vol. 36, 1.

**Loh, W.Y. and Shih, Y.S. 1997.** Split selection methods for classification trees. *Stat. Sin.* 7, 1997, pp. 815-840.

**Long, J., Shelhamer, E and Darrell, T.,. 2015.** Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2015, pp. 3431-3440.

**Lübker, T. and Schaab., G. 2009.** Optimization of parameter settings for multilevel image segmentation in GEOBIA. *Proceedings of the 2009 ISPRS Hannover Workshop High-Resolution Earth Imaging for Geospatial Information.* 2009.

**Manen, S., Guillaumin, M. and L. J. V. Gool. 2013.** Prime object proposals with randomized prim's algorithm. *ICCV.* 2013.

**Matsuyama, T. and Hwang, V. 1990.** SIGMA: A knowledge-based aerial image understanding system. *Plenum Publishing Corporation.* 1990.

**Matsuyama, Takashi. 1987.** Knowledge-based aerial image understanding systems and expert systems for image processing. *Geoscience and Remote Sensing, IEEE Transactions.* 1987, Vol. 3, pp. 305-316.

**McKeown, D. M., Harvey, W. A. and McDermott, J. 1985.** Rule-based interpretation of aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1985, Vol. 7, pp. 570-585.

**Moran, Emilio Federico. 2010.** Land cover classification in a complex urban-rural landscape with QuickBird imagery. *Photogrammetric engineering and remote sensing.* 2010, Vol. 76, 6.

**Mortensen, E. N. and W. A. Barrett. 1995.** Intelligent scissors for image composition. *SIGGRAPH.* 1995, pp. 191-198.

**Mortensen, E., et al. 1992.** Adaptive boundary detection using ‘live-wire’ two-dimensional dynamic programming. *Computers in Cardiology.* 1992, pp. 635-638.

**Mostajabi, Mohammadreza, Yadollahpour, Payman and Shakhnarovich, Gregory. 2015.** Feedforward semantic segmentation with zoom-out features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2015, pp. 3376-3385.

**Mumford, David y Shah, Jayant. 1989.** Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics.* 1989, Vol. XLII, 5, págs. 577-685.

**Myint, S.W., et al. 2011.** Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment.* 15, 2011, Vol. 5, pp. 1145-1161.

**Nelder, J. A. and Mead, R. 1965.** A Simplex Method for Function Minimization. 1965, Vol. 7, 4, pp. 308-313.

**Neubert, M., Herold, H. and Meinel, G. 2008.** Assessment of Remote Sensing Image Segmentation Quality. *International Archives of photogrammetry, Remote Sensing and Spatial Information Sciences.* August 6-7, 2008, Vol. XXXVIII, 4/C1, p. 5.

**Noh, Hyeonwoo, Seunghoon, Hong and Bohyung, Han. 2015.** Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision.* 2015, pp. 1520-1528.

**—. 2015.** Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision.* 2015, pp. 1520-1528.

**Otsu, N. 1979.** A threshold selection method form grey-level histograms. *IEEE Transactions on Systems Man and Cybernetics.* 9, 1979, Vol. 1, pp. 62-66.

**Pedrini, H. and Schwartz, W. 2008.** Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações. *Thomson Learning.* 2008.

**Petersen, M. E., De Ridder, D. and Handels, H. 2002.** Image processing with neural networks - A review. 2002, Vol. 35, 10, pp. 2279-2301.

**Pignalberi, G., et al. 2003.** Tuning Range Image Segmentation by Genetic algorithm. 2003, pp. 780-790.

**Pinheiro, P.H. and Collobert, R. 2014.** Recurrent convolutional neural networks for scene labeling. *ICML*. 2014.

**Pinho, C.M.D, et al. 2012.** Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis. *International Journal of Remote Sensing.* 2012, Vol. 33, 19, pp. 5973-5995.

**Platt, Rutherford V. and Rapoza, Lauren. 2008.** An Evaluation of an Object-Oriented Paradigm for Land Use/Land Cover Classification. *The Professional Geographer.* 2008, Vol. 60, 1, pp. 87-100.

**Quinlan, R. 1993.** *C4.5: Programs for Machine Learning.* CA, USA : Morgan Kaufmann Publishers: San Mateo, 1993.

**Ren, X. and Malik, J. 2003.** Learning a classification model for segmentation. *ICCV.* 2003, pp. 10-17.

**Rocha, C. H. B. 2007.** *Geoprocessamento: Tecnologia Transdisciplinar.* s.l. : UFJV, 2007.

**Rodriguez, Juan José, Kuncheva, Ludmila I. and Alonso, Carlos J. 2006.** Rotation forest: A new classifier ensemble method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 10, 2006, Vol. 28, pp. 1619-1630.

**Rother, C., et al. 2004.** interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* 2004, Vol. 23, 3.

**Sande, K. E. A. van de, et al. 2011.** Segmentation as selective search for object recognition. *ICCV.* 2011, pp. 1879-1886.

**Schiewe, J. 2002.** Segmentation of high-resolution remotely sensed data-concepts, applications and problems. *The International Archives of the Prothogrametry, Remote Sensing and Spatial Information Sciences.* 2002, Vol. XXXIV, pp. 358-363.

**Schmidt, M. 2007.** UGM: A Matlab toolbox for probabilistic undirected graphical models. *http://www.cs.ubc.ca/~schmidtm/Software/UGM.html.* [Online] 2007.

**Senthilkumaran, N. and Rajesh, R. 2009.** Edge detection techniques for image segmentation–a survey of soft computing approaches. *International journal of recent trends in engineering.* 2, 2009, Vol. 1.

**Shi, J and Malik, J. 1997.** Normalized cuts and image segmentation. *CVPR.* 1997.

**Shotton, J., Winn, J., Rother, C. and Criminisi, A. 2006.** Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. [ed.] Springer Berlin Heidelberg. *Computer Vision–ECCV.* Jan 1, 2006, pp. 1-15.

**Simonyan, Karen and Zisserman, Andrew. 2014.** Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.* 2014.

**Smith, Geoffrey M. y Morton, R. Daniel. 2010.** Real World Objects in GEOBIA through the Exploitation of Existing Digital Cartography and Image Segmentation. *Photogrammetric Engineering & Remote Sensing* . 2010, Vol. 76, 2.

**Story, Michael and CONGALTON, Russell G. 1986.** Accuracy assessment-A user\'s perspective. *Photogrammetric Engineering and remote sensing.* 3, 1986, Vol. 52, pp. 391-399.

**Szeliski, R. 2011.** *Computer Vision - Algorithms and Applications, Texts in Computer Science.* s.l. : Springer, 2011.

**Thoma, Martin. 2016.** A survey of semantic segmentation. *rXiv preprint arXiv.* 2016, Vol. 1602, 06541.

**Tiede, D., S, Lang, F. Albrecht and Hölbling., D. 2010.** Object based class modeling for cadastre constrained delineation of geo-objects. *Photogrammetric Engineering and Remote Sensing.* 2010, Vol. 76, 2, pp. 193-202.

**Tilton, J and Lawrence, W. 2000.** [ed.] IEEE Press n International Geoscience and Remote Sensing Symposium IGARSS-2000. New York : s.n., 2000, pp. 733-773.

**Tilton, J and Lawrence, W. 2000.** Interactive analysis of hierarchical image segmentation. [ed.] IEEE Press n International Geoscience and Remote Sensing

Symposium IGARSS-2000. *International Geoscience and Remote Sensing Symposium IGARSS-2000, IEEE Press.* 2000, pp. 733-773.

**Ulusoy, Ilkay and Christopher, M. Bishop. 2005.** Generative versus discriminative methods for object recognition. IEEE Computer Society Conference on. *Computer Vision and Pattern Recognition.* 2005, Vol. 02, pp. 258-265.

**Van Rijsbergen, C. 1979.** *Information Retrieval.* 2nd. s.l. : Dept. of Computer Science, Univ. of Glasgow, 1979.

**Vantaram, S. R. and Saber, E. 2012.** Survey of contemporary trends in color image segmentation. *Journal of Electronic Imaging.* 2012, Vol. 21, 4, pp. 040901-1-040901-28.

**Vapnik, Vladimir Naumovich and Vlamimir, Vapnik. 1998.** *Statistical learning theory.* New York : Wiley, 1998.

**Vedaldi, Andrea and Soatto, Soatto. 2008.** Quick shift and kernel methods for mode seeking. *European Conference on Computer Vision (ECCV).* 2008.

**Veksler, O., Boykov, Y. and P. Mehrani. 2010.** Superpixels and supervoxels in an energy optimization framework. *Computer Vision - ECCV 2010 - 11th European Conference on Computer Vision.* september 5-11, 2010, pp. 211-224.

**Vieira, M.A., et al. 2012.** Object based image analysis and data mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sensing of Environment.* 2012, Vol. 123, pp. 553-562.

**Visa, A., Valkealahti, K. and Simula, O. 1991.** Cloud detection based on texture segmentation by neural network methods. 1991, Vol. 2, pp. 1001-1006.

**Vishwanathan, S., et al. 2006.** Accelerated training of conditional random fields with stochastic gradient methods. *Proc. of the 23rd international conference on Machine learning.* 2006, pp. 969-976.

**Wang, P., et al. 2013.** Structure-sensitive superpixels via geodesic distance. *International Journal of Computer Vision.* 2013, Vol. 103, 1, pp. 1-21.

**Weszka, J. S. and Rosenfeld, A. 1979.** Histogram modification for threshold selection. *IEEE Transaction on Systems Man and Cybernetics 9.* 38, 1979, Vol. 52.

**Whittaker, Joe. 1990.** *Graphical models in applied multivariate statistics.* s.l. : Wiley Publishing., 1990.

**Yang, W., et al. 2010.** Semantic Segmentation of Polarimetric SAR Imagery using Conditional Random Fields. 2010, pp. 1593-1596.

**Yi, Faliu and Moon, Inkyu. 2012.** Image segmentation: A survey of graph-cut methods. *Systems and Informatics (ICSAI).* 2012, pp. 1936-1941.

**Zeiler, M.D., et al. 2010.** Deconvolutional networks. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference.* 2010, pp. 2528-2535.

**Zhang, Chenxi, Wang, L. and Yang, R. 2010.** Semantic segmentation of urban scenes using dense depth maps. *In Computer Vision–ECCV 2010.* 2010, pp. 708-721.

**Zhang, J. 2010.** Edge Detection in Glass Fragmentation Images Based on One Order Differential Operator. *Computer Engineering and Applications (ICCEA).* Second International Conference, March 19-21, 2010, Vol. 2, pp. 591-594.

**Zhang, Y. 2001.** A review of recent evaluation methods for image segmentation. *Int. Symp. on Signal Processing and its Applications (ISSPA).* 2001, pp. 148-151.

**—. 1996.** A survey on evaluation methods for image segmentation. *Pattern Recognition.* 1996, Vol. 29, 8, pp. 1335-1346.

**Zhang, Yimeng and Chen, Tsuhan. 2012.** Efficient inference for fully-connected crfs with stationarity. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE.* 2012.

**Zhu, Hongyuan, et al. 2016.** Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *Journal of Visual Communication and Image Representation.* 34, 2016, p. Journal of Visual Communication and Image Representation.