Graph cuts for fast optimization in Markov Random Field based Remote Sensing image analysis

Abdulmain Karimov

March, 2010

Graph cuts for fast optimization in Markov Random Field based Remote Sensing image analysis

by

Abdulmain Karimov

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Geoinformatics

Thesis Assessment Board

Chair:	Ms Dr. Ir. W. Bijker
External examiner:	Dr. Ir. B.G.H. Gorte
Supervisor:	Dr. V.A. Tolpekin
Second supervisor:	Prof. Dr. A. Stein

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Abstract

Optimization of energy is a challenging issue in Markov Random Field (MRF) based remote sensing image analysis. Traditional energy minimization methods such as Iterated Conditional Modes (ICM) and Simulated Annealing (SA) are widely used to deal with this problem. ICM does not provide a globally best estimation but it concentrates on local area and it does not provide optimal solution. SA is used to find globally optimal solutions for MRF based image analysis problems which allow approximating the global minimum of energy function that produces better quality of solutions, but at it takes long computational time to approximate the global minimum. Therefore, in order to address this problem faster energy minimization methods from vision are proposed. The applicability of graph based methods such as swap-move and expansion-move algorithms in MRF based remotely sensed images have been proposed and studied in this research.

A number of methods are proposed that address multi label (class) image classification problem, which make use of class separability measures. Based on these measures classification trees are constructed. Each level of tree nodes represents a particular class. These methods show how the sequence (order) of tree nodes can have an impact on classification result. The most appropriate method among them is selected that best represents the reality. In addition, a method is proposed that used to optimize the smoothness values based on "trial and error" method. Its main issue is to identify the optimal smoothness value, i.e. the value that is most suitable for specific spectral classes. The results show that these smoothness values are more sensitive to spectral classes that are least separated, and less sensitive to those that most separated. The "least" and "most" separated issues are based on divergence - class separability measures.

The results of swap-move and expansion-move algorithms are compared with MLC, ICM and SA. Different smoothness parameters are chosen and tested for each energy minimization method. The results of swap-move and expansion-move algorithms are similar to SA annealing with logarithmic cooling schedule, which is considered able to approximate the global optimal solution. In terms of computation, these algorithms outperformed the SA algorithms and in some cases ICM also. In comparison with MLC, the swap-move and expansion-move algorithms produce better results.

In conclusion, swap-move and expansion-move algorithms show that they are applicable in MRF based remote sensing image classification. Their performances, both in terms of computation and classification accuracy are considered sufficient to address the research problems. Although in some cases they are not able to fully detect trees, mainly due to similarity in spectral properties of trees and grassland classes. In addition, the proposed methods for construction of classification tree and optimising smoothness value showed their significant importance in addressing a multi label image classification problem.

Key words: *Markov Random Field, local optimization, global optimization, graph cuts, swap-move, expansion-move, performance evaluation, computational time, Simulated Annealing (SA), logarithmic cooling schedule, exponential cooling schedule, classification tree, tree nodes, optimizing of smoothness value, least separated class, most separated class .*

Acknowledgements

I acknowledge and express my special gratitude to my supervisor Dr. Valentyn Tolpekin whose scholarly advice, help and continuous encouragement have contributed significantly to the completion of this study. I want to thank him for providing me with constructive critics which has immensely impacted the result of this thesis.

A special thanks to Professor Dr. Alfred Stein for his comments, recommendations and advices that helped me a lot to overcome some academic challenges while completing this study. Thank you so much!

Also, my extensive gratitude in none particular order to Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Yuri Boykov and other researchers and scholars for their incredible works in developing and improving the graph based energy minimization methods and also for their generosities for making the codes available online for other researchers. Without their academic contributions as a mean of source, this research would not have been possible to carry out.

A big thank to Juan Pablo Ardila Lopez, a PhD student at Earth Observation Department of ITC faculty, University of Twente for providing me with satellite image and the reference data for my study area. Also his advises and supports on image analysis related issues were very useful in the research process.

Moreover, I am indebted to the course director, lecturers and all my fellow students for their invaluable input and for being a great source of support to me during my study. I am grateful to my new friends for creating friendly academic environment while sharing the lab which has helped me to complete this research project.

I really appreciate the help of the ITC IT department for installing the required software and helping me in obtaining the satellite images for my research. Also, I am appreciative of the services of ITC library staff for making available all the high quality information, including the recommended literatures and other relevant text books during my studies and thesis.

The Huygens Scholarship Programme made possible my study in the Netherlands, which has assisted me to improve my knowledge skills on Geoinformatics. At the same time it was a great opportunity to learn and experience the diverse culture and traditions of Dutch people. Therefore, I express my deep gratitude for the member of this Scholarship Programme for their generosities and kind supports which has played a vital role in both my studying and research project.

Finally, I want to say my special thanks for my family members for their love, supports and affection during my studies in the Netherlands. This work would not have happened without their advices, courage and supports.

Thank you all.

Table of contents

List of figures

List of tables

1. Introduction

1.1. Background and Problem statement

Many organizations such as national mapping agencies, cartographic institutions, and geological survey agencies are interested in extracting land cover information from remote sensing images that can be used to plan and manage the environment both at local and global scales. Number of image processing and analysis techniques proposed that help to extract information form these images. In hard classification approaches each pixel is associated with a single land cover class. Using these techniques the classification results may still contain uncertainty, because of the occurrence of mixed pixels in coarse spatial resolution images. Soft classification techniques address these problems and their results are more appropriate to represent the land cover than the one produced by hard classification. But even with soft classification techniques it is still difficult to represent the reality. More promising approaches for remote sensing image analysis are developed which not only considering spectral signatures but also contextual information. The spatial context is very important for the visual interpretation of images. The suitable use of context allows the elimination of possible ambiguities, the recovery of missing information and correction of the errors. Using the concept of the context, pixels are considered to have a relationship with their neighbours. This relationship is treated as being statistically dependent.

Markov Random Fields (MRF) are considered as relevant contextual image analysis tools. The use of both spectral and contextual information produces better results for image analysis. Contextual information can be defined as how the probability of presence of one object (or objects) is affected by its (their) neighbours. During the process of analysis they are able to integrate the use of contextual information associated with the image data.

One of the important tasks in MRF based image analysis methods is a minimization of an energy function that lead to better quality of solutions. Simulated annealing for energy minimization in MRF based image analysis is considered as an appropriate approach to deal with this problem. SA was proposed first by Kirkpatrick [1]. "It is a stochastic algorithm used for combinatorial optimization and it simulates the physical annealing procedure in which a physical substance is melted and then slowly cooled down in search of a low energy configuration" [2]. For simplicity, throughout this proposal I refer to Simulated annealing for energy minimization in MRF based image classification as just SA. There are no analytical methods that are able to find globally optimal solutions for MRF based image analysis problems, except for very simple ones. The main reason for that is the huge number of variables, i.e pixels in the images. Simulated annealing allows approximating the global minimum of energy function that lead to better quality of solutions. There are some advantages and disadvantages of using SA approach. The advantages are following: simulated annealing is an intelligent random search method and it can deal with many constrains, nonlinear models and noisy data. Also it is robust, flexible and has an ability to approach a global optimality [2]. The main disadvantages of this approach is that it requires long computations; in order of tens of thousands of iterations (depending

on image analysis problem) are required to approximate the global minimum [3]. Computation speed of SA is proportional to the number of pixels in the image.

In the field of computer vision, several algorithms have been developed and proposed that are much faster than simulated annealing. Graph cuts algorithms are among those algorithms. Graph cuts are considered as efficient energy minimization algorithms in computer vision applications [4]. These algorithms are widely used in image restoration, segmentation, voxel occupancy and stereo [5]. Boykov and Kolmogorov made a significant contribution in improving and developing new graph cut algorithms [5]. They studied and made an experimental comparison of min-cut/max-flow algorithms for energy minimization in computer vision. The implementation part is available upon request for research purposes.

The graph cuts methods pose conditions on energy function. They can be applied in nominal resolution classification approach in remote sensing images, because they have been applied graph cuts on a problem with the similar prior energy function [5]. They use α - expansion-move algorithm which allows updating the pixel value not in pixel by pixel based but for a group of pixels at once. It is one of the main reasons that lead to faster performance than in SA approach. Perhaps this is an opportunity to obtain the segments, group of spatial adjacent pixels from the same land cover class, which means that the classification results can be represented as objects. Objects are usually more informative than pixels for the users. This gives sufficient motivation for exploring and investigating the applicability of these methods in remote sensing image analysis, which can be considered and addressed in the proposed research problem.

1.2. Research identification

A number of studies have been carried out in computer vision, where several energy minimization algorithms such as graph cuts are developed to solve the image analysis problems. Until now these algorithms did not receive sufficient attention in MRF based remote sensing image analysis literature.

1.2.1. Research objectives

The overall objective of this research is to explore and study the graph cuts based energy minimization algorithms in MRF based remote sensing image classification and compare the results with existing energy minimization algorithms such as Simulated Annealing (SA) and Iterated Conditional Modes (ICM).

The following are the specific research objectives:

- to explore and analyze the applicability of graph cuts algorithms in MRF based Remote Sensing image classification.

- to identify the most appropriate graph based algorithms that can be applied in MRF based Remote Sensing image classification.

- to evaluate the performance of the applied algorithms and compare the results with ICM and SA.

1.2.2. Research questions

- What does the applicability means in terms of energy minimization in MRF based remote sensing image analysis?
- Graph cuts algorithms are mainly applied on regular colour photographs (RGB) and medical images. Are the algorithms applicable for multiband remote sensing images?
- Graph cuts algorithms are widely used to solve binary image segmentation problems. How to address multi label (class) MRF based remote sensing image classification problem?
- SA produces sufficient enough classification accuracy. Do the graph cuts algorithms achieve similar to SA classification accuracy?
- ICM is considered fast energy minimization method. What is the performance (in term of computation time) of the applied algorithms compared to ICM?

1.2.3. Innovation aimed at

The research is entirely innovative since it aims at improving the performance of remotely sensed image classification by applying faster energy minimization algorithms from vision.

1.2.4. Related work

There are no directly related works have been done in regards to this research problem. Although several algorithms have been proposed and tested in computer vision applications that have some similarity with remote sensing image classifications. The details on these algorithms are provided in the literature review chapter.

1.3.   Project setup

In this section, the new approach is proposed to address the research problem.

1.3.1. Method adopted

 Following is the proposed method that can be considered during the thesis:

- **Exploring and investigating the algorithms**: The first step is to study, explore and investigate the existing energy minimization algorithms both in MRF based remote sensing image classification and computer vision applications. Understanding the mathematical background of both fields is one of the important tasks at this stage.

- **Applicability of graph cuts algorithms in MRF based remote sensing image classification and identifying the most appropriate algorithms**: In order to find out the applicability of the graph cuts algorithms for MRF based remote sensing image analysis, it is important to analyse the energy minimization functions, because graph cuts approaches are not applicable for all energy functions [6]. Next step is to identify the most appropriate algorithms that consider those energy functions which are applicable in MRF based remote sensing image analysis. Modification or improvement of graph cuts algorithms will be considered if necessary.

- **Optimization of smoothness parameter.** Optimal smoothness parameter value is very sensitive to characteristics of land cover classes. At this stage the most appropriate smoothness parameter value for each individual land cover class will be defined and tested. The "trial and error" method is considered in order to obtain an optimal smoothness parameter for each land cover class.

- **Construction of classification tree.** Since the graph cuts algorithms are used to solve binary image segmentation (object of interest and background) problem, in this thesis a method is proposed to apply sequences of binary classification that addresses a multi label classification problem.

- **Remote sensing dataset**: at this stage remote sensing dataset will be selected for a specific area where the algorithm will be applied and tested. Reference data for that area will be needed for validation purposes. The information on dataset and reference data requirements is described in sec 4.1.

- **Applying and testing the algorithms**: The graph based energy minimization algorithms as well as the SA and ICM algorithms are applied on MRF based remote sensing images.

- **Performance evaluation**: After applying the identified energy minimization algorithms on the remote sensing images, the results will be assessed. The aim of this research is not only to assess the performance (in term of computation time) but also the quality of classification (in terms of energy minimization) will be assessed.

- **Experimental comparison**: In the last stage the classification results of all algorithms will be compared.

1.4. Structure of the thesis

The first chapter describes the background, problem statement, research identification and adopted method. The second chapter discusses about the MRF based image analysis both in remote sensing environment and computer vision. It provides number of previous related works on energy minimization problem in both fields. Several related works on TS-MRF models are described that used to improve the results of remotely sensed image classification. Also some energy minimization functions that are used for optimization problems are discussed. Chapter 3 provides detail information about the existing energy minimizations methods in both fields including the mathematical background that can be used for this research. Chapter 4 provides detail information study area and adopted method, including the implementation part. Chapter 5 provides the results. Charter 6 discuss about the results and provides comparison analysis of the applied methods. The last chapter concludes and provides recommendations including future work.

2.   Literature Review

This chapter reviews the previous works related to energy minimization problems both in vision and remote sensing fields. The first part of the chapter describes energy minimization methods to solve labelling problem in image analysis. A number of related works on energy minimization problem for remotely sensed image analysis based on MRF models is described for different application such as image segmentation and contextual image classification. Two widely used energy minimization algorithms such as Simulated Annealing (SA) and Iterated Conditional Modes (ICM) are discussed including their comparison analysis. The second part of this chapter provides some related works on energy minimization methods in computer vision. Several graph based algorithms, such as min-cut/max-flow, swap-move and expansion-move are discussed including the group of energy functions that are applicably via graph cuts. Finally, a summary on chapter is provided.

2.1.   Markov Random Fields (MRF) as a contextual image analysis tool

The contextual constrains are very important for the interpretation of images. They can be derived from spectral, spatial and temporal properties [7]. Markov Random Fields theory provides a consistent approach for modelling context dependent entities such as image pixels and correlated features. The suitable use of context allows the elimination of possible ambiguities, the recovery of missing information and correction of errors.

MRF is widely used in remote sensing image processing and analysis such as image segmentation and classification, image restoration and reconstruction problems. MRF is considered as a useful tool for characterizing contextual information [8], [9]. Using the concept of the context, pixels are considered to have a relationship with their neighbours. This relationship is treated as being statistically dependent [7].

Number of studies has been carried out with MRF models for image analysis which are discussed in the following subsections.

2.1.1.   MRF based models for Remotely Sensed image analysis

Many researchers used MRF based model to address different issues related to remotely sensed image analysis, those include texture analysis, image segmentation and classification, de-noising, pattern recognition and other related problems. Different methods are proposed in the literature that addresses such problems. This section describes number works related to MRF based models for remotely sensed image analyzes.

 Solberg *et al* [10] proposed a general model based on MRF for multisource classification of remotely sensed data where they used and tested a model for synthetic aperture radar (SAR) images, fusion of optical images and GIS ground cover data. MRF model was used which can exploit a spatial and temporal class dependency properties of the same scene. The temporal aspect of the data was included that allowed the proposed model to the class changes between acquisition dates of different images. The proposed method was applied on fusing Landsat TM and multitemporal ERS-1 SAR images and

in addition GIS ground cover maps were used for land use classification. The proposed method showed that MRF model achieved a high accuracy results.

Melgani and Serpico [11] considered an approach to spatio-temporal contextual image classification based on MRF model which allows improving the classification accuracy as well as the reliability during the classification process by better exploration and using of temporal information. The method introduced by the authors is based on the concept of "minimum perturbation" which was implemented with pseudo inverse technique of minimization of the sum of squared errors. Multi-temporal dataset (Landsat TM and SAR) was used and the experimental results showed an improvement in terms of classification accuracy using the MRF based model.

Tso and Olsen [12] proposed a MRF based method which make use of both contextual information and multiscale fuzzy line process for classification of remotely sensed imagery. IKONOS panchromatic and multispectral data is used to study the applicability of the method. The results show that the proposed method, based on the MRF model with the multiscale fuzzy line process, successfully generated the patch-wise classification patterns, and improved both the accuracy and visual interpretation.

Amador [13] described a region extraction algorithm based on the MRF concept. A heuristically developed energy functional is presented and used with the MRF in an efficient and accurate manner. Since the MRF used in this work is defined using the polar coordinate system, a very large search space exists for radial lengths and sites. To aid in pursuing these radial sites, a combinatorial optimization technique known as Tabu Search is exploited. Also provided is an extensive empirical study on aerial imagery and parts detection.

2.1.2. Markov Random Fields for image segmentation

MRF models are often used to solve image segmentation problem. Number of existing MRF based image segmentation methods are discussed in this section.

Kim *et al.* [14] provided an unsupervised image segmentation method, which makes use of hierarchical distributed genetic algorithm to solve the computationally intensive problem of Markov Random Field based models. Their results show that the method is effective in segmenting the real images.

Arques [15] proposed an approach to the model based on Markov random field (MRF) based image segmentation. The author defined the energy function with robust features and their integration. To check the suitability of this method the author compared the robust features with classic ones. In this method, first an image was segmented into a set of disjoint regions and the adjacent graph. In order to apply the method, Markov Random Field mode was first defined on the corresponding adjacent graph. Simulated annealing was used to minimize the energy function.

Unsupervised segmentation algorithms based on hierarchical MRF models was presented in [16] which allows solving the segmenting problems of both noisy and textured images. These algorithms are able to find the number of classes, their associated model parameters and generate a corresponding segmentation of the image into these classes which is achieved according to the MAP criterion. The experimental results showed improvement in solution with sufficient accuracy.

Hu and Fahmy [17] presented two segmentation techniques, supervised and unsupervised. They used MRF to model the textured images which is formed by combining the binomial model for textures and the multi-level logistic model for region distributions. The proposed algorithm is able reach the global maxima of the posteriori distribution in case of supervised segmentation. In the case of unsupervised segmentation they proposed a new parameter estimation scheme which estimates the model parameters directly from a given image.

Many other researchers used similar MRF based models for image segmentation problems. [18] proposed a multi scale segmentation algorithm for SAR images and they applied a hierarchical two-level MRF to improve the segmentation results. Similarly, [19] proposed a new SAR image segmentation algorithm based on the mixture context and the wavelet hidden-class-label MRF. They studied different methods and made a comparison analysis. Using the new method a new MAP classification was obtained. The experimental results showed that their method outperformed several other segmentation methods. In [20] and [21] MRF based models for remote sensing image segmentation were used for similar problems and the results are provided.

2.1.3.   Tree-Structure Markov Random Fields models

In order to achieve good segmentation results it's not sufficient to rely on observed data only, like in case of non-contextual segmentation algorithms but it's also important to consider the prior information about the image which allows better representing the reality. For this reason Tree-Structured (TS) – MRF segmentation algorithms are important. Aiazzi *et al*. [22] used TS- MRF based on priori model which takes into account special dependencies within image through the conditional probability that pixel belongs to a given class, given the class of its neighbours. The method was used to solve the MAP estimation problem and also to estimate a set of unknown parameters, given the data and the model. The experiments showed that the method was able to extract the content from SAR images with sufficient accuracy.

Similarly, Gaetano *et al*. [23] used a TS-MRF for segmentation of multitemporal remotely sensed images. They proposed a method to build automatically the underlying tree structure of the model based on metric which allows comparing the class features to establish hierarchical relationship among classes.

Another algorithm based on a TS-MRF model that carries out the unsupervised classification of remote sensing images was proposed in [24]. Due to MRF model, it considers spatial dependencies and it achieves fast computation results because only binary MRF's were used. The results provide useful information about the segmentation process.

D'Elia *et al*. [25] presented image segmentation algorithm based on a TS-MRF binary model. Binary segmentation obtained solution to a MAP estimation problem. Because of binary fields and tree structure algorithm were used, the method produced good results and allowed to address the cluster validation problem. A split-and-merge procedure was developed to improve segmentation accuracy in addition spatially adaptive MRF model was used. The method was applied on multispectral images it showed quite good performance both in term of segmentation and map smoothness.

A TS-MRF model for supervised segmentation which uses of prior knowledge on the number of classes and their statistical features was introduced in [26]. This allows generalizing the model so that the binary MRFs associated with the nodes can be adapted freely together with their local parameters, to better fit the data. In addition, it allows defining a suitable likelihood term to be coupled with the TS-MRF prior so as to obtain a precise global model of the image. The method was applied on SPOT images and the results showed that the proposed algorithm achieved better performance with compare to other MRF based algorithms.

2.1.4.   Energy minimization algorithms: Iterated Conditional Modes (ICM)

Maximum-a-Posteriori (MAP) estimation is often applied on image segmentation and one of the main issues is that it leads to the combinatory optimization problem. Therefore, several MAP estimation algorithms are proposed in literature. Among the proposed algorithms Iterated Conditional Modes (ICM) is widely used to deal with this problem. This technique was introduced by Besag (1986) [27]. "This iterative procedure incorporates knowledge about the underlying scene by the choice of a 'neighbourhood system', weight function and smoothing parameter. Broadly speaking this method exploits the tendency of adjacent pixels to have the same colour" [27]. ICM does not provide a globally best estimation but it concentrates on local area.   There are advantages as well as disadvantages of this technique. The main advantage is that computationally it is considered as a fast method and the disadvantage is that its accuracy is concerned. Several studies have been carried out to improve the algorithm.

A generalized version of ICM algorithm for image enhancement was developed which utilizes the characteristic of MRF in modelling the contextual information embedded in image formation [28]. They introduced a new MRF model with a second-order neighbourhood system that allows to extracts contextual information both from the intensity levels and the relative position of neighbouring cliques. In addition to that they introduced an outlier rejection method which depends on each candidate's contribution to the local variance. The method was tested on image restoration problem [28].

Magnussen *et al.* [29] have been studied a "contextual classification of Landsat TM images to forest inventory cover types" where the context of Landsat TM images forest stands are a cluster of homogeneous pixels. ICM algorithm was applied on contextual classification of forest cover types Landsat Images and the classification accuracy was assessed. The results of ICM were compared with Maximum Likelihood where ICM were best which improved the overall accuracy by four to six percentage points (statistically significant).

Many other similar techniques are proposed in [30], [14], [31] and [32] where ICM was used and compared with other energy minimization methods.

2.1.5.   Energy minimization algorithms: Simulated Annealing (SA)

Simulated annealing (SA) was introduced first by Kirkpatrick [1]. "It is a stochastic algorithm used for combinatorial optimization and it simulates the physical annealing procedure in which a physical substance is melted and then slowly cooled down in search of a low energy configuration" [2]. SA is often used to find globally optimal solutions for MRF based image analysis problems. It allows approximating the global minimum of energy function that produces better quality of solutions. But at the same time it takes long time to approximate the global minimum. Number of researches used SA to address the energy minimization problems which are described below.

Modestino and Zhang [33] described an MRF model-based approach to automated image interpretation. They used a region-based approach where an image is first segmented into of disjoint regions that form the nodes of an adjacency graph. Image interpretation was achieved by assigning object labels to the segmented regions using domain knowledge. Using the proposed method the interpretation of labels is modelled as an MRF on the corresponding adjacency graph. Then the image interpretation problem is formulated as a MAP estimation given domain knowledge and region-based measurements. In order to achieve an optimal MAP interpretation Simulated Annealing (SA) was used to for best realization. The performance of this approach was tested by applying the SA algorithm on real-world and synthetic images which are described in the paper.

Ingber in [3] made a comparison analysis of Simulated Annealing (SA) algorithm with Simulated Quenching (SQ). The author used modified faster algorithms Simulated Quenching (SQ) which shows that it's statistically promising to provide an optimal solution. The results showed that SQ is faster then SA without effecting the accuracy.

Lamotte *et al.* [34] presented comparative study of several optimisation algorithms based that are on Simulated Annealing (SA). The authors introduced a new method on random descent. The study included ICM, the Metropolis algorithm and Gibbs Sampler. They considered several criteria's such as the convergence speed, the quality of the optimum obtained with a new energy function and the total computing time necessary for image restoration. They applied the algorithms on the special case of restoration of images disturbed by a Gaussian noise. Their results allow decreasing computation time.

Nasab *et al.* [35] presented modified implementations of SA for image segmentation problems. The segmentation procedure is based on MRF model for describing regions within an image. SA is used as an iterative approach for computing a set of labels with MAP probability which normally take long computation time. They proposed a random cost function (RCF) for computing a posterior energy function in SA. The proposed modified SA method depicts more robust performance for image segmentation than standard SA at the same computational cost. The comparison analyses are provided which describes the performance of the new algorithm both in terms of computational cost and segmentation accuracy.

Many other researches have been carried out to study SA algorithm for energy minimization problems in different applications. Hurn, M and Jennison, A [36] addressed an image reconstruction problem as well as the computational problems arising in determining the MAP estimate. They used ICM and SA algorithms to deal with the problem. The algorithms were applied on SAR image and the results are provided with comparison analysis.

2.2.   Energy minimization via graph cuts in vision

Many problems in vision are formulated in term of energy minimization. Energy minimization is used to solve the pixel labelling problem in application such as image restoration and segmentation, image synthesis, stereo and motion etc.

Y. Boykov and V. Kolmogorov [5] provided an experimental comparison of min-cut and max-flow algorithm for energy minimization in vision. In this paper they have compared the performance of the standard algorithms with the newly developed min-cut max-flow, which works faster then any other methods in vision then.

Still, energy minimization remains the main issue in vision because of a wide range of energy functions. In some cases is possible to compute the exact global minimum. In some other cases Y. Boykov, *et al.* [45] showed how to compute the local minimum in a strong sense that is within a known factor of a global minimum.

V. Kolmogorov and R. Zabih [6] have studied different energy minimization functions and provided a precise characterization of energy functions class that can be minimized by graph cuts. Also they provided a condition for any energy function with binary variables that can be minimized by graph cuts.

2.2.1.　　　Graph cuts for Markov Random Fields (MRF)

Markov Random Fields are used as generative models to solve the labelling problem in image processing in computer vision. They are widely used in modelling low level vision problem such as image restoration, image segmentation, texture analysis, etc. Graph cuts are considered effective approaches for solving the energy minimization problems in vision. Some examples are described below how Graph cuts methods can be applied to address MAP-MRF estimation problems.

Boykov *et.al.* [5] proposed methods to solve MAP-MRF using graph-cut algorithms and they showed that MAP-MRF estimation is equivalent to min-cut problem on a graph. This equivalent makes graph cuts extremely important. They also showed that using max-flow min-cut graph algorithms it is possible to solve some class of energy functions with MAP-MRF framework within a known factor of global minimum.

Lempitsky *et al.* [37] proposed a new approach to the optimization of multi-labelled MRFs. They used the LogCut algorithm which proved itself as an efficient algorithm for multi-label energy minimization. This algorithm was able to reduce the time complexity from linear in the size of label space to logarithmic. For application such as image restoration, optical flow and high resolution stereo they showed that this approach gives fast computational time.

Kohli and Torr [38] presented a new fully dynamic algorithm for the st-mincut problem which can be used to find MAP solutions for certain dynamically changing MRFs rapidly. Their method is generic and finds exact solutions for all dynamic problems which can be formulated as sub-modular energy functions of binary variables. The results showed that their algorithm is substantially faster than the best known static st-mincut algorithm. They have demonstrated how their method can be used to perform efficient image segmentation in video sequences in a manner much faster than previously possible.

2.2.2.　　　Binary image segmentation with Graph cuts

Graph cuts algorithms have been used in a wide range of problems in computer vision. They are often used for medical image segmentation as well as for video segmentation using stereo disparity cues [39]. Boykov and Jolly [40] were the first who proposed and tested a binary graph cuts algorithms for object segmentation.

In [5] several algorithms were tested and applied on binary image segmentation. Boykov and Funka-Lea proposed a technique used for object extraction via graph cuts [4]. This technique can be applied to object of interests in images and also to volumes of any dimension. They showed that it finds a

globally optimal binary segmentation of N-dimensional image under appropriate constrains. The authors tested min-cut max flow algorithms on 2D and 3D segmentation. As results they showed that this technique was able to efficiently re-computes the optimal solution. This approach in addition allows effective editing of segments if necessary.

2.2.3.   Energy minimization algorithms: the min-cut max-flow algorithm

The min-cut max-flow algorithms are considered as an increasingly useful tool for exact or approximate energy minimization in vision [5]. These algorithms are used in [41], [42], [6], and presented with different polynomial time's complexity. Different categories of polynomial algorithms for min-cut max-flow problems on directed weighted graphs with two terminals are described by Boykov and Kolmogorov [5]. These algorithms belong either to "push-relabel" or to "augmenting path". Boykov and Kolmogorov proposed a new min-cut max-flow algorithm [5] and it belongs to the category of "augmenting path". They showed that this algorithm outperformed the already existing category of algorithms.   The details on implementation of the min-cut max-flow algorithms are described in [5].

2.2.4.   Energy minimization algorithms: swap-move and expansion-move algorithms

Boykov et al. [39] introduced the most popular energy minimization algorithms, called swap-move algorithm and expansion-move algorithm. Both algorithms are able to compute the global minimum of the binary labelling problem in vision which allows producing a lower energy labelling. Both algorithms are described in [4]. Boykov and Jolly [45] studied the applicability of these algorithms for some energy functions including their differences which are explained in this paper.

Szeliski et al. [39] made a comparative study of energy minimization algorithms. They have studied several algorithms such as iterated conditional modes (ICM) [27], swap-move and expansion-move algorithm [42], max-product loopy belief propagation (LBP) and finally tree-reweighted message passing (TRW) [43] algorithms. They have created a set of low level energy minimizations problem in different applications, such as stereo matching, photomontage, binary image segmentation, image de-noising and inpainting. The results showed that the modern energy minimization algorithms such as swap-move and expansion-move are better than ICM because they close to compute the global minimum [4]. The results show that in terms of computational time, expansion-move performed the best among the modern energy minimization methods.

2.3.   Summary

The chapter provided sufficient information related to the current research problem. It described the last improvements as well as the challenges of energy minimization problems both in remote sensing and vision fields. The existing energy minimization methods in both fields, their mathematical background, including their applicability in MRF based remote sensing image analysis have been studied. Comparative analysis of existing energy minimization algorithms of both fields have been carried out and the appropriate method was adopted that better address the current research problem. Certain types of energy functions were studied that are applicable for MRF based remote sensing image analysis.

Based on the study, swap-move and expansion-move algorithms are chosen among the graph based algorithms that use to address the current research problem. During the study most attention paid to the type of energy function that is used for graph cuts methods.

In addition, other methods such as Tree Structured - MRF based models are studied that can be combined with the adopted method and use for multi label image classification.

3. Markov Random Fields in contextual image analysis

3.1. Markov Random Fields

There is a strong relationship between the Bayesian statistics and Markov Random Field (MRF) theories, which is described in this chapter.

Bayesian theory has a profound influence on statistical modelling. It consists of two key elements, the prior and conditional probability density function (p.d.f.) and the classification can be expressed in terms of maximum a posteriori (MAP) by using booth elements [7]. But in practice, it is challenging task to use MAP estimates, because the prior information or data distribution information may not always be available. Maximum Likelihood (MLC) criterion can be used in case if the knowledge of data distribution is available but not the prior information about the data. This criterion is widely used in remote sensing image classification. The classification results can be improved by modelling the prior p.d.f. and using the class-conditional p.d.f. to establish MAP estimates [7].

Context is very important in image interpretation which can be derived from spectral, spatial and temporal properties [7] and [10]. Contextual information produces better classification results and it allows reducing the ambiguity and recovering missing information.

Markov Random Field (MRF) considered as a useful framework for characterising the contextual information and widely used to image segmentation and restoration problems [8], [9] and etc. It is also part of probability theory that characterizes the local contextual relationship of physical phenomena [44] and it has been used in statistical physics to describe the interaction between neighbouring particles of different phenomena [45]. The statistical dependence between pixels is defined based on their neighbourhood system. In other words, pixels are considered to have a relationship with their neighbours using the concept of context. This brings the concept of smoothness prior model which allows producing smooth image classification pattern [7].

MRF theory including its formulation described below is adapted from [7] with some minor changes.

"Let a set of random variables $d = \{d_1, d_2, ..., d_m\}$ is defined on the set S containing m number of sites in which each random variable $y_i (1 \leq i \leq m)$ takes a label from label set L. The family d is called random field. The set S is equivalent to an image containing m pixels; d is a set of pixel DN values; and the label set L depends upon the application. The label set L is equivalent to a set of the user-defined information classes, e.g. $L = \{$water, forest, pasture, or residential areas$\}$, while in case of boundary detection, the label set $L = \{$boundary, non-boundary$\}$. There are many kinds of random field models describing ways of labelling the random variables"[7]. Markov Random Field as one of a special type of random fields is described in the next paragraph.

Based on the definition of random field, the configuration w for the set S as $w = \{d_1 = w_1, d_2 = w_2, ..., d_m = w_m\}$ is defined, where $w_r \in L (1 \leq r \leq m)$. For convenience, the

notation of w can be simplified to $w = \{w_1, w_2, ..., w_m\}$. A random field with respect to the neighbourhood system is a Markov Random field if its probability density function satisfies the following three properties;

Positivity: $P(w) > 0$, for all possible configurations of w, it has a non-zero probability and $P(w)$ is the probability of given dataset w

Markovianity: $P(w_r \mid w_{s-r}) = P(w_r \mid w_{Nr})$, this defines the neighbourhood system which can be interpreted as follow, membership value of pixel r is strongly dependent on it neighbouring pixels.

Homogeneity: $P(w_r \mid w_{Nr})$ is the same for all site r, for all pixels probability is dependent on neighbourhood pixels regardless of the pixel location.

The neighbourhood system used in image analysis defines the first-order neighbours of a pixel as the four pixels sharing a side with the given pixel, as shown in Figure 1a. Second-order neighbours the four pixels having the corner boundaries with the pixel of interest, as shown in Figure 1b. Similarly, higher-order neighbours can be extended same way. Up to five neighbourhoods order is shown in Figure1c. More specifically, when the sites from a regular rectangular lattice, as do pixels in two dimension image, the site $r = (i, j)$ has four first-order neighbours, donated by $N(i.j) = \{(i-1, j), (i+1, j)(i, j-1)(i, j+1)\}$.
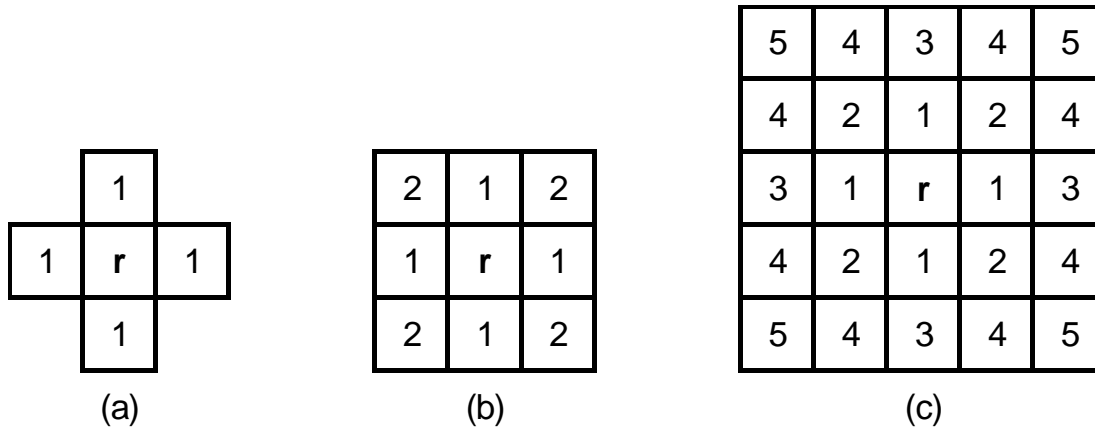
Figure 1. Nneighbourhood ordering for pixel provided in [6]. (a), (b) and (c)
depict the first, second and fifth order neighbourhood system, respectively.

The second order neighbourhood system is sufficient for some image analysis problems as suggested in previous MRF modelling study [46]. In the current research the first order neighbourhood system is constructed.

3.1.1. Gibbs Random Field

Gibbs Random Field (GRF) is considered as a popular parametric random field model. It is applicable and widely used in many areas including image analysis, where they have become an efficient tool to provide statistical models for images. GRF's characterize the global properties on an image. The probability density function then can be specified in [7] as follow:

$$P(w) = \frac{1}{Z}\exp\left[-\frac{U(w)}{T}\right] \tag{3.1}$$

where $P(w)$ the probability of w, $U(w)$ is the energy function, T is the temperature, Z is partition function that can be expressed as:

$$Z = \sum_{all-possible-configuration}\exp\left[-\frac{U(w)}{T}\right] \tag{3.2}$$

where Z is the sum of all possible configurations, in practice it is not computable except for very simple cases. This difficulty complicates sampling and estimation problem [7].

From (3.1) it can be observed that maximizing $P(w)$ is equivalent to minimizing the $U(w)$.

$$U(w) = \sum_{c \in C} V_c(w) \tag{3.3}$$

In equation (3.3) C is a clique, which is subset of image and within a clique all pairs are mutually neighbours. $C = C_1 \cup C_2 \cup C_3...$, is the collection of all possible cliques, $V_c(w)$ is the potential function [7]. Figure 2 shows all possible combinations of pixel for first order in our particular case.

$$\alpha \qquad \beta_1 \qquad \beta_2 \qquad \beta_3 \qquad \beta_4$$

(a) (b) (c)
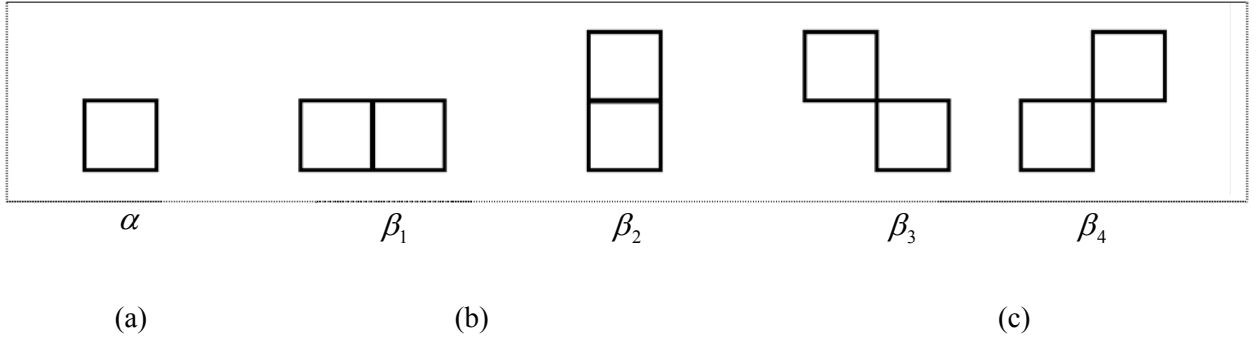
Figure 2, patterns (a) to (c) show possible cliques with the neighbourhood system on the first order.

The complexity of computation increases by increasing the neighbourhood system which also leads to growing the cliques. The energy function in (3.3) can be expanded as follow in order to understand easy:

$$U(w) = \sum_{\{r\}\in C_1} V_1(w_r) + \sum_{r,r'\in C_2} V_2(w_r, w_r') + \sum_{\{r,r',r''\in C_3\}} V_3(w_r, w_{r'}, w_{r''}) + ... \tag{3.4}$$

Each of C_1 and C_2 represents a single site clique (Figure2a) and pair-site clique (Figure2b and c).

### 3.1.2.	MRF-GRF Equivalence

An Markov Random Fields (MRF) are defined in terms of local properties, in other words, the label assigned to a pixel is influenced only by its neighbours. The Gibbs Random Fields (GRF) are defined in terms of global properties of an image. which is interpreted as the label given to a specific pixel is affected by the labels given to all pixels [7]. According to the Hammersley-Clifford theorem for every MRF a unique GRF exists as long as the GRF is defined in terms of cliques on neighbourhood system. The proof of the theorem is described by number of researchers including Tso and Mather [7]. They show that MRF-GRF equivalence offers an appropriate way to address MRF based image analysis problems.

## 3.2.	Energy minimization

### 3.2.1.	MAP Probability Solution

The theoretical part is adapted from [47] with some minor changes. Many problems in image analysis are formulated in term of energy minimization. Energy minimization is used to solve the pixel labelling problem in different applications such as image restoration and segmentation etc. MAP solution can be obtained only by minimizing the global posterior energy. The posterior energy itself consists of prior and conditional energy function. In accordance to the Bayesian formulae, the MAP solution can be represented following:

$$P(w \mid d) = \frac{P(d \mid w)P(w)}{P(d)}$$

(3.5)

where, w is the membership value and d is a given dataset. The posterior probability can be maximized as follows:

$$w = \arg\max\{p(w \mid d)\}$$

(3.6)

Equation (3.1) shows that the MAP estimate is equivalent to the minimization of global energy function and can be expressed as:

$$\hat{w} = \arg\min\{U(w \mid d) + U(w)\}$$

(3.7)

where, $\hat{w}$ is the optimal class membership value after minimizing the global energy function, $U(w \mid d)$ is the conditional energy and $U(w)$ prior energy function and the global posterior energy function can be defined as follow:

$$U(w \mid d) = U(d \mid w) + U(w)$$

(3.8)

An additional parameter of λ is added to equation (3.8) which controls the balance between the two energy functions and the value of λ ranges between 0 and 1.

$$U(w \mid d) = (1 - \lambda) \cdot U(d \mid w) + \lambda \cdot U(w)$$

(3.9)

Next, it is required to minimize the global posterior energy function in order to get the MRF-MAP estimate. Simulated Annealing (SA) and Iterated Conditional Modes (ICM) are proposed, which are able to find out the MRF-MAP solution. For this research both SA and ICM are used and comparison analysis will be provided.

3.2.2.   Iterated Conditional Modes (ICM)

Iterated Conditional Modes (ICM) algorithm was proposed by Besag, which allows to maximize the local conditional probabilities sequentially [44]. It uses a deterministic greedy or steepest strategy to perform local search [27]. First it starts with an estimate of labelling, then for each pixel it chooses the label given the largest decrease of energy function. This process is repeated until convergence [39]. It is quite fast technique and is able to find a local minimum. The theoretical background of ICM technique is adapted from [44] with minor changes.

Given the data d and the other labels $f_{S-\{i\}}^{(k)}$, the algorithm sequentially updates each $f_i^{(k)}$ into $f_i^{(k+1)}$ by maximizing $P(f_i \mid d, f_{S-\{i\}})$, the conditional probability with respect to f_i .

Following assumptions are made by calculating $P(f_i \mid d, f_{S-\{i\}})$: First of all, the observation components $d_1, ..., d_m$ are conditionally independent given f and each d_i has the same known conditional density function $p(d_i \mid f_i)$ depend only on f_i . Thus

$$p(d \mid f) = \prod_i p(d_i \mid f_i) \tag{3.10}$$

Second assumption describes that f depends on the labels in the local neighbourhood, which is concept of Markovianity. Using these assumptions and the Bayes theorem, it follows that

$$P(f_i \mid d, f_{S-\{i\}}) \propto p(d_i \mid f_i) P(f_i \mid f_{N_i}) \tag{3.11}$$

It can be mentioned that $P(f_i \mid d, f_{N_i}^{(k)})$ is much easier to maximize than $P(d \mid f)$, which is the main issue that ICM addresses.

Maximising (3.11) is equivalent to minimizing the corresponding posterior potential using the below provided rule

$$f_i^{(k+1)} \leftarrow \arg\min_{f_i} V(f_i \mid d_i, f_{N_i}^{(k)}) \tag{3.12}$$

where

$$V(f_i \mid d_i, f_{N_i}^{(k)}) = \sum V(f_i \mid d_i, f_{i'}^{(k)}) + V(d_i \mid f_i) \tag{3.13}$$

An example of discrete restoration formulated in Section 2.2.2 [44], the posterior potential for (2.19) can be considered as follow

$$V\left(f_i \mid d_i, f_{N_i}\right) = \left(f_i - d_i\right)^2 / \sigma + v_{20} \sum_{i' \in N_i} \left[1 - \delta\left(f_i - f_{i'}\right)\right] \tag{3.14}$$

where $\sum_{i' \in N_i}\left[1 - \delta\left(f_i - f_{i'}\right)\right] = \#\left\{f_{i'} \neq f_i \mid i' \in N_i\right\}$ is the is the number of neighbouring sites whose labels

$f_{i'}$ differ from f_i. For discrete $\Psi, V\left(f_i \mid d_i, f_{N_i}\right)$ is evaluated with each $f_{i'} \in \Psi$ and the label causing the lowest $V\left(f_i \mid d_i, f_{N_i}\right)$ value is chosen as the value for $f_i^{(k+1)}$. When applied to each i in turn, the above defines an updating cycle of ICM. The iteration continues until convergence. The convergence is guaranteed for the serial updating and is rapid [27].

The ICM result depends very much on the initial estimator f^0. It is still a challenge to known how to set the initialization properly to obtain a good solution. A natural choice for f^0 is the maximum likelihood estimate

$$f^0 = \arg\min_f p\left(d \mid f\right) \tag{3.15}$$

when the noise is identically, independently distributed Gaussian. Obviously, ICM can be applied to problems where f_i takes a continuous value. In minimizing (2.22) for continuous restoration, for example, one needs to maximize

$$V\left(f_i \mid d_i, f_{N_i}\right) = \left(f_i - d_i\right)^2 + \lambda \sum_{i' \in N_i} g\left(f_i - f_{i'}\right) \tag{3.16}$$

for each f_i. In order to achieve this, it is necessary to solve $\dfrac{dV\left(f_i \mid d_i, f_{N_i}\right)}{df_i} = 0$.

In according to the MRF algorithm, two neighbouring sites should not be updated simultaneously. The "coding method" [8] is incorporated into ICM in order to parallelize the iteration. Using codings, S are partitioned into several sets such that no two sites in one set are neighbours (Section 6.1.3) [48]. Therefore, all f_i on a single coding can be updated in parallel.

3.2.3. Simulated Annealing (SA)

Simulated Annealing (SA) is an opposed to deterministic algorithms. It is a stochastic algorithm used to find a good optimization problem by trying randomly variation of current solution [48]. A worse variation is accepted as the new solution with a probability that decreases as the computation proceeds. The slower the cooling schedule, or rate of decrease, the more likely the algorithm is to find an optimal or near-optimal solution.

A system is considered where any f in the configuration space F has following probability:

$$P_T(f) = [P(f)]^{1/T} \tag{3.17}$$

where, $T > 0$ is the temperature parameter. $P_T(f)$ is a uniform distribution on F, when $T \to \infty$; for $T = 1, P_T(f) = P(f)$; as $T \to 0, P_T(f)$ is concentrated on the peak(s) of $P(f)$. This explains how samples of f distribute in F.

```
initialize T and f
repeat
      randomly sample f from N(f) under T;
      decrease T;
until (T → 0;)
return f;
```

Figure 3, the Simulated Annealing (SA) algorithm[44]

Figure 3 shows the description of SA algorithm, where it applies a sampling algorithm similar to Metropolis algorithm or Gibbs sampler [9] and able to decrease the temperature values of T. First of all T is set very high and f is set to a random configuration. The sampling $P_T(f) = e^{-E(f)/T} / \sum_f e^{-E(f)/T}$ is according to Gibbs distribution at a fixed T. T is decreased accordance to a carefully chosen schedule after the sampling converges the equilibrium at current T. This will continue until the T is close to 0. The cooling schedule plays an important role which is specified by a decrement function and a final vale [49].

Geman and Geman [9] developed two convergence theorems where the first theorem concerns convergence of the Metropolis algorithm and the second one is about SA, which states that the decreasing sequence of temperatures satisfies following:

$$\lim T^{(t)} = 0 \tag{3.18}$$

and

$$T^{(t)} \geq \frac{m \times \Delta}{\ln(1+t)} \tag{3.19}$$

where $\Delta = \max_f E(f) - \min_f E(f)$, the system then converges the global minimum regardless of the initial configuration $f^{(0)}$. These conditions are sufficient but not for the convergence. In practice equation (3.19) is very slow, therefore faster schedules need to be used and some examples are provided in [49] with their experiments.

3.3. Graph Cuts

3.3.1. Min - cut max - flow algorithm

Graph cuts algorithms have been studied in computer vision in the last years and they still remain an active research area in this field. Research has been done to develop and improve methods for energy minimization in vision. Graph theory together with a number of optimization methods are provided in this charter.

The min-cut/max-flow algorithms from combinatorial optimization are provided in [41] that can be used to minimize number of energy functions in vision [5]. These energies as well as some graph based methods [42], [50], [51], [52], [53] and [54] can be represented as (3.20). The theoretical part adapted from [5] with some minor changes.

$$E(L) = \sum_{p \in P} D_p(L_p) + \sum_{(p,q) \in N} V_{p,q}(L_p, L_q)$$

(3.20)

where $L = \{L_p \mid p \in P\}$ is called a labelling of image P, $D_p(.)$ is a data penalty function, $V_{p,q}$ is called an interaction potential, which encourage spatial coherence by penalizing discontinuities between neighbouring pixels, and N is a set of all pairs of neighbours pixels [5]. Figure 4 shows an example of image labelling problem.

(a)   (b)
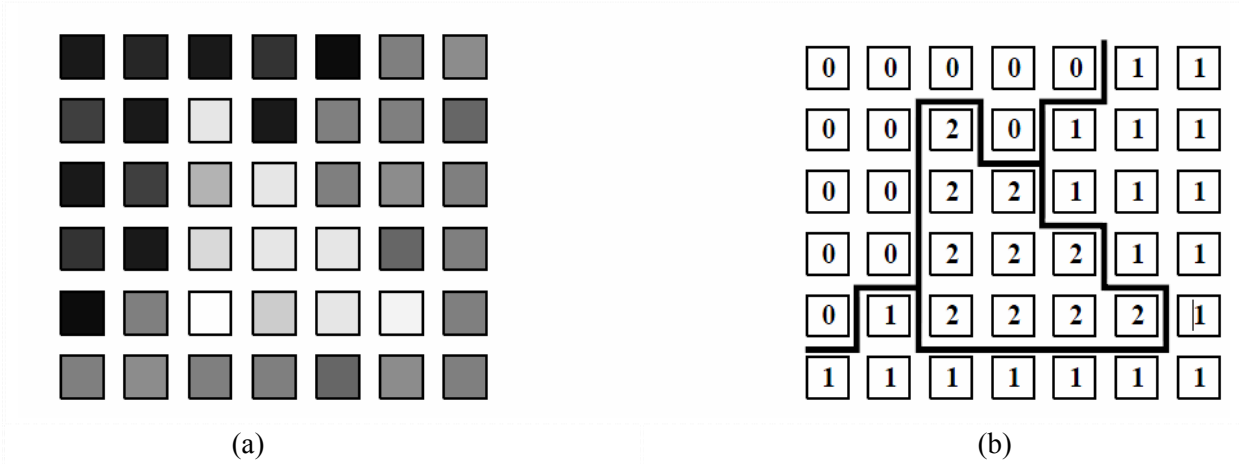
Figure 4, an example of image labelling[5].

An image in Figure 4(a) is a set of pixels P with observed intensities I_p for each $p \in P$. In the case of Figure 4 b, labelling L assigns label $L_p \in \{0,1,2\}$ to each pixel $p \in P$. Such labels can represent object index in segmentation. In case of graph based methods it is assumed that a set of feasible labels at each pixel is finite. Thick lines in Figure 4 (b) show labelling discontinuities between neighbouring pixels [5].

Greig *et al.* [41] show that the minimum cost of the graph produce a globally optimal binary labelling L in the case of Potts model of interaction in (3.20).

Further in this section some basic facts and theory on graphs are provided which are described in [5]. A directed weighted graph $G = (V, E)$ which consists of a set of nodes V and a set of directed edges E connecting them is considered in [5]. The nodes correspond to pixels in this case. The graph

contains of additional special nodes that are called terminals. These terminals correspond to the set of labels that are assigned to pixels. They are *source, s* and *sink, t* and they correspond to the set of labels that can be assigned to pixels. Figure 5(a) shows an example of a two terminal graph that can be used to minimize the Potts case of energy (3.20) on $3x3$ image with two labels. All edges in the graph are assigned some weight or cost. A cost of a directed edge (p,q) may differ from the cost of the reverse edge (q,p) [5]. There are two types of edges in the graph: *n-links* and *t-links*. *N-links* connect pairs of neighbouring pixels which represent a neighbourhood system in the image. Cost of *n-links* corresponds to a penalty for discontinuity between the pixels which are derived from the pixel interaction term $V_{p,q}$ in energy (3.20). *T-links* connect pixels with terminals (labels). The cost of a *t-link* connecting a pixel and a terminal corresponds to a penalty for assigning the corresponding label to the pixel which is derived from the data term D_p in the energy (3.20). Figure 5 shows that *t-links* are shown in red and blue, but *n-links* are in yellow

(a) A graph G (b) A cut on G
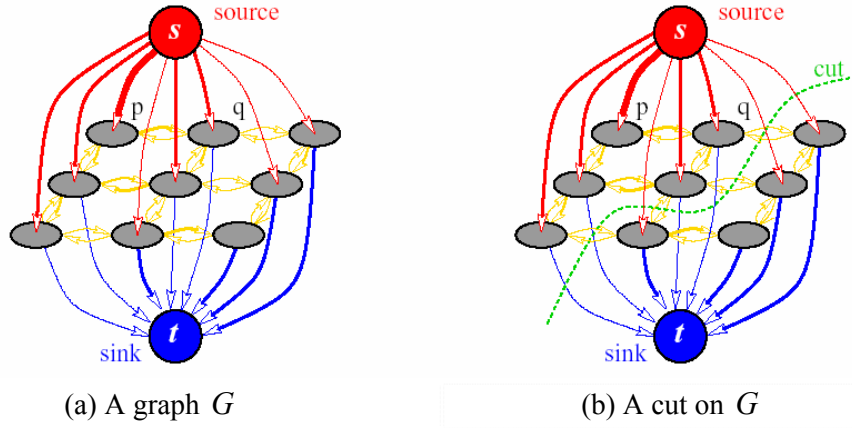
Figure 5. Example of a directed capacitated graph. Edge costs are reflected by their thickness. A similar graph-cut construction was first used in vision by Greig et al. [49] for binary image restoration.

Min-cut (or max-flow) algorithm is used to help with energy minimization over image labelling. This algorithm is used for fast energy optimization problem and it shows very good computational time. The proposed graph cuts algorithms such as swap-move and expansion-move are based on min-cut algorithm. The mathematical background of min-cut and max-flow provided below is described in [5].

An s/t cut C (can be called as cut) is a partitioning of the nodes in graph into two disjoint subset S and T such that the source s is in S and the sink t is in T. An example of cut is shown in figure 5(b). The cost of cut is $C = \{S,T\}$ is the sum of cost of boundary edges (p,q) such that $p \in S$ and $q \in T$. The cost is "directed" as it sums up weights of directed edges specifically from S to T. The minimum cut problem on a graph is to find the cut with minimum cost among all other cuts [5]. One of the results in combinatorial optimization is that the minimum s/t cut problem can be solved by finding the maximum flow from source s to sink t. "The theorem of Ford and Fulkerson [55] states that a maximum flow from s to t saturates a set of edges in the graph dividing the nodes into two disjoint parts $\{S,T\}$ corresponding to a minimum cut. Thus, min-cut and max-flow problems are equivalent. In fact, the maximum flow value is equal to the cost of minimum cut" [5].

Boykov and Kolmogorov [5] made an experimental comparison of several algorithm where they proposed a new min-cut/max-flow algorithm, which outperformed the existing algorithms. Detail implementation of the new algorithm is described in the same paper.

3.3.2. Swap-move and expansion-move algorithms

The swap-move and expansion-move algorithms are introduced in [42] are able to find the local minimum of energy function (3.20) with respect to large moves, namely α-expansion-move and α-β-swap, which allows a large number of pixels to change their labels simultaneously. These algorithms are similar in their structure. Their main is based on the graph cut method from combinatorial optimization. Both of them are used to compute the global minimum of a binary labelling problem. They are able to produce lower energy [39]. Swap-move for a pair of labels α, β takes some subset of a pixel given the label α and assigns them a label β and vice versa. This algorithm (swap-move) finds a local minimum such that there is no more swap-move for any pair of labels α-β that produces a lower energy labelling. In a case of expansion-move algorithm, it finds a local minimum such that no expansion-move for any label α, yields a labelling with lower energy.

Boykov *et al.* [42] showed that the expansion-move algorithm can be applied to any energy where V_{pq} is a metric and the swap-move can be applied to any energy where V_{pq} is semimetric. Kolmogorov and Zabih [6] considered the above conditions and showed that the expansion-move can be used if for all α, β and γ

$$V_{pq}(\alpha,\alpha)+V_{pq}(\beta,\gamma)\leq V_{pq}(\alpha,\gamma)+V_{pq}(\beta,\alpha) \tag{3.21}$$

in other words, expansion-move algorithm can be used if the binary energy for expansion-move algorithm is regular, using Kolmogorov's terminology.

For the swap-move algorithm can be used if for all labels α and β

$$V_{pq}(\alpha,\alpha)+V_{pq}(\beta,\beta)\leq V_{pq}(\alpha,\beta)+V_{pq}(\beta,\alpha) \tag{3.22}$$

in other words, swap-move algorithm can be used if the binary energy for the swap-move algorithm step is regular, due to   algorithm Kolmogorov's terminology [6].

The main computation cost of graph cuts depends on computing the minimum cut which is done via max-flow [39]. The detail information of the expansion-move and swap-move algorithm, including their lemmas, theorems, proofs and implementations are described and discussed in [42].

1. Start with an arbitrary labelling f
2. Set success : $= 0$
3. For each pair of labels $\{\alpha, \beta \subset \Psi\}$

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

    3.2 If $E(\hat{f}) < E(f)$, set $f := \hat{f}$ and success : $= 1$
4. If success = 1 goto 2
5. Return f

---

1. Start with an arbitrary labelling f
2. Set success : $= 0$
3. For each pair of labels $\{\alpha \subset \Psi\}$

    a. Find $\hat{f} = \arg\min E(f')$ among f' within one *α-expansion-move* of f

    b. If $E(\hat{f}) < E(f)$, set $f := \hat{f}$ and success : $= 1$
4. If success = 1 goto 2
5. Return f

Figure 6. Swap-move algorithm (top) and expansion-move algorithm (bottom) [50].

4. Study area, data preparation and methods

4.1. Study area and dataset

In order to test the applicability of the adopted method, two study areas are proposed with different conditions and purposes. In the case of the first study area, Landsat TM (30m spatial resolution) images and the second case Quick Bird Images (2.4m spatial resolution) are considered. Brief introduction about study areas, their needs and characteristics are provided.

4.1.1. Study area - 1

Lake Sarez is located in the Pamir Mountains of Tajikistan. It was created about almost 100 years ago when a strong earthquake triggered a massive landslide that, in turn, became a huge dam along the Murghob River, which is called the Usoi Dam. The lake is located at an elevation greater than 3000m. The length of the lake is 61 km and the depth is 500 m and holds an estimated 17 cubic km of water. "The area experiences considerable seismic activity and scientists fear that part of the right bank may slump into the lake, creating a huge wave that will top over and possibly breach the natural dam. Such a wave would create a catastrophic flood downstream along the Bartang, Panj and Amu Darya Rivers, perhaps reaching all the way to the Aral Sea. Currently, Central Asian governments, as well as the World Bank and the UN are monitoring the dam closely, and have proposed gradually lowering the lake   level   as   a   preventive   measure" (http://eol.jsc.nasa.gov/EarthObservatory/Lake_Sarez,_Tajikistan.htm).

The study area is situated in Easter part of Pamir, which covers about 24km from west to east and 24km south to north. Figure 7 shows the location of study area. The information including some statistics on Sarez Lake is provided in NASA earth observatory website above. The proposed method can be useful for classification of the lake and its surrounding area. The reference data for this study area is not available and the performance of the applied methods is assessed based on visual interpretation of the classification results.

4.1.2. Study area - 2

Quickbird images (2.4m spatial resolution) of Boothoven area, Enschede is provided to extract tree crown information from these images. Figure 6 shows the study area. An area of 400 x 400 pixels size of Boothoven is used to test the applicability of the graph based energy minimization algorithms - swap-move and expansion-move as well as the traditional energy minimization methods – Maximum Likelihood (ML), Simulated Annealing (SA) and Iterated Conditional Modes (ICM). For this study area the reference data is available to validate and quantify the accuracy of the applied methods.

The image including, including its reference data is provided by Boom en Beeld Project, http://www.boomenbeeld.nl/.

The metadata includes the following: Quickbird images: 2.4m spatial resolution, acquisition date: 26 Sep, 2006, four bands, R, G, B, NIR. The reference data is a product of DKLN 2008.
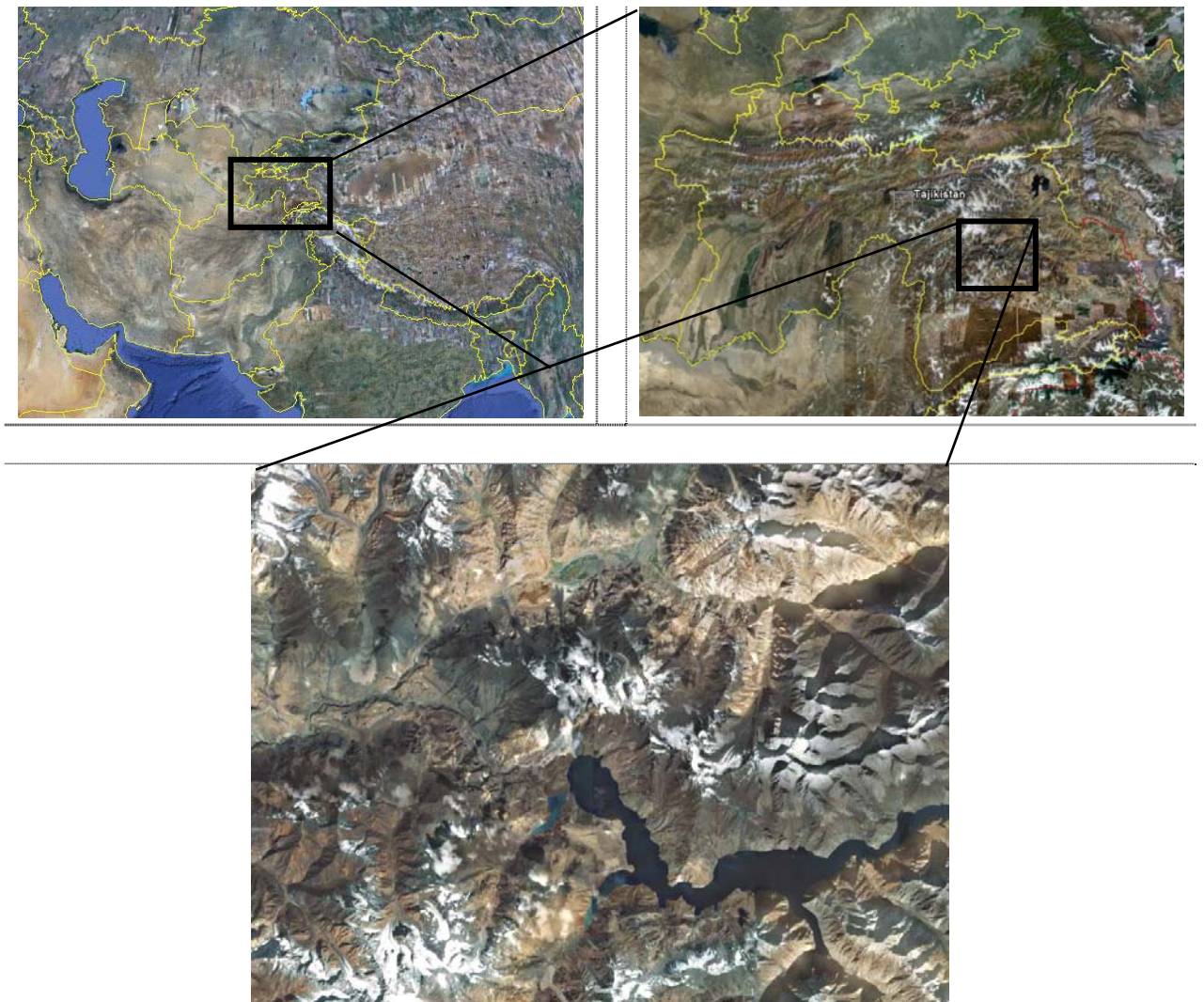
Figure 7. Sstudy area – 1, Source: Google Earth,

a) Google Earth   *b)Quick bird image of Enchede, source: Boom en Beeld*
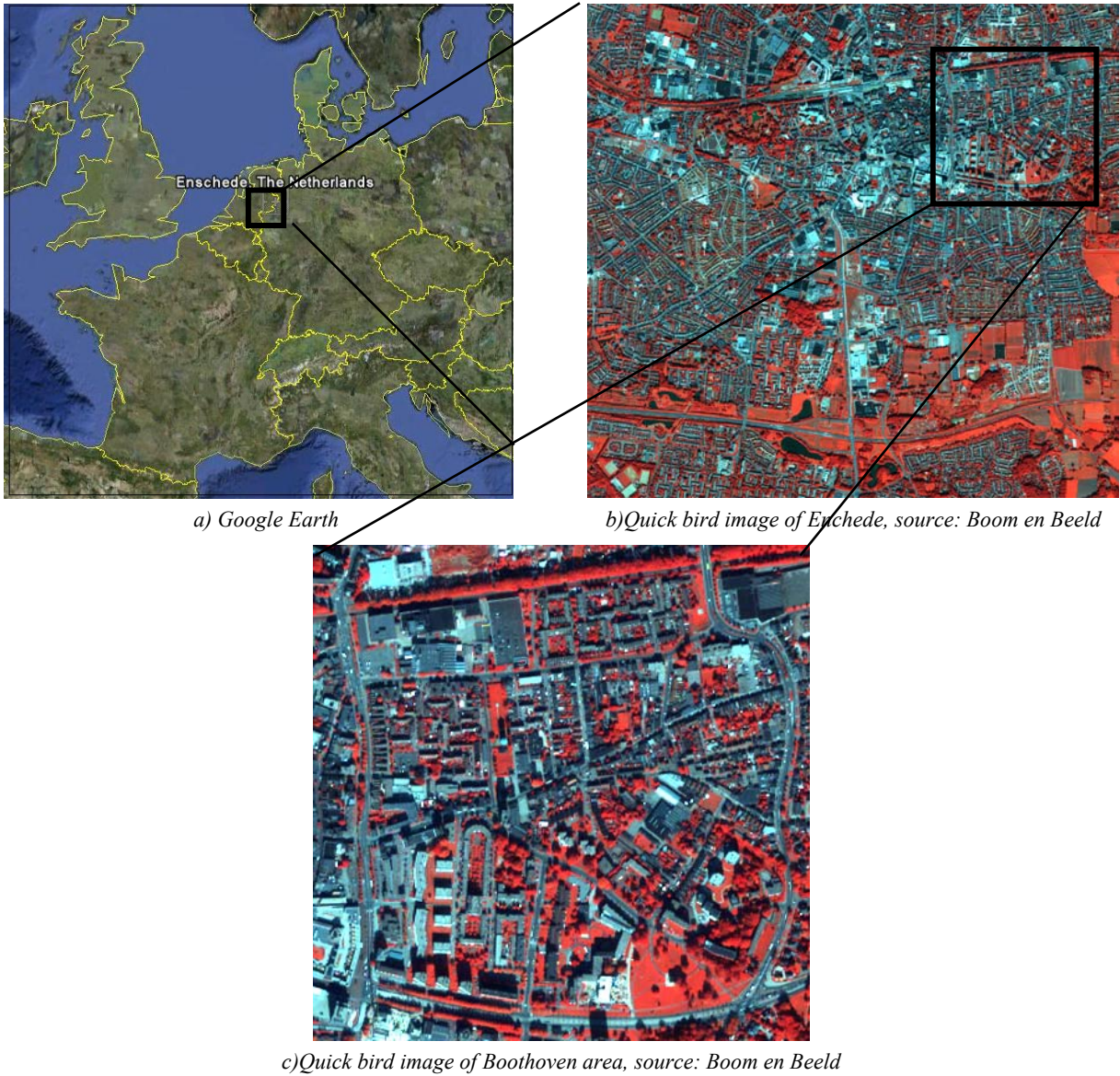
c)Quick bird image of Boothoven area, source: Boom en Beeld
Figure 8. Image of study area - 2.

4.1.3.   Selecting spectral classes and defining their training set

Eight spectral classes including their training set are defined for the study area - 1. This study area includes following spectral classes: lake, snow, landslides, vegetation, baresoil, rock, river 1 and river 2. In this case classes "landslides", "river 1" and "river 2" they have following characteristics: landslides contain mixture of rock and soil, river 1 and river 2 are spectrally different, that's why they are considered separate classes. The rest of the classes are described in a standard way. Appendix (A) shows the training sets on Landsat TM image of Sarez Lake area and appendix (B) shows the defined land cover and their training sets. Based on the defined training set the *mean* and *covariance values* are calculated which are provided in appendix (C).

Similarly, eight classes including their training sets are defined and provided for the study area – 2. These classes are following: "bright trees", "dark trees", "dark grass", "bright grass", "shadow", "impervious", "shadow vegetation" and "shrubs". The *mean* and *covariance values* for the second study area are provided in appendix (D). For this study only the first two tree classes are considered and used for classification.

4.1.4. Reference data generation

In order to validate the accuracy of applied methods normally reference data is used. The reference data for study area – 1 is not available. Therefore, Aster images are used for validation purpose, where visual interpretation is carried out. The reference data for study area – 2 is available. This data contains information about "tree crown" and "vegetation". For this particular case, "tree crown" reference data is used for validation of the applied methods. This reference data is generated from high spatial resolution (0.25m) aerial photograph. Some differences (changes) in reference data are observed. These differences might have an impact on the classification results, therefore they need to be considered and addressed in a proper way. In order to avoid this problem, manual editing is applied using ENVI software. In addition, degrading is applied on vegetation mask as well as on tree crown reference data, because of different spatial resolution. Figure 8 shows the result of manual editing in case of vegetation mask, *blue area* is the original reference data and the *yellow area* is the manual edited one. Quickbird images contain shadow in vegetation area, where in reference data (vegetation mask) this area is considered as vegetation. The vegetation mask contains also the information on tree crown.
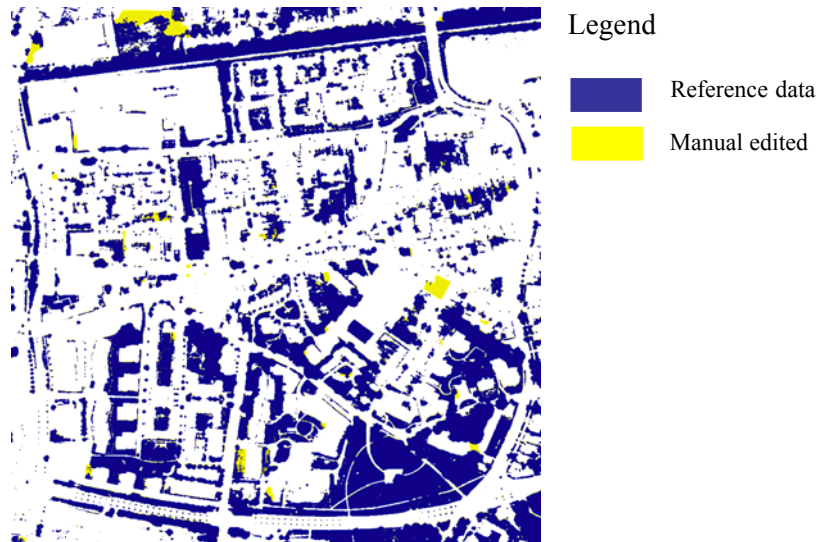
Figure 9. The reference tree crown data

Figure 9 shows the reference data for *tree crown*. This reference data does not contain all the trees. At the same, time the reference data contain small trees and due to limitation of spatial resolution these trees are not visible on Quickbird images. Another problem which has a direct impact on classification result is that some the trees are covered by shadow, both in urban and non-urban areas. These limitations should be taken into account when accuracy assessment is done.
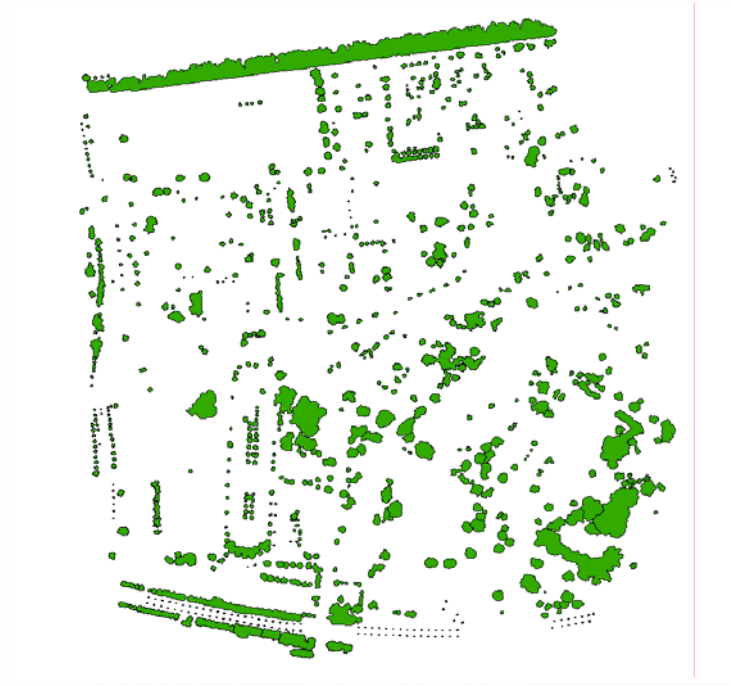
Figure 10. Reference data for tree crown

4.2. Methods

The adopted method for this research is described in details the following sections.

4.2.1. Energy minimization

In order to minimize the energy function it is important to describe following terms using graph based energy minimization methods. These terms are described in following section.

4.2.2. Setting data and smoothness terms

Energy type of (3.20) is considered for defining the data and smoothness costs which is described in section 3.1.1.

In order to apply the energy minimization algorithms on binary image segmentation problem, it is important to prepare an input file. In accordance to graph cut theory for energy function type (3.20), data and smoothness costs should be computed for each individual pixel in the image. The content of input file includes following: number of rows and columns, number of labels, for binary case there 2 labels, a label for object of interest (foreground) and a label for the rest part of the image (background), the ignorance part, data costs for each single pixel and the last part is the smoothness costs.

Figure 11 shows how datacots for classes "lake" and "non-lake" are computed. The datacosts are likelihood energies that computed from normal distribution, where Mahalanobis distance is used. In the case of class "non-lake", the least cost among the rest of the classes (seven classes) except the class of interest ("lake") is computed. This procedure applies for all "class of interest". Figure 11 (a) represents datacost for class "lake" and Figure 11 (b) represents data cost for class "non-lake" or background. The dark image (Figure 11 (a)) represent a low cost, where the bright image (Figure 11 (b)) represents a high cost. The main idea of graph based method is to find out the energy with least

cost. This cost or energy can be related to probability, low cost means high probability and high cost means law probability. Following example is provided: in the case of low cost (Figure 11 (a)) the probability of pixel belonging to class "lake" is high and inverse in the case of Figure 11 (b). The type of datacosts for study area – 1 is integer, where the type of datacosts for study area – 2 is floating point. In terms of smoothness cost, first-order neighbourhood system is used and $(\lambda = const)$, it can have any range of value.

<div align="center">

a) Datacost (lake)   b) Datacost (non-lake)
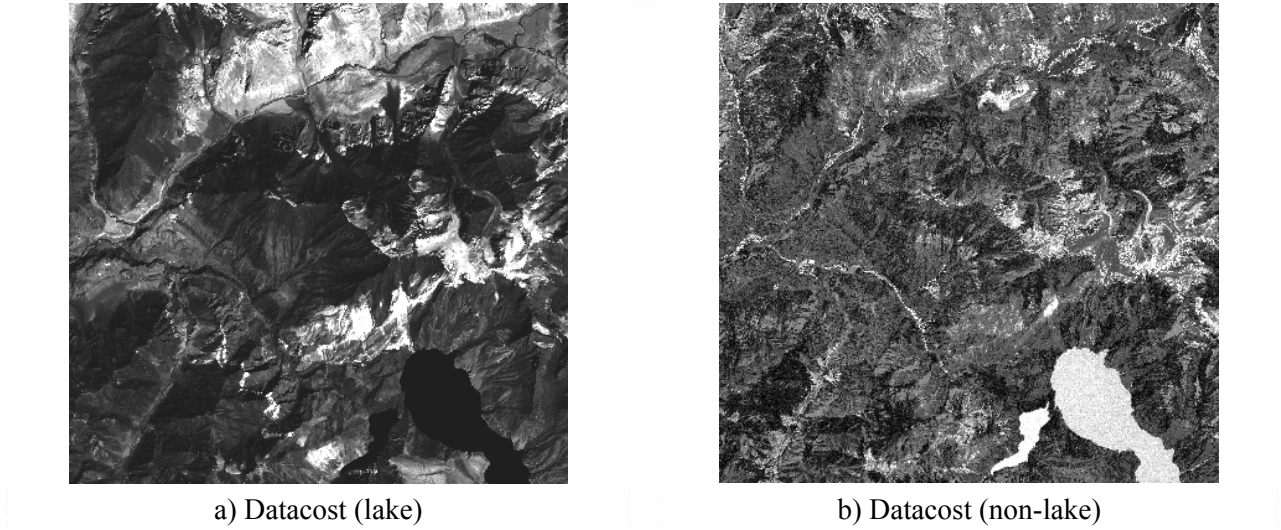
Figure 11. Data-costs for classes "lake and non-lake"

</div>

4.2.3.   Specifying and optimizing the energy

In order to optimize the energy in is necessary to specify it, which means, the number of labels and pixels, neighbourhood system, the data and smoothness terms needs to be specified. Two constructors are used one for grid graph case and another one for general graph. In all cases it is considered that the pixels go between *0, …, number of pixels-1*, and the labels between *0, …, number of labels-1* [6]. The details on constructors are described in library which is availably through the following website: http://vision.middlebury.edu/MRF/. Certain type of functions is used for labelling problem which is provided in graph cuts library.

4.2.4.   Optimizing of smoothness parameter's value

Before applying binary segmentation it is important to consider the smoothness parameter (λ) values that are optimal for certain spectral class. The adopted method depends on class separability measures, - divergence in this case. The aim is to find the most appropriate and/or optimal smoothness value that can best represents a particular spectral class. There is a relationship between smoothness values and class separability measures. Based on divergence report, some spectral classes are least separated from other classes and some other are most separated. Based on this report it is observed that following spectral classes; "lake", "snow" and "vegetation" belong to those classes that are most separated from other classes and classes "river -1", "river 2", "baresoil", "landslides" and "rock" belong to those that are spectrally least separated from other classes. In other words, those classes that are most separated from other, easy to differentiate and classes that are least separated difficult to differentiate from other spectral classes. Based on these factors it can be expected that different smoothness value might lead to different classification results, which gives a sufficient reason to study and find out the optimal smoothness value for each spectral class. A "trial and error" method is used to test and find out the

optimal smoothness value for each spectral class. All possible combinations of smoothness values are tested and applied on each class individually and the most appropriate that better represents the particular class is selected.

4.2.5. Designing the classification tree

Since the graph cuts algorithms are applied to solve binary image segmentation (to extract the object of interest as foreground and the rest as background) problems, it is important to find out a proper method which can address a multi label image classification using these algorithms. For this purpose, a method is proposed that used to construct a classification tree. Different tree structure with MRF based models are proposed in literature ([56], [24], [25],[26] and [57]) which are discussed in literature review chapter. The method proposed for this study is a slightly different to those discussed in chapter 2.

The proposed method is based on class separability measures. The method makes use of class separability measures. It is important to consider these measures before construction of the classification tree. This method considers three statistical measures that are derived from class separability report. These are the max, mean and min values which are described below. In addition, an inverse tree nodes order is proposed.

Class separability is a measure that provides statistics on similarity between classes. Four quantitative measures are widely used for class separability. These measures are following: Divergence D, Transformed Divergence TD, Bhattachatyya distance - B and Jeffries-Mutasita distance – JM [58]. For this particular study Divergence is selected. Mathematical expression of this measure is described below.

For any pair of class a and b the above measure is defined following.

Divergence:

$$D_{ab} = \frac{1}{2}(\mu_a - \mu_b)'\left(C_a^{-1} + C_b^{-1}\right)(\mu_a - \mu_b) + \frac{1}{2}Tr\left[\left(C_a - C_b\right)\left(C_b^{-1} - C_a^{-1}\right)\right], \tag{4.1}$$

$Tr[A]$ is a trace matrix A. Divergence takes a value from 0 to ∞. If the two classes have shared values, it means that there is in no difference between these classes based in spectral information.

These measures are used for the study area - 1. Training set for eight classes is defined, which is used to estimate the mean vector μ_a and covariance matrix C_a of the normal distribution for each class a. The class separabity measures using Divergence are calculated and provided in Table 1.

Class name	Lake	Snow	Landslides	Vegetation	Baresoil	Rock	River1	River2
Lake	0	34691	18887	9042	9552.3	2118.3	2002.2	7188
Snow	34691	0	744	2665	2579.8	2144.6	1213.7	188
Landslides	18887	744	0	305	113.1	106.2	165.0	114
Vegetation	9042	2665	305	0	404.6	606.0	395.1	1059
Baresoil	9552	2580	113	405	0.0	80.6	72.8	379
Rock	2118	2145	106	606	80.6	0.0	61.6	474
River1	2002	1214	165	395	72.8	61.6	0.0	477
River2	7188	188	114	1059	379.0	473.9	477.3	0

Table 1. Class separability table (Divergence).

These measures are used to construct the classification tree. The *max*, *mean*, and *min* values per classes are computed from Table 1 and provided in Table 2.

Class name	Lake	Snow	Landslides	Vegetation	Baresoil	Rock	River1	River2
Lake		34691	18887	9042	9552.3	2118.3	2002.2	7188
Snow	34691		744	2665	2579.8	2144.6	1213.7	188
Landslides	18887	744		305	113.1	106.2	165.0	114
Vegetation	9042	2665	305		404.6	606.0	395.1	1059
Baresoil	9552	2580	113	405		80.6	72.8	379
Rock	2118	2145	106	606	80.6		61.6	474
River1	2002	1214	165	395	72.8	61.6		477
River2	7188	188	114	1059	379.0	473.9	477.3	
	34691	34691	18887	9042	9552.3	2144.6	2002.2	7188
	11926	6318	2919	2068	1883.2	798.7	626.8	1411
	2002	188	106	305	72.8	61.6	61.6	114

Table 2. The max, mean and min values of divergence table

The results of these measures (*max*, *mean* and *min*) from second part of the Table 2 are ordered in a proper way. Decreasing order is considered in this particular case. Table 3 shows this order for each measure.

Max values	34691	34691	18887	9552	9042	7188.4	2144.6	2002.2
Mean values	11926	6318	2919	2068	1883	1411.3	798.7	626.8
Min values	2002	305	188	114	106	72.8	61.6	61.6

Table 3. Ordering class based on max, mean and min values

After putting in order all the results in Table 3, each value is assigned to its original class. In other words the values in Table 3 are represented by class name (Table 4). This order or sequence for each measure (max, mean and min) is used to construct a classification tree. Each level of sequence represents node in tree. An addition, an extra order (inverse order to min value) is proposed. These orders are used to test and find out the most appropriate order that better produces classification result.

Max values	Snow	Lake	Landslides	Baresoil	Vegetation	River2	Rock	River1
Mean values	Lake	Snow	Landslides	Vegetation	Baresoil	River2	Rock	River1
Min vlaues	Lake	vegetation	Snow	River 2	Landslides	Baresoil	Rock	River1
Inv, ord.m.v	River1	Rock	Baresoil	Landslides	River 2	Snow	vegetation	Lake

Table 4. Defined order (sequence) for classification tree

All four cases are proposed for the current method. First, constructed classification tree based on maximum values Figure 12(a), second based on mean values Figure 12(b), third based on min values Figure 12(c) and last the inverse order to min value Figure 12(d) are defined and proposed. From

Figure 12 it is observed that the sequence of nodes for classification tree is different in all cases. This difference can have a possible impact on classification results, which is addressed and studied in this research. All cases need a careful study and the most appropriate method will be identified. The results based on these methods are provided in chapter 5.

Figure 12 shows four proposed methods for classification trees which are constructed based on Table 4. "O of I" next to the class name means – an object of interest and "Root" is from where construction of classification tree starts.

a) based on *max value*

b) based on *mean value*

c) based on *min value*

d) based on *inverse order to min value*
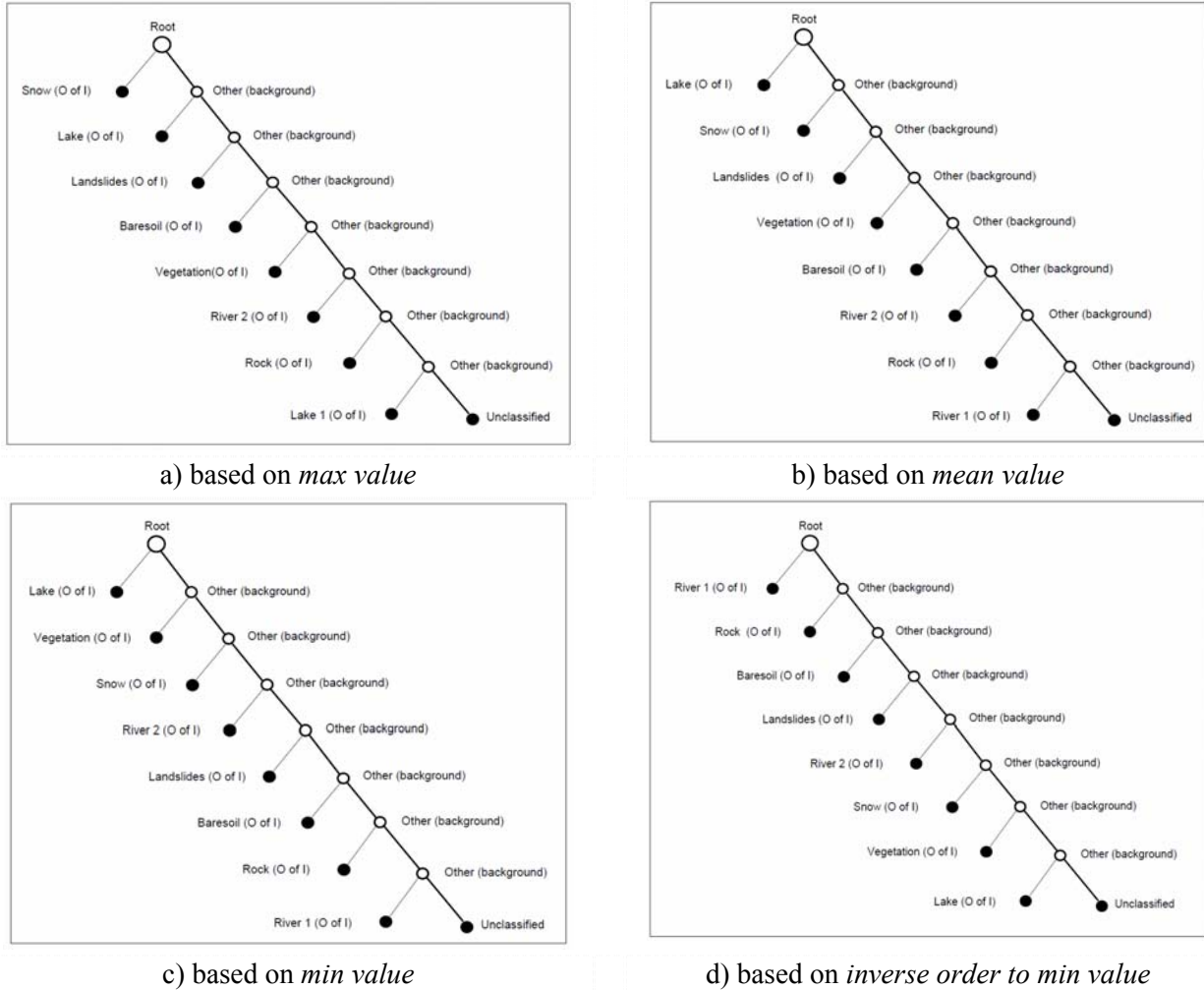
Figure 12. Defined classification tree for binary image classification.

4.2.6. Applying the energy minimization algorithms

After defining the above methods, both construction of classification tree and optimizing the smoothness parameter it is necessary to apply the energy minimization algorithms (ICM, SA, Swap-move and expansion-move) on both Landsat TM and Quickbird mages. Graph cuts optimization software is used to check the applicability of swap-move and expansion-move algorithms on MRF based remotely sensed image analysis. In addition, ICM and SA algorithms are applied on the same problem.

4.2.7. Performance evaluation and accuracy assessment

The last stage of the adopted method provides information about the methods that applied to evaluate the performance of the energy minimization algorithms, both in term of computation and the accuracy assessment. In order to check the performance of all algorithms, it is necessary to find out the most appropriate measures to evaluate them. Numbers of accuracy assessment methods are proposed in literature and the applicability of each method depends upon the specific purpose. The performance evaluation is different between the study areas.

Study area – 1: For first study area the quality of classification results is assessed using visual interpretation by comparing the results to Aster images. In terms of computation, the algorithms are able to provide information on computational time. Based on these measures the methods are compared.

Study area – 2: In this case, confusion matrix or error matrix is used to quantify the classification accuracy. At the same time, visual interpretation of the results is carried out. In addition, object based analysis is considered, where ArcGIS functionality such as raster calculators are used to evaluate the quality of object based analysis.

4.2.8. Hardware and Software

Desktop computer with following specifications is used for implementation of the adopted methods: Intel(R) Core (TM) 2 Duo CPU, 3.00GHz, 4 GB of RAM. Following operating systems and software are used: MS-WIN XP and LINUX. In a case of Linux, Virtual Machine is used, where 1 GHz processor and 2GB of RAM are allocated to this system. ERDAS IMAGINE, ENVI, ArcGIS as well as R and Code Blocks for programming part are used. In addition to these graph cuts algorithms including MRF codes are used, which are available in C++ language. The graph cuts codes are developed by number of people from vision [5, 39, 42]. The C++ codes for SA and ICM are provided by my first supervisor.

.

5. Results

5.1. Classification results of Landsat TM image

First part of the chapter provides the results of energy minimization algorithms for binary image classification on Landsat TM image (Lake Sarez). In this case the results of all energy minimization methods are visually compared. This chapter demonstrates how the smoothness parameters have an impact on classification results.

The results of ICM algorithm as well as the classification results of swap-move and expansion-move algorithms with different smoothness parameters are provided.

5.1.1. Classification results using Iterated Conditional Modes (ICM)

ICM algorithm is applied on binary image segmentation with defined land cover classes. Binary image segmentation is applied with different smoothness parameters. The aim of this method is to observe and study the influence of smoothness parameter on particular spectral class. Figure 13 shows binary image classification based on min value, where smoothness parameter value $\lambda = 1$ is used. The black area on the map is an object of interest (foreground) and the white is a background in this case.

| a) Lake | b) vegetation |
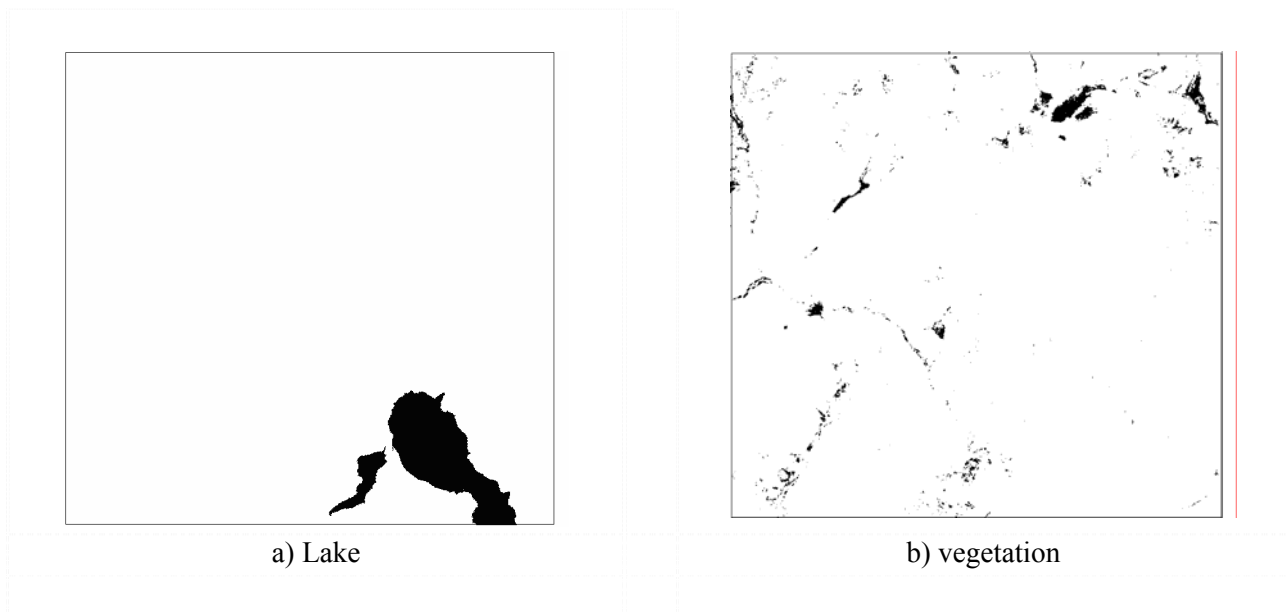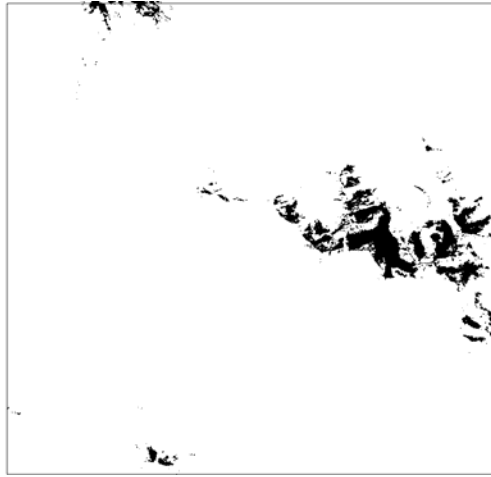
c) Snow
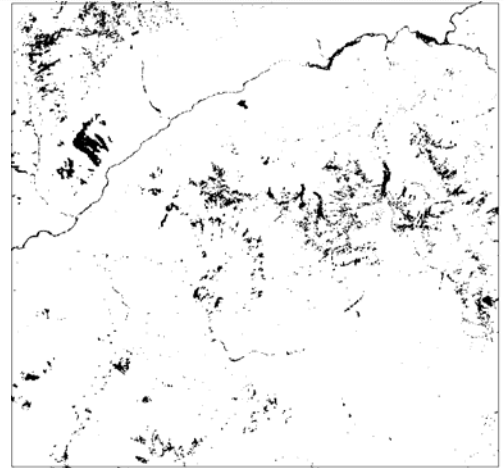
d) River 1
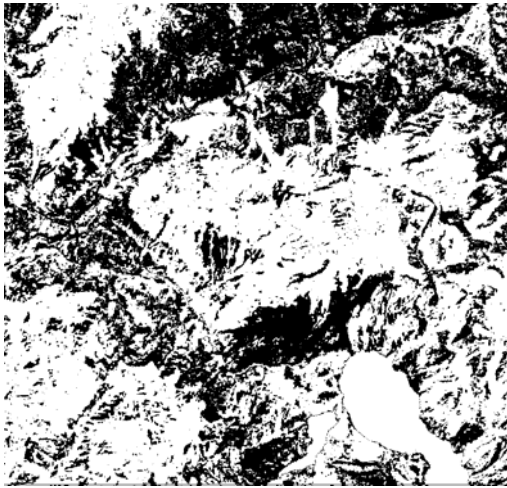
e) Landslides
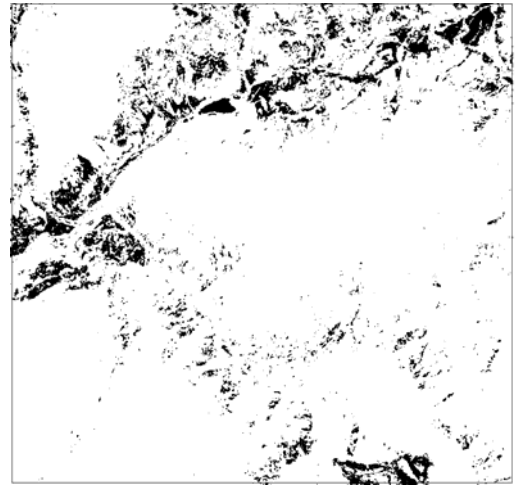
f) Baresoil
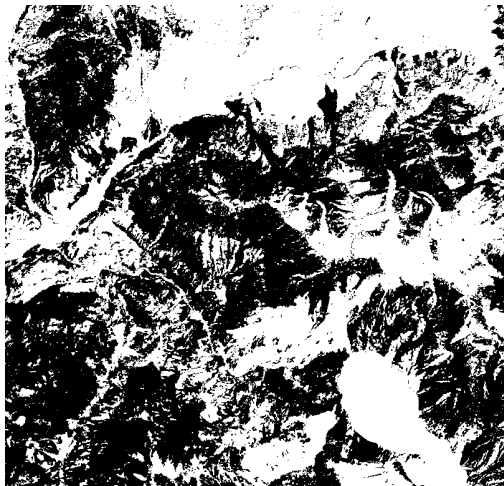
g) Rock
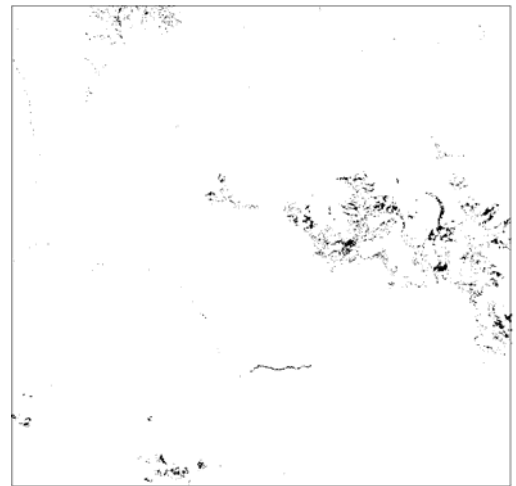
h) River 2

Figure 13. Binary Segmentation using ICM algorithm based on "min value" method ($\lambda = 1$)

Similarly, the same way, binary image segmentation is applied using the other three methods. Figure 14 provides final classification results using all the adopted methods. As mentioned above, the aim of constructing different tree structure is to find out the most appropriate classification tree that produces better results.

Table 7 shows computation time (in seconds) of binary image segmentation using ICM algorithms.

Class	energy	time(s)
Lake	7671198	1.66
Vegetation	7681090	0.33
Snow	7874489	0.85
river 1	7683173	0.49
Landslides	7738470	0.48
Baresoil	7817700	0.67
Rock	7720894	0.68
River 2	7684220	0.48

Table 5. Computation time of ICM algorithm

5.1.2. Overlap analysis

By combining the binary classification results in Figure 13 and assigning a specific colour to particular class, final classified map is obtained (Figure 13). The results can be slightly different from one method to another. Unclassified objects are appearing in most cases (Figure 15). A study has been carried out during the research is to find out the overlap area in a final classified map. For this case study the results of swap-move algorithms are presented.

The aim of overlap is to find out whether a single pixel is classified in more then one class or not. Figure 14 show the results of overlap analysis. Figure 14 (a) shows the sum of all classes, the black area in the image is the sum of all pixels that are assigned to particular class (classified pixels) and the white area is unclassified objects, in other words, pixels that do not belong to any of defined classes. Histogram (b) shows the proportion of pixels that are classified and not classified (unclassified). This histogram counts the sum of pixels that are classified and pixels that are remained unclassified. Number 1 shows the sum of pixels that are assigned to a particular class and number 0 shows unclassified pixels. From the histogram it can be observed that there are no pixels equal to number 2, which means that each pixel is classified only once and to one class. From the Figure 14 t can be concluded that there is no overlap between the classes. Also it can be concluded that the unclassified objects belong to those range of spectral signatures that are not considered during the collecting training sets.
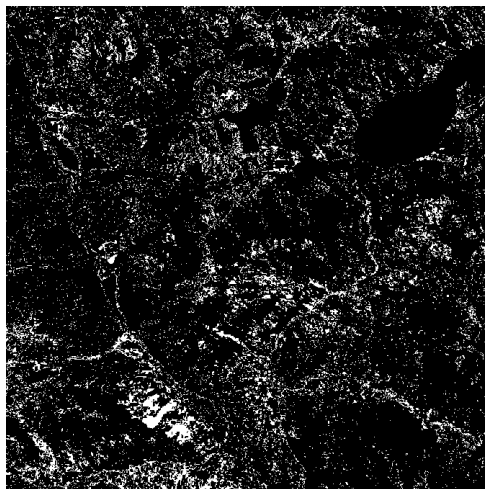
a) sum of all classified classes   b) Histogram of class overlap
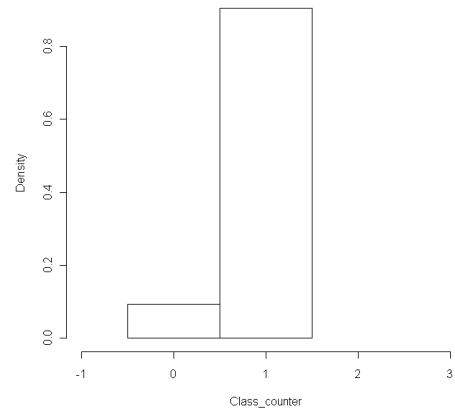
Figure 14. Overlap analysis

5.1.3.   Influence of classification tree order

When a binary classification is done, they results can be used to produce a final classification using the proposed methods. The results in Figure 15 show the influence of defined classification trees on the classification results.

a) method based on *max value*

b) method based on *mean value*

c) method based on *min value*

d) method based on *inverse tree nodes order to min value*
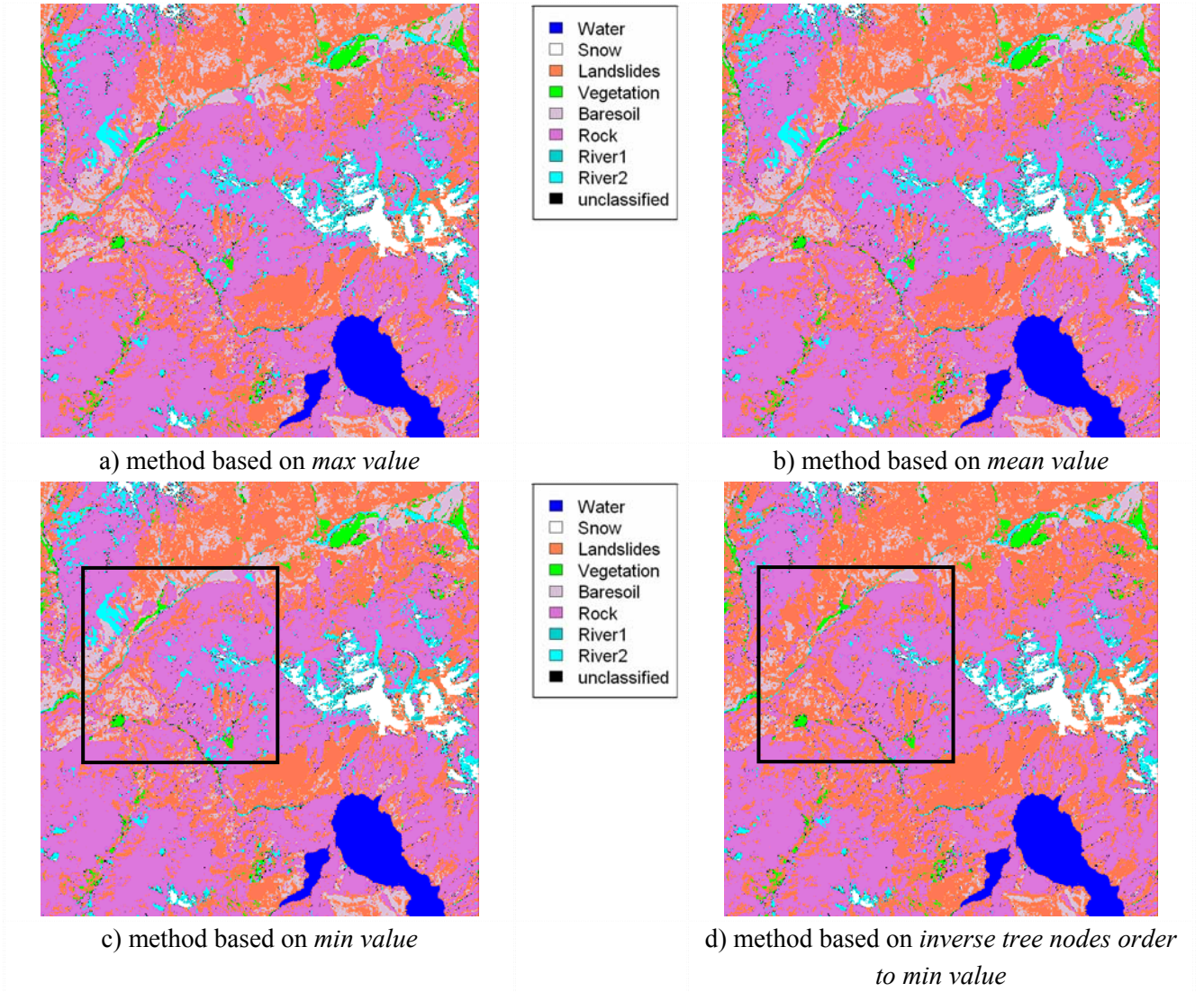
Figure 15. Final classification results using the adopted methods ($\lambda = 1$)

From the Figure 15, it is observed that the first three methods (Figures 12 a, b, c) produce identical results. The last method based on an "inverse tree nodes order" (Figure 10 d) produces different to the three previous results, which means that the sequence of tree nodes has an impact on classification result. From the first three methods it can be concluded that the class separability measures produces identical results. Therefore, for further comparison the last two methods are used, method based on min values (Figure 15c) and its inverse tree nodes order (Figure 15d).

Figure 16 shows the difference between (by zooming to a certain area on the image) the min value (Figure 16 c) and its inverse order (Figure 16 d).

c) method based on *min value*   d) method based on *inverse order to min value*
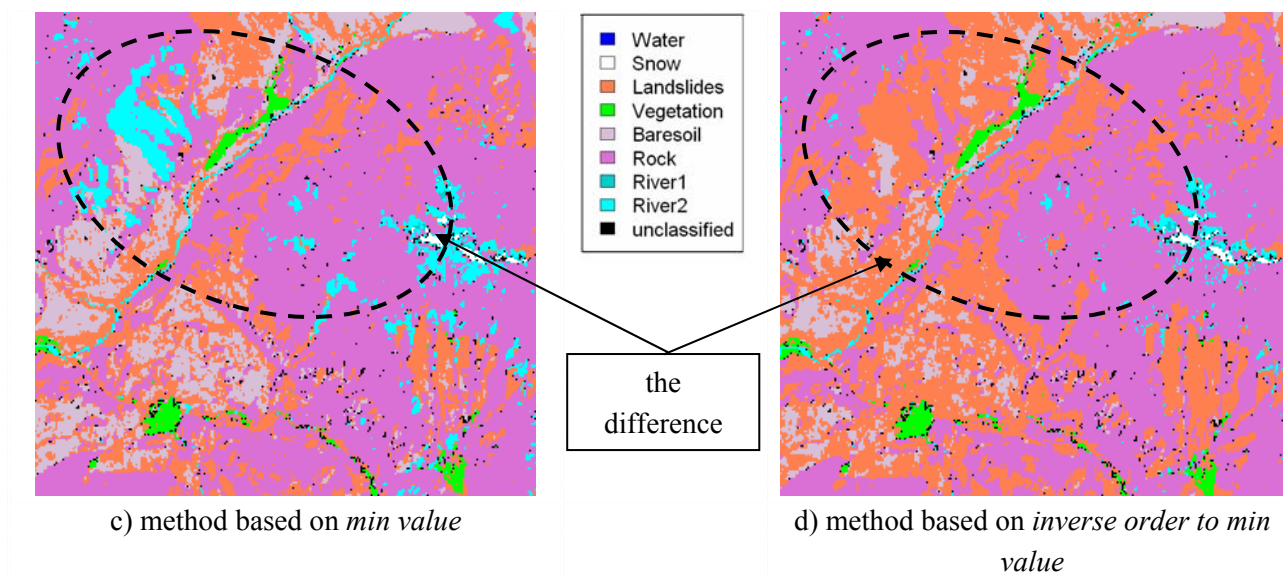
Figure 16. Classification results, min value (Figure 15c) versus its inverse order (Figure 15d) ($\lambda = 1$)

Some changes (commission errors) are observed in case of class "river 1" and "baresoil". In case of class "vegetation" no changes are observed. Class "vegetation" is the only class that represents most separated class in this small area of an image. The rest of the classes in this case belong to least separated classes. The classification results can be related to two factors, smoothness value and the classification tree. Both of these factors can have an impact on Classification results. Using a different smoothness values the classification results look different. En example of applying different smoothness parameter using binary segmentation is provided next subsection.

5.1.4.   Classification results using swap-move and expansion-move algorithms

Similarly to ICM algorithm, binary image segmentation is applied using the swap-move and expansion-move algorithms. The final results of binary image classification using these algorithms are provided in this subsection. Figure 17 illustrate the final classification results using swap-move and expansion-move algorithms, the min values (based on Figure 10c) and its inverse order (based on Figure 10 d).

a) based on *min value*   b) based on *inverse order to min value*
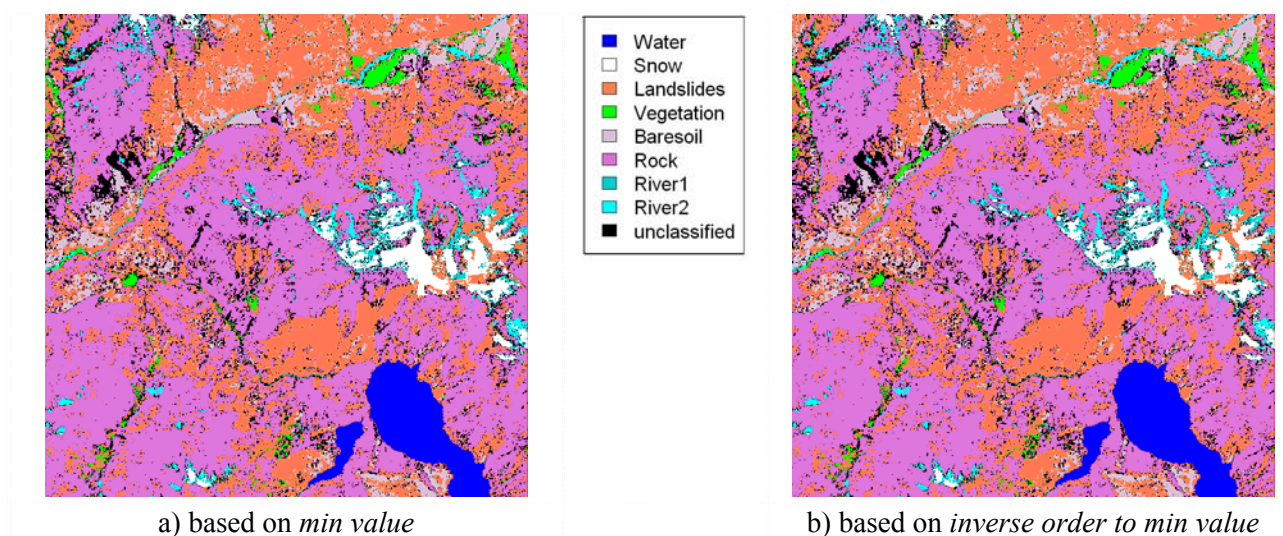
Figure 17. Classification results using swap-move algorithm ($\lambda = 1$)

In the case of graph cuts algorithm the difference is not much in comparison with ICM algorithm. The differences are assessed based on visual interpretation. Figure 18 shows the detail where some minor differences are visible.

a) based on *min value*   b) based on *inverse order to min value*
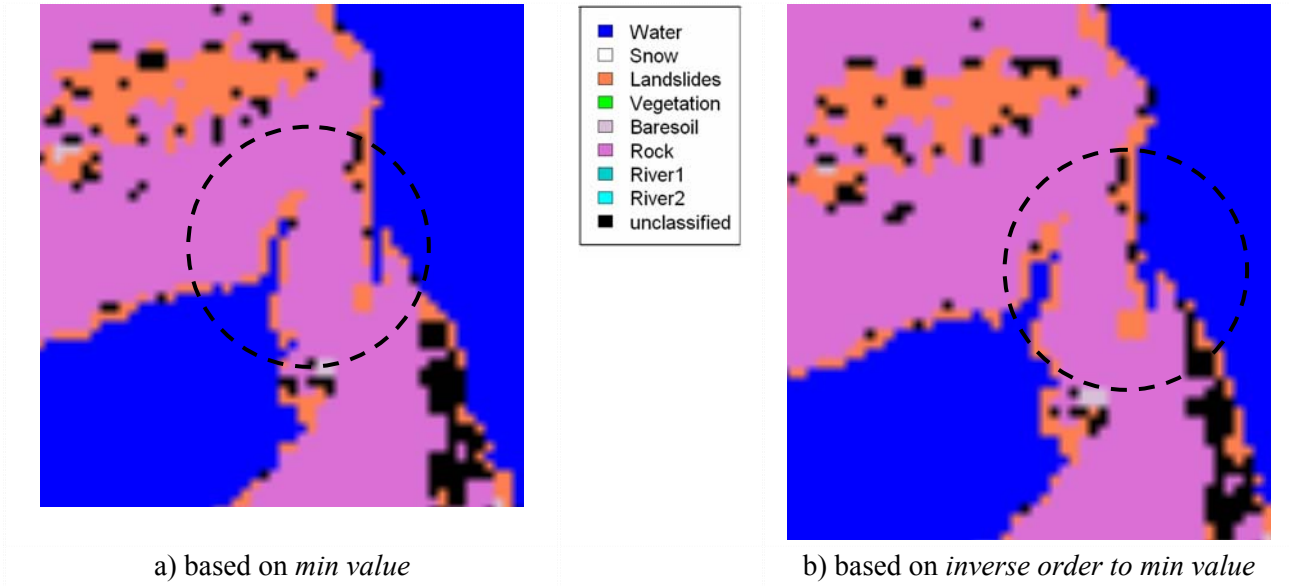
Figure 18. Classification results, using swap-move algorithm ($\lambda = 1$)

Similarly to the swap-move algorithm classification results are provided for expansion-move same way. Figure 19 shows the results of both methods.

a) based on *min value*   b) based on *inverse order to min value*
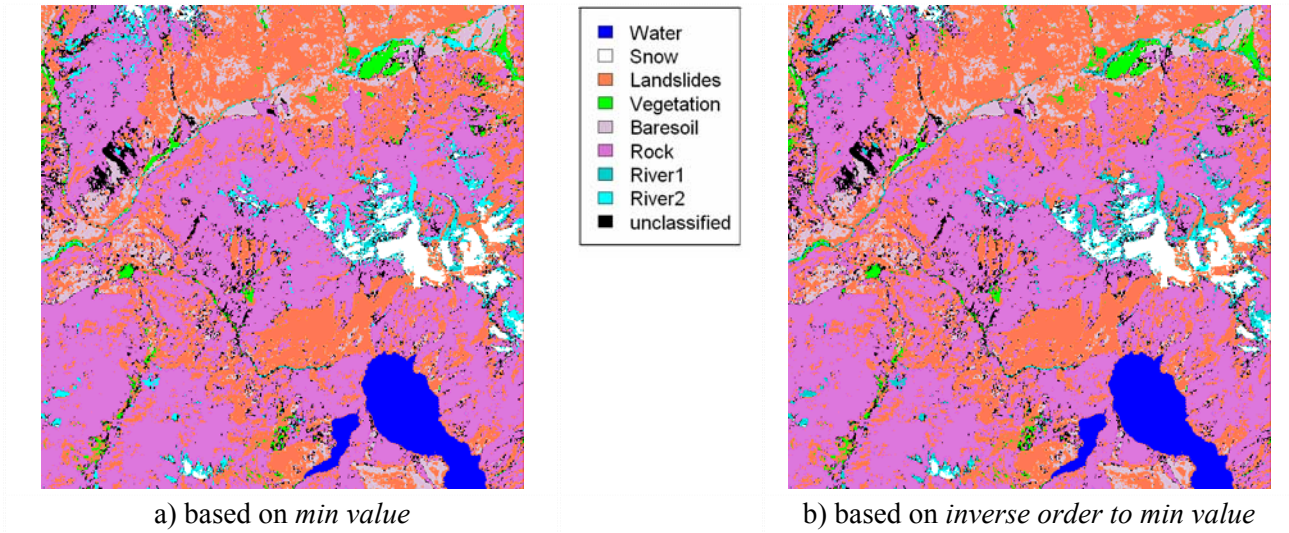
Figure 19. Classification results based on min value and its inversed order ($\lambda = 1$)

Figure 20 shows some minor differences between the methods in more details. The differences are observable particularly in the edges of classes.

| a) based on *min value* | b) based on *inverse order to min value* |
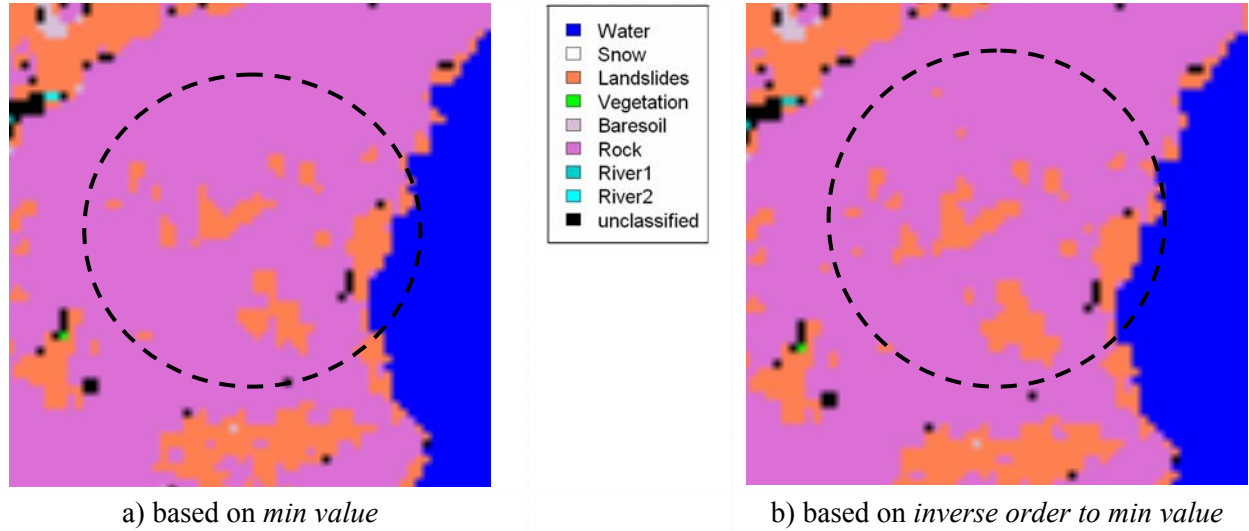
Figure 20. The difference between min value and it inverse method ($\lambda = 1$)

Based on the achieved results (Figures 15, 17 and 20) it is observed that each method produces different classification result, which depends on the sequences of the tree nodes. An example of binary image classification with different smoothness parameters is provided. The choice of most optimal smoothness value depends on users. Some user may need smooth images and some may not.

Binary image segmentation with different smoothness parameters on class "lake" and "landslides" is applied (Figure 21). These two classes are selected to study the impact of different smoothness values on classification results. Class "Lake" represents those spectral classes that are easy to differentiate from other classes and "landslides" represents those classes that are difficult to differentiate from other. This can be related to the class separability measures, discussed above. The results of binary classification are provided in Figure 21 – 22.

| a) $\lambda = 1$ | b) $\lambda = 3$ | c) $\lambda = 5$ |

d) $\lambda = 10$ e) $\lambda = 20$ f) $\lambda = 50$
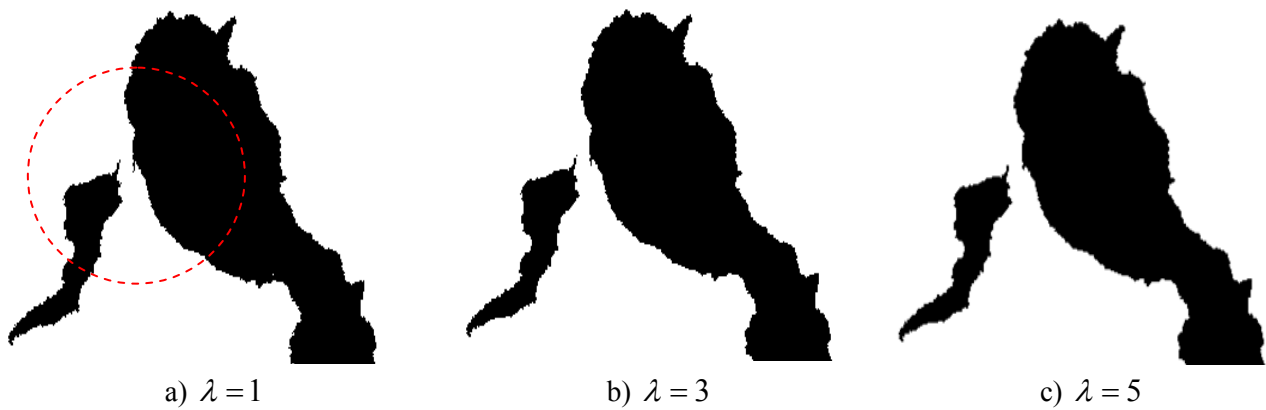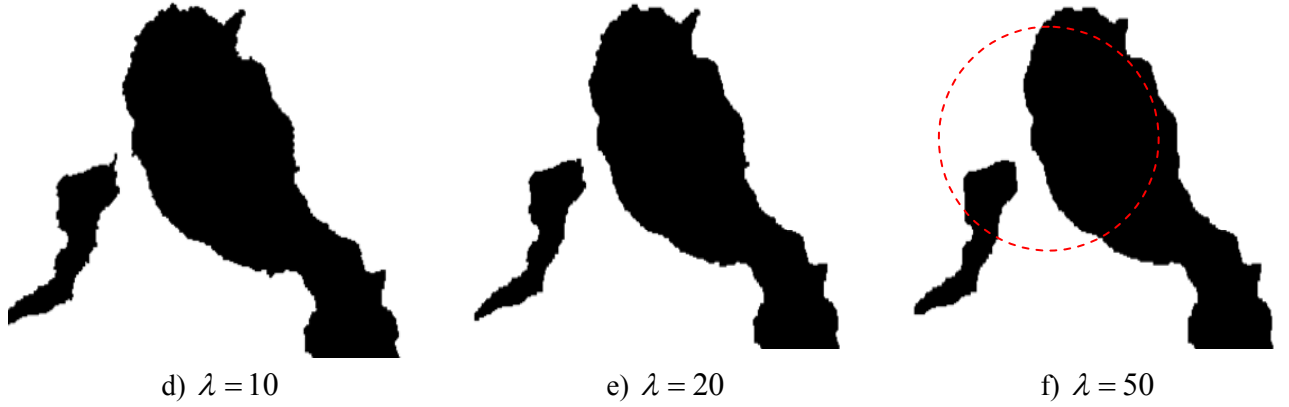
Figure 21. Binary classification results of class "lake" using swap-move algorithm

Figure 21 shows binary image segmentation results with different smoothness parameters using swap-move algorithm. By setting the smoothness parameter to larger value, more smoothing of class "lake" is observed. If we compare the results of (*a*) with (f) in Figure 21 (with smoothness parameter $\lambda = 1$ and $\lambda = 50$)*,* then the difference in results can be easily observed, although the difference between the first three smoothness values is not much observable. Following can be predicted from Figure 21: by assigning larger smoothness value, over smoothing can be observed and class "lake" might not preserve its shape furthermore.

Figure 22 provides the results of binary image classification in the case of class "landslides". The results are different in comparison with class "lake".

a) $\lambda = 1$ b) $\lambda = 3$ c) $\lambda = 5$

d) $\lambda = 10$ e) $\lambda = 20$ f) $\lambda = 50$
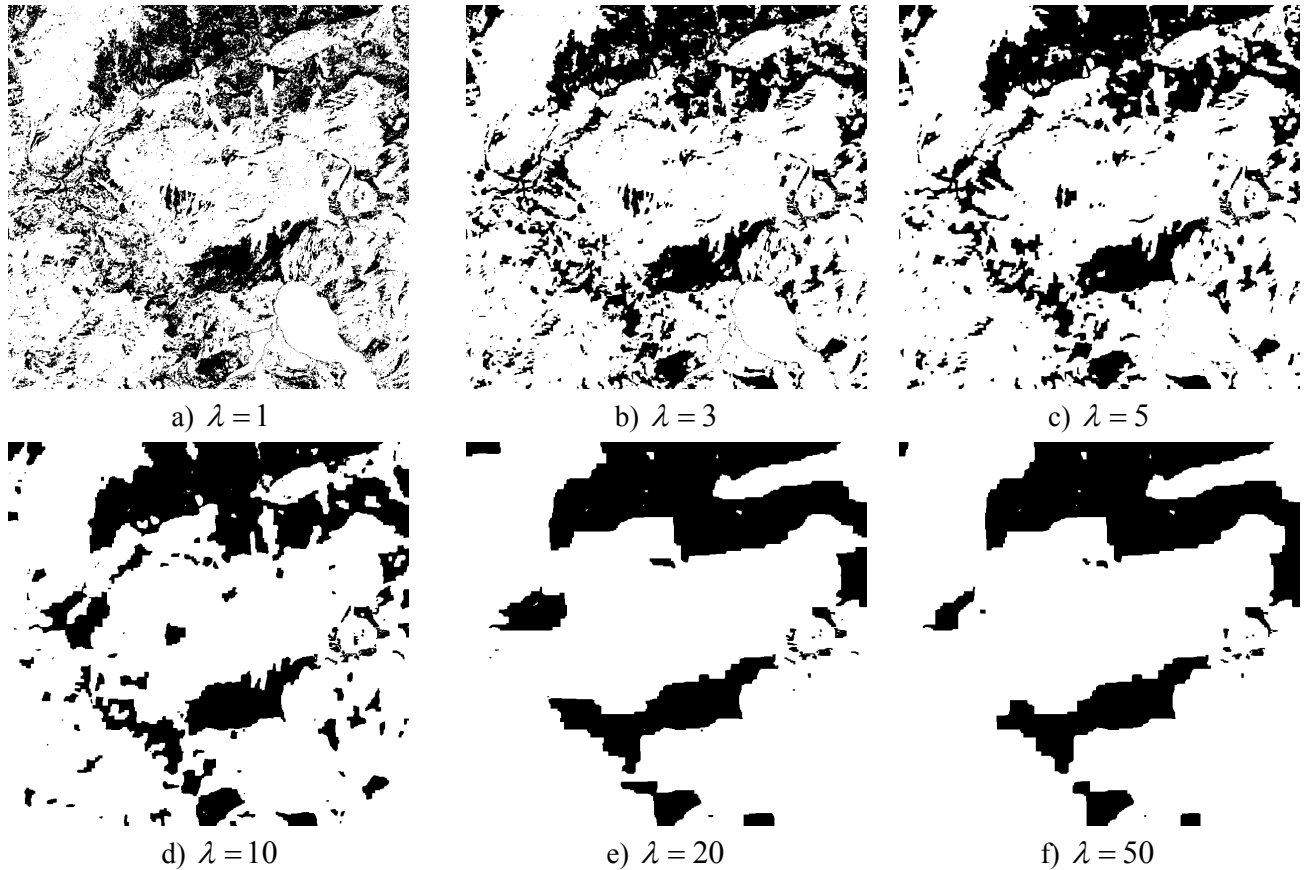
Figure 22. Binary classification results of class "landslides" using swap-move algorithm

From Figure 21 – 22 following can be observed: the smoothness parameters have different impact on both classes. In the case of class "landslides" the smoothness values are very sensitive, where in case of class "lake" they are less sensitive. Over smoothing is observed and the objects on the image do not preserve their shape in case of $\lambda = 50$. The same situation can be observed for other spectral classes

This paragraph compares all applied algorithm based on Aster images visual interpretation. The results of each algorithm are compared with original Aster image. It is not the ideal method to compare the results, because in this case it is not possible to quantify the accuracy but it still gives an idea about the quality of classification results.

Figure 23 shows the classification results of ICM algorithm in comparison with the Aster image. Particularly, the difference in the results can be observed around the class "river". The "river" is shown in *"cyan colour"*. If we compare Figure 23 (a) and Figure 23 (b) then it is observed that some commission errors are visible in 23 (b) where some objects from "river" are wrongly classified. In the case of 23 (a) omission errors for this class are observable, where quite numbers of pixels are missing in this class.

a) method based on *min value (figure 10c)*

b) method based on *inverse order to min value*

c) original Aster image
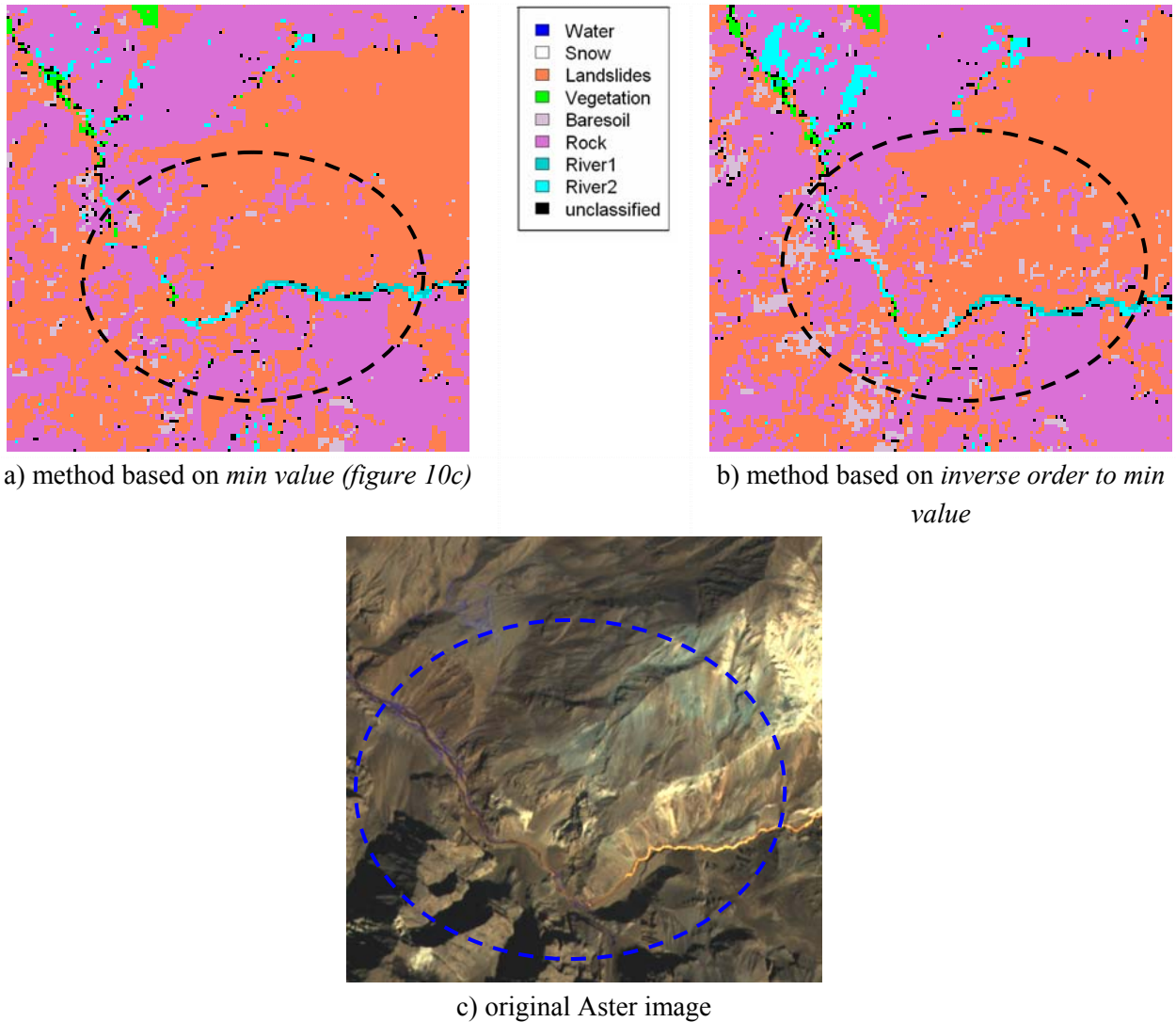
Figure 23. Comparison analysis based on visual interpretation ($\lambda = 1$)

The following figure shows the results in details of both methods based on swap-move algorithm.

a) based on *min value*   b) based on *inverse order to min value*

c) original Aster image
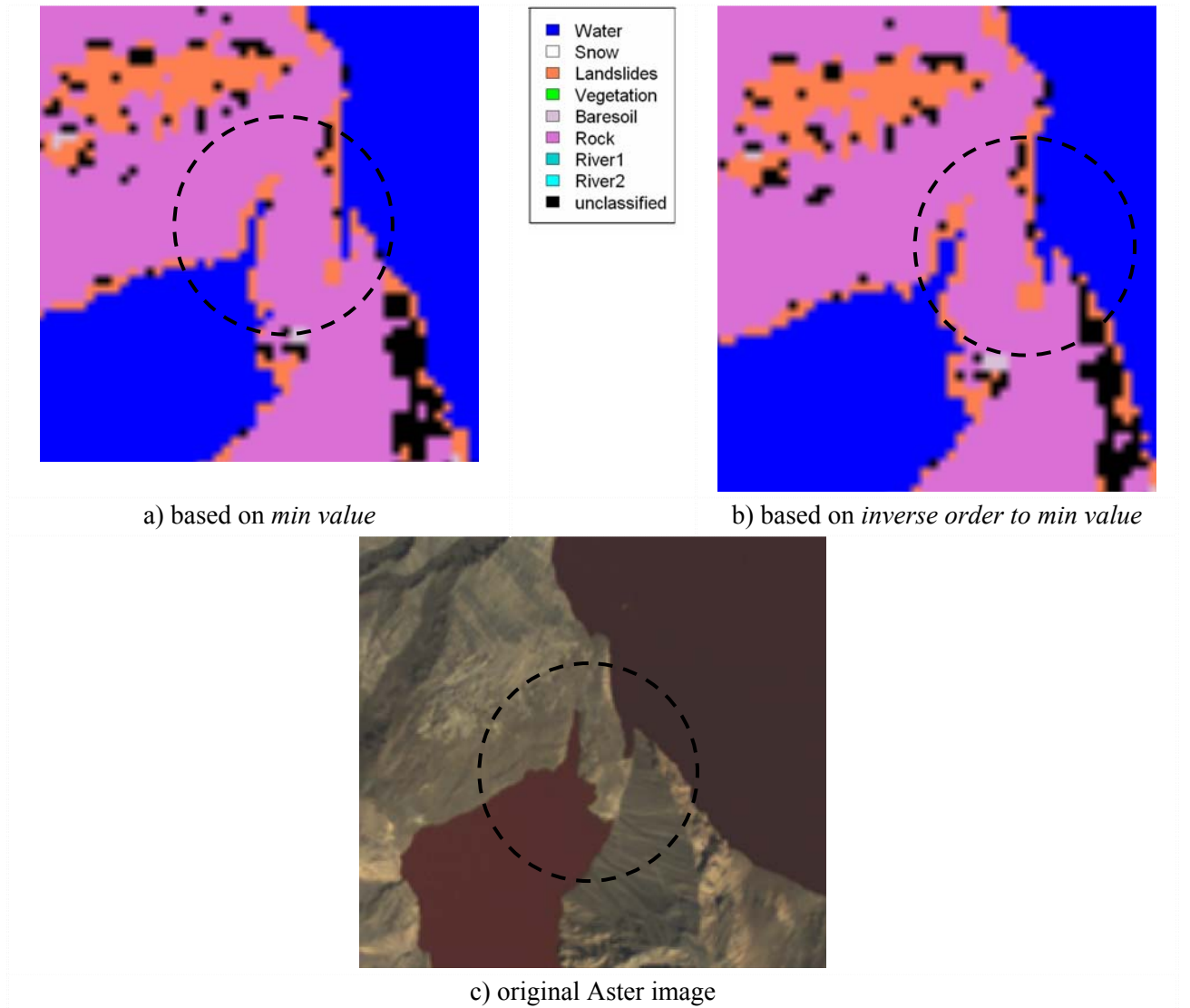
Figure 24. Comparison analysis based on visual interpretation ($\lambda = 1$)

Similarly, the same method is applied to make comparison analysis based on visual interpretation of expansion-move algorithm.

Except visual interpretation of the classification results and compare them with the original Aster image it is important to consider the computation time of applied algorithms, because this is one of the main problems that is addressed in this research. Computation time is proportional to number of interactions that the specific algorithm takes. In addition, energy values are considered, where the least energy should provide better classifi8cation results in according to the theory of energy optimization methods.

Table 6 shows computation time (in seconds) versus the energy of binary classification for all individual classes using the applied energy minimization algorithms.

Algorithm	energy	time
ICM	7671198	0.05
Swap	7671198	0.17
Expansion	7671198	0.12

a) lake

Algorithm	energy	time
ICM	7681090	0.33
Swap	7681054	0.89
Expansion	7681054	0.64

b) vegetation

Algorithm	energy	time
ICM	7874489	0.85
Swap	7836528	0.98
Expansion	7836528	0.73

c) snow

Algorithm	energy	time
ICM	7683173	0.49
Swap	7683019	0.89
Expansion	7683019	0.94

d) river 1

Algorithm	energy	time
ICM	7738470	0.48
Swap	7733128	0.92
Expansion	7733128	0.96

e) landslides

Algorithm	energy	time
ICM	7817700	0.67
Swap	7808464	0.95
Expansion	7808464	0.7

f) baresoil

Algorithm	energy	time
ICM	7720894	0.68
Swap	7712131	0.94
Expansion	7712131	0.99

g) rock

Algorithm	energy	time
ICM	7684220	0.48
Swap	7683815	0.89
Expansion	7683815	0.95

h) river 2

Table 6. Energy versus time (in second) for each class

The computation time comparison is presented in Table 6 for each individual class. Similarly, Table 7 shows the final computation time that individual algorithm took.

Algorithm	time
ICM	4.03
Swap	6.63
Expansion	6.03

Table 7. Computational time (in seconds)

The last step in this stage is to compare the quality of binary image classification based on visual interpretation among all the energy minimization methods considering the smoothness parameters applied above. Figure 25 shows the results of all applied algorithms based on the *min value* classification tree. The results of all energy minimization algorithms are compared with maximum likelihood (MLC) algorithm.
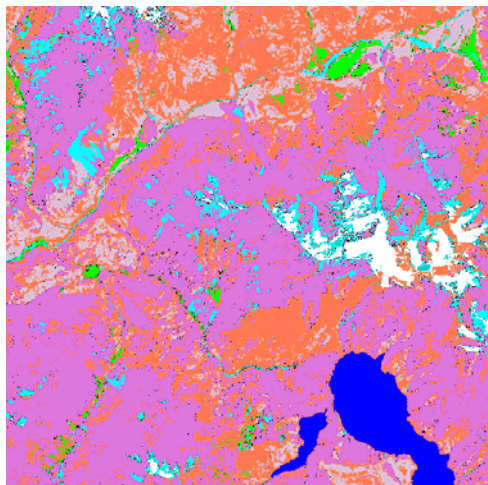
a) ICM algorithm

b) swap-move algorithm

c) expansion-move algorithm

d) Maximum likelihood

e) original Landsat image
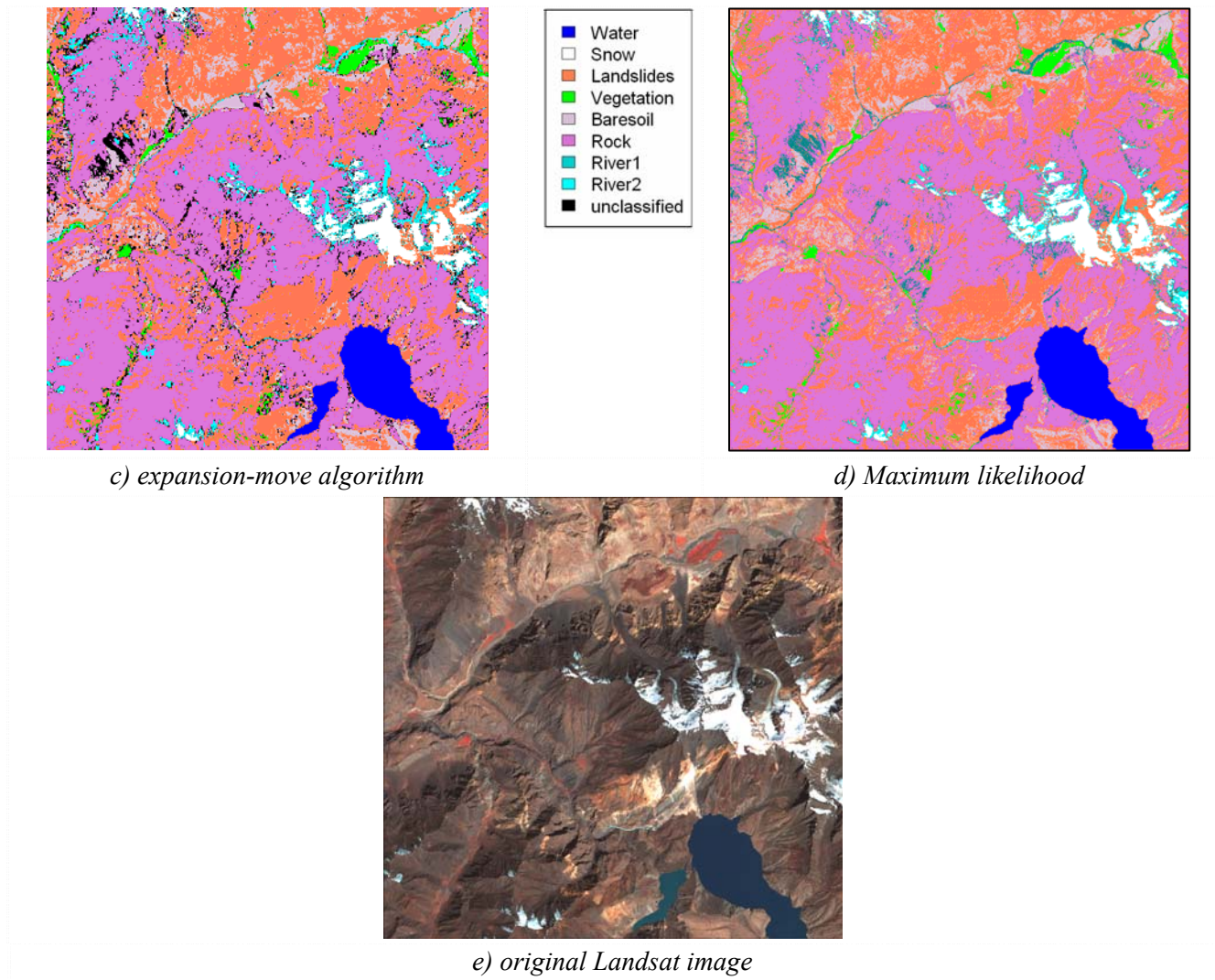
Figure 25. Visual comparison of the classification results based on min value method ($\lambda = 1$)

The results in Figure 25 look different in all cases, MLC produces smooth classification results. ICM result is close to MLC where swap-move and expansion-move algorithms produces classification with more unclassified objects. Expansion-move algorithm produce less unclassified objects than swap-move algorithm. The results are visually compared with the original Landsat image.

5.2.   Classification results of Quickbird image of Boothoven, Enschede

The second part provides the results based on methods applied on Quick Bird image of Boothoven area. The results are al all energy minimization methods are assessed based on visual interpretation, object based analysis and the classification accuracy is quantified using the error matrix. The results of "tree crown" are provided. Different smoothness values are tested and used for image classification. The most appropriate smoothness value that produces better classification result is identified for each energy minimization method. Next section provides some results based on object analysis. Chapter 6 discusses in details the results of each individual algorithm.

5.2.1. Classification results of object based analysis

This section provides the results of binary image classification based on visual and object analyses. The results of expansion-move algorithms are provided with different smoothness values. It is important to mention that the procedure is same for other methods.

Object based analysis is done based on visual interpretation of the classification results. The total number of objects both from classified and reference image that intersect each other are calculated. This comparison is done for every defined smoothness parameter. Discussion on the results is provided.

The classification results of expansion-move algorithm with different smoothness is provided in Figure 26. These comparative analyses are provided using ArcMap software.

a) intersect results, $\lambda = 0.05$ *b) classification results , $\lambda = 0.05$*

c) detailed comparison, $\lambda = 0.05$
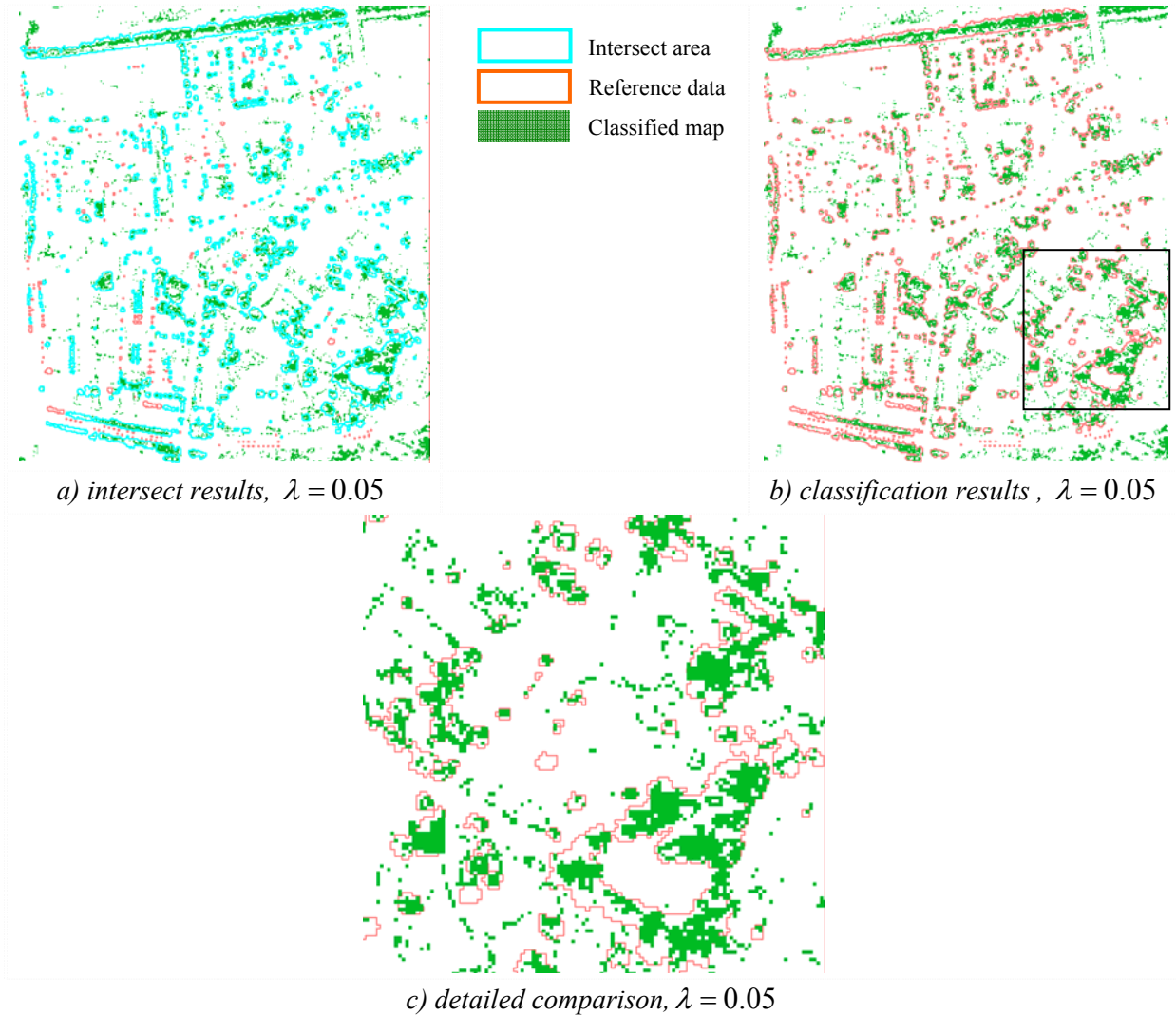
Figure 26. Expansion-move algorithm, classification results using $\lambda = 0.05$.

Figure 26(a) shows intersect results between the reference image and the classified map using ArcMap software, where the total number of objects is identified. The intersect results shows that 458 out of 653 objects intersect with each other, which is about 70.13%. In this particular case the meaning of

intersect is considered as follows: two objects intersect, if at least they have one common pixel. In some cases the objects are very big and in some other cases they are very small, even a size of single pixel. This gives an object based analysis which is normally easy for users. Figure 26 (b) shows the visual comparison of classification results with the reference data. Figure 26 (c) shows more details where the affect of smoothness parameter is observable. In Figure 26 (c) more commission errors are observed as the same time the trees are not fully detected using the applied smoothness value.

Similarly, visual image analysis is provided based on expansion-move algorithm where $\lambda = 1.3$ (Figure 27).

a) intersect results, $\lambda = 1.3$

b) classification results, $\lambda = 1.3$

c) detailed comparison, $\lambda = 1.3$
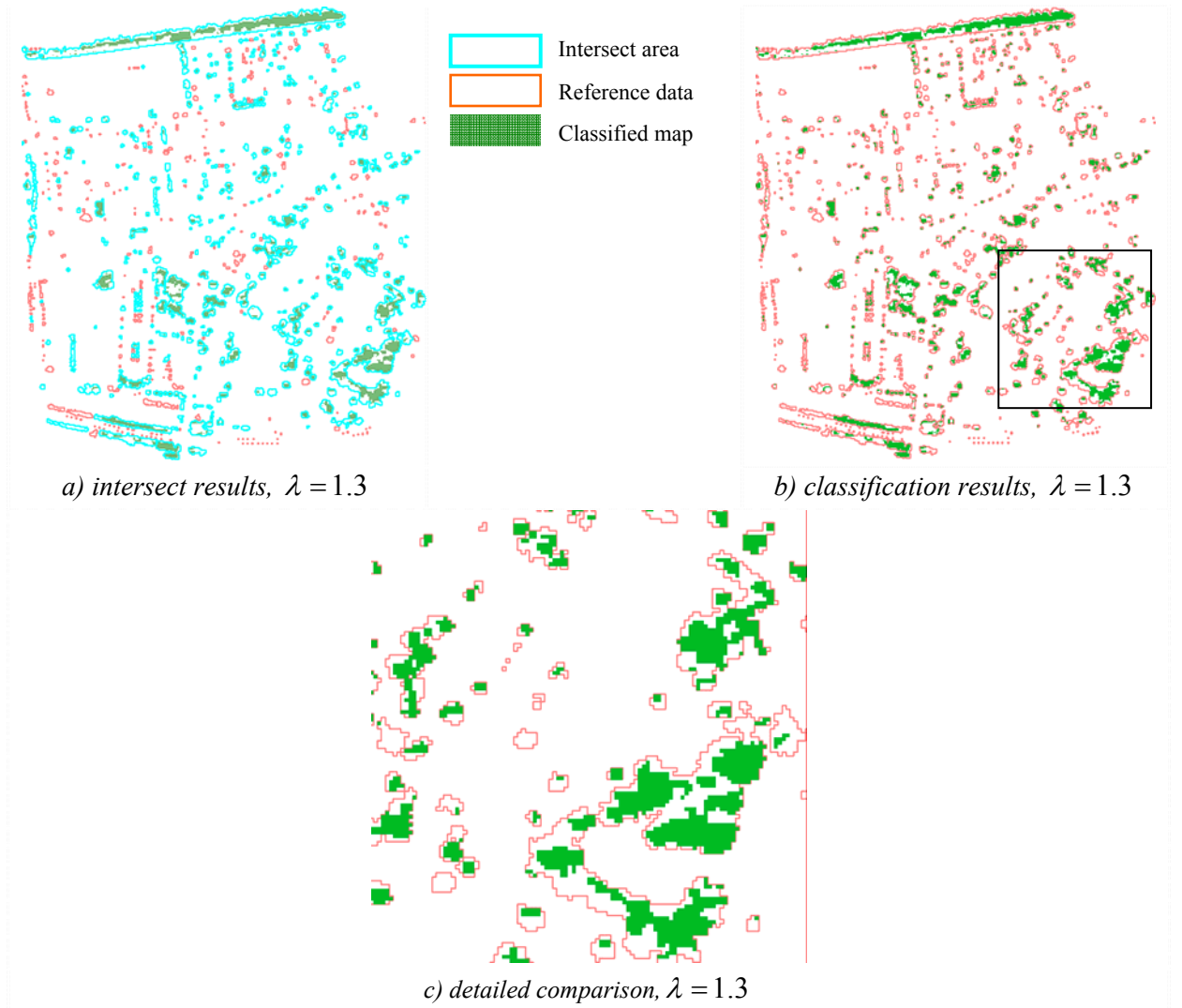
Figure 27. Expansion-move classification results.,

The results produced in Figure 27 are different in comparison with Figure 26. It is observed that by increasing the smoothness value classification results look smoother than with smaller smoothness value, at the same time the number of objects are decreasing. No commission errors are visible using $\lambda = 1.3$.

Legend:
- Intersect area
- Reference data
- Classified map

a) intersect results, $\lambda = 5$

b) classified versus the reference map $\lambda = 5$

c) detailed comparison $\lambda = 5$
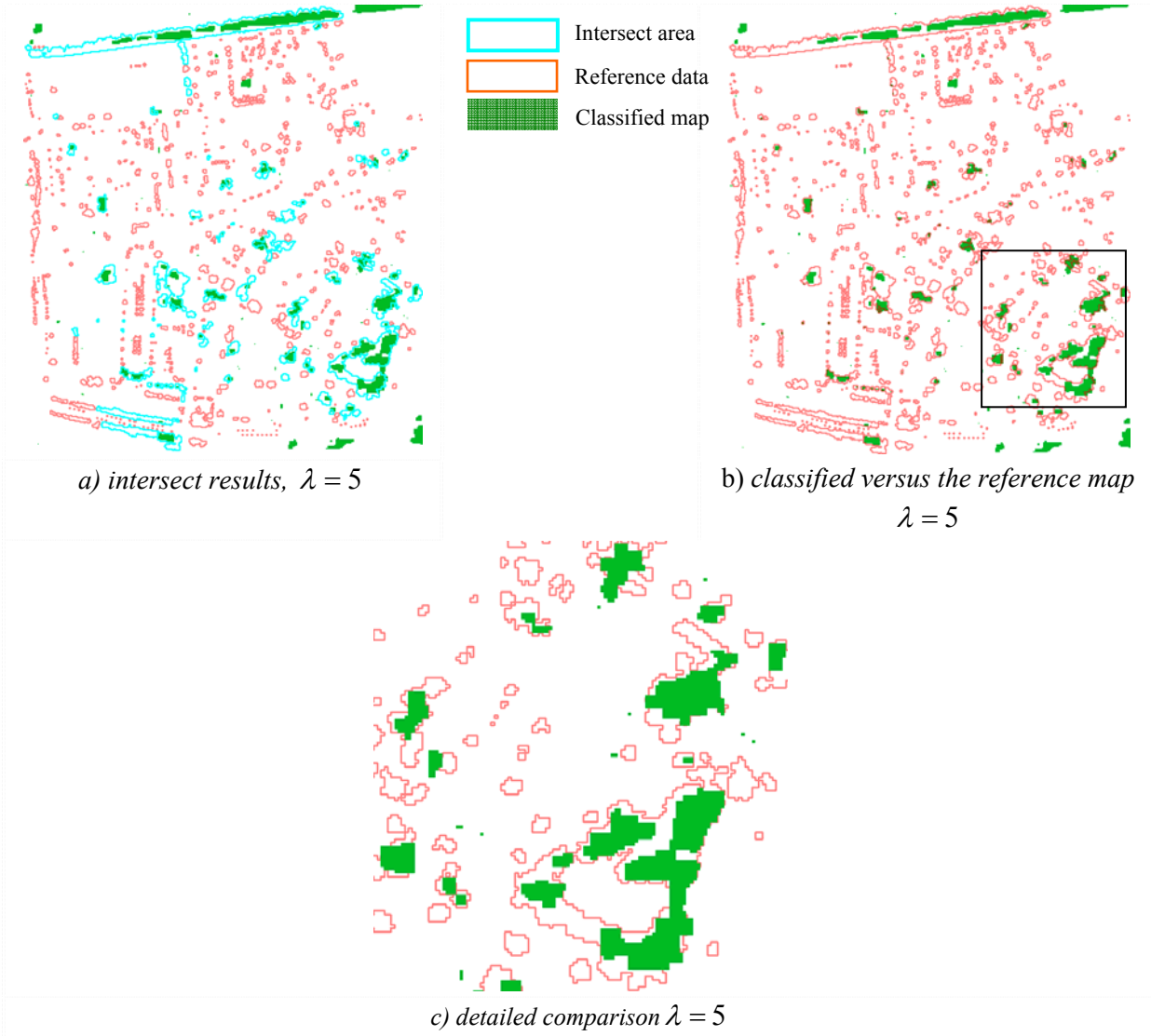
Figure 28. Expansion-move algorithm, classification results using $\lambda = 5$

Similar effect is observed for other methods (SA and ICM) which are not presented in this case. The aim is to show the effect of smoothness value on classification result. It is important to mention that by assigning same smoothness value, all energy minimization methods provide the same number of intersected objects. It can be mentioned that the size of intersect objects are different. In addition, the results of classification accuracy using the error matrix for all methods are provided in the next subsection.

Visual comparison of classification results based on Maximum Likelihood (MLC), Simulated Annealing with logarithmic schedule and expansion-move algorithms are provided. In addition, the results are visually compared with Quickbird images. The comparison analysis is provided based on visual interpretation. Figure 29-30 shows the classification results of MLC , SA with logarithmic schedule and expansion-move algorithms with $\lambda = 5$ are provided. It is observed that expansion-move algorithm provide better classification results in this case. The spatial context is better considered using this algorithm.

a) MLC versus SA Log, $\lambda = 5$

b) MLC versus SA Log, $\lambda = 5$ *(detailed)*
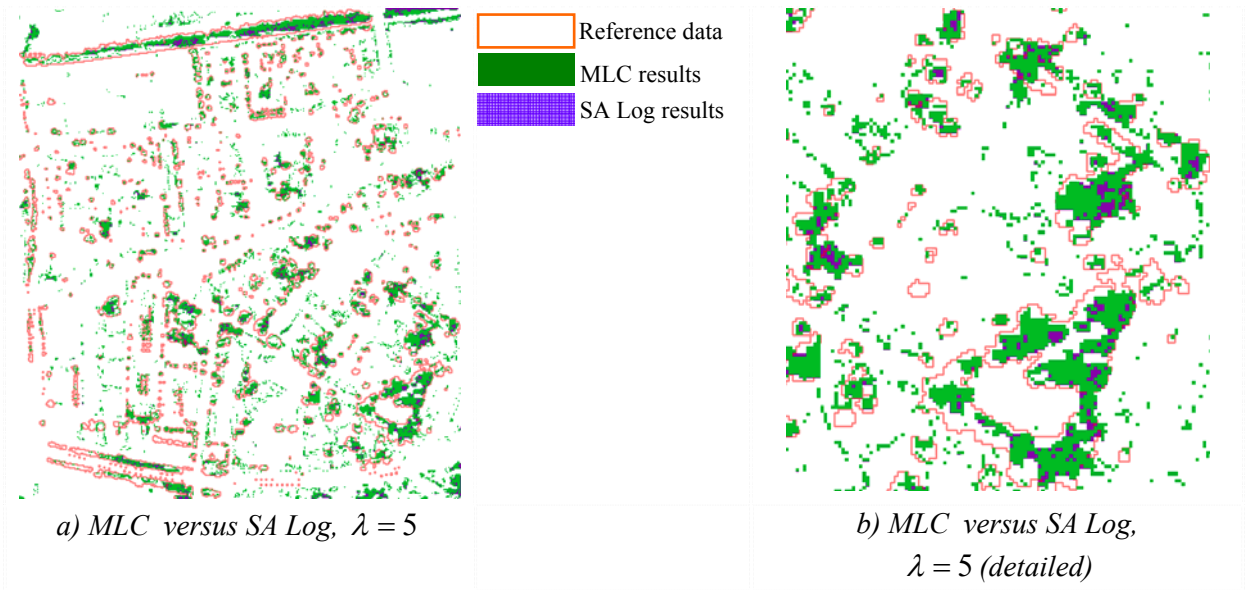
Figure 29. Classification results: Maximum Likelihood versus SA_log, $\lambda = 5$

a) MLC versus Expansion-move, $\lambda = 5$

b) MLC versus Expansion-move, $\lambda = 5$ *(detailed)*
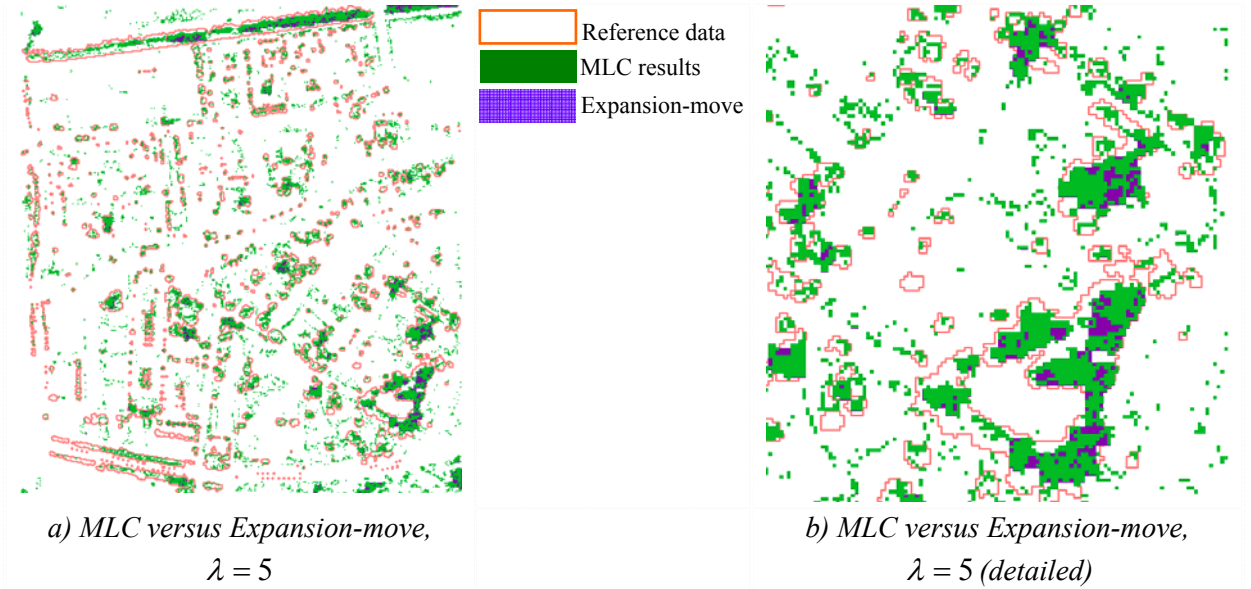
Figure 30. Classification results: Maximum Likelihood versus Expansion-move, $\lambda = 5$

Figure 31 compares the results of MLC classification with three different smoothness values of expansion-move algorithm. In addition, the original image of Buickbird and the trees that classified as shadow are presented. The trees in shadow area are not included in classification, but it is important take into account this problem, because spectral signatures of these trees are different with tree crown. These shadow trees are quite visible in Quickbird image (Figure 31 (f)).

a) MLC result   b) Expansion-move, $\lambda = 0.05$ c) Expansion-move result, $\lambda = 1.3$

d) Expansion-move, $\lambda = 5$ e) Expansion-move result, $\lambda = 1.3$ f) Original Quickbird image

	Reference data
	Classified map
	Trees in shadow
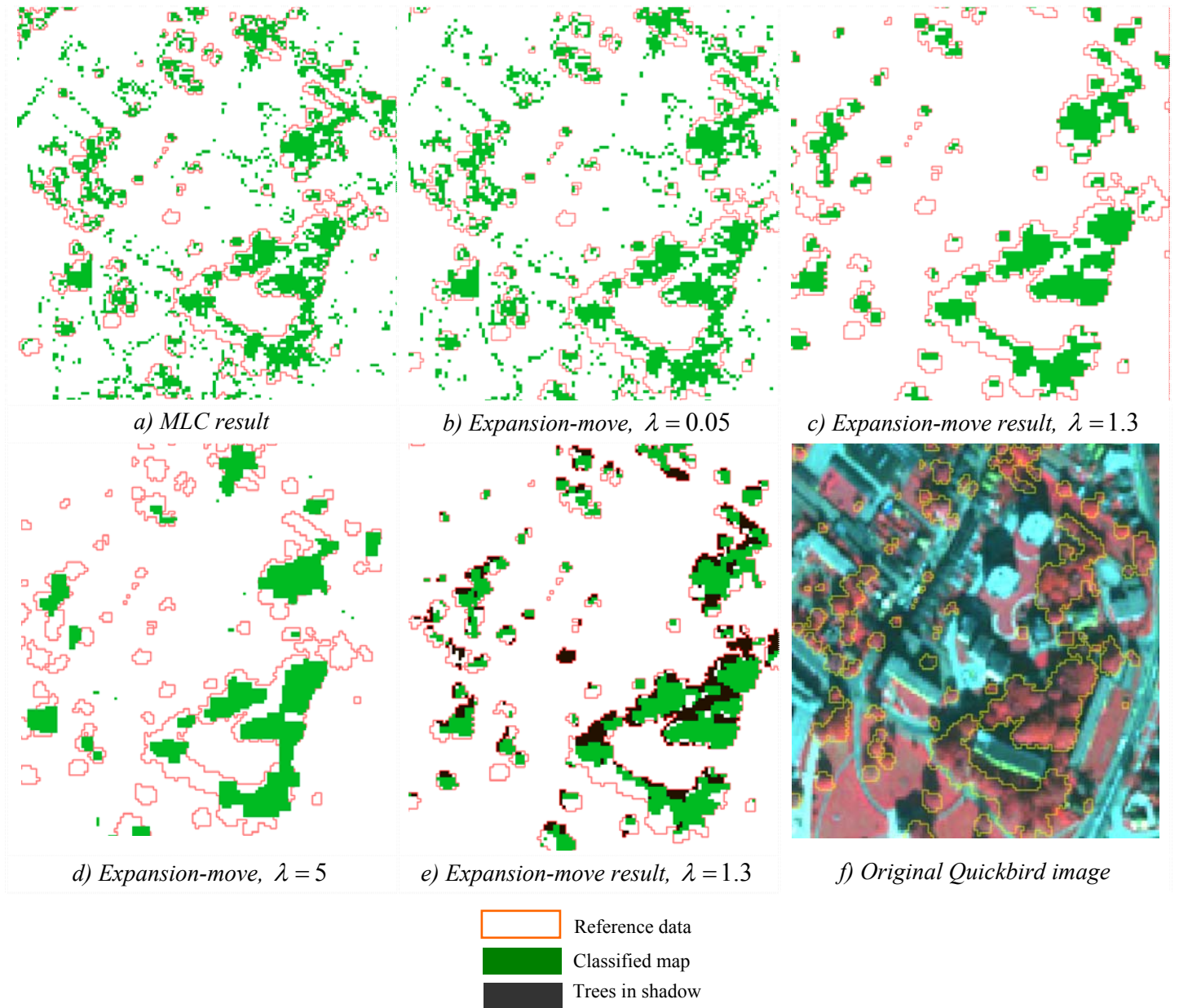
Figure 31. Classification results, MLC versus Expansion-move

From Figure 31 following is observed: the results of expansion-move algorithm (Figure 31 c) where $\lambda = 1.3$ look the best results in this case, because it represents better the tree crown if we compare it with the original Quickbird image (Figure 31 f). In the case of MLC (Figure31a) the classification results look noisy, more commission errors are observed. Figure 31 (e) shows those trees in shadow that are not detected by any algorithms.

5.2.2. Classification accuracy using the error matrix

In this section classification results of all energy minimization methods are provided using the error matrices. Table 8 shows classification results of ICM algorithm that assessed using the error matrix with following smoothness values: $\lambda = 0.05$, $\lambda = 0.5$, $\lambda = 1.3$ and $\lambda = 5$.

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.32	4.51	8.71	53.97	57.68
Non-Tree	57.68	95.49	91.29	92.98	4.51
Total	100	100	100		
Producer's accuracy	42.32	95.49		Overall Acc = 89.6%	
Errors of Comission	46.03	7.02		Kapp = 0.42	

a) ICM classification results ($\lambda = 0.05$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.20	4.23	8.45	55.52	57.80
Non-tree	57.80	95.77	91.55	92.82	4.23
Total	100	100	100		
Producer's accuracy	42.20	95.77		Overall Acc = 89.8 %	
Errors of Comission	44.48	7.02		Kapp = 0.42	

b) ICM classification results ($\lambda = 0.5$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	41.61	3.80	8.00	57.77	58.39
Non-tree	58.39	96.20	92.00	92.95	3.80
Total	100	100	100		
Producer's accuracy	41.61	96.20		Overall Acc = 90.1 %	
Errors of Comission	42.23	7.05		Kapp = 0.43	

c) ICM classification results ($\lambda = 1.3$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	34.19	2.09	5.66	67.16	65.81
Non-tree	65.81	97.91	94.34	92.25	2.09
Total	100	100	100		
Producer's accuracy	34.19	97.91		Overall Acc = 90.8 %	
Errors of Comission	32.84	7.75		Kapp = 0.40	

d) ICM classification results ($\lambda = 5$)

Table 8. ICM Classification results (error matrix)

The results of ICM algorithm are different in all cases. There is almost not difference in case of overall accuracy but in case of user and producer's accuracy some differences are observable. Kappa values are in the range from 0.40 to 0.43, not much difference.

Table 9 -10 show the results of Simulated Annealing algorithm both with Exponential and Logarithmic cooling schedule for the same smoothness values.

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.31	4.51	8.71	53.97	57.68
Non-Tree	57.69	95.49	91.29	92.98	4.51
Total	100	100	100		
Producer's accuracy	42.31	95.49		verall Acc = 89.6 %	
Errors of Comission	46.02	7.02		Kappa = 0.41	

a) SA_Exp classification results ($\lambda = 0.05$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.06	4.21	8.41	55.56	57.94
Non-tree	57.80	95.77	91.55	92.97	4.21
Total	100	100	100		
Producer's accuracy	42.06	95.79		Overall Acc = 89.8 %	
Errors of Comission	44.44	7.03		Kapp = 0.42	

b) SA_Exp classification results ($\lambda = 0.5$)

(Percent)	Ground Truth			User's accuracy	Error of Omission
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	41.39	3.72	7.91	58.15	58.61
Non-tree	58.61	96.28	92.09	92.93	3.72
Total	100	100	100		
Producer's accuracy	41.39	96.28		Overall Acc = 90.2 %	
Errors of Comission	41.85	7.07		Kapp = 0.43	

c) SA_Exp classification results ($\lambda = 1.3$)

(Percent)	Ground Truth			User's accuracy	Error of Omission
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	35.10	2.13	5.79	67.34	64.90
Non-tree	64.90	97.87	94.21	92.34	2.13
Total	100	100	100		
Producer's accuracy	35.10	97.87		l = 90.9 %	
Errors of Comission	32.66	7.66		Kappa = 0.41	

d) SA_Exp classification results ($\lambda = 5$)

Table 9. SA_Exp Classification results (error matrix)

(Percent)	Ground Truth			User's accuracy	Error of Omission
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.34	4.52	8.72	53.96	57.66
Non-Tree	57.66	95.48	91.28	92.98	4.52
Total	100	100	100		
Producer's accuracy	42.34	95.48		Overall = 89.6 %	
Errors of Comission	46.04	7.02		Kappa = 0.42	

a) SA_Log classification results ($\lambda = 0.05$)

(Percent)	Ground Truth			User's accuracy	Error of Omission
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.03	4.21	8.41	55.53	57.97
Non-Tree	57.97	95.79	91.59	92.98	4.21
Total	100	100	100		
Producer's accuracy	42.03	95.79		Overall Acc = 89.8 %	
Errors of Comission	44.47	7.03		Kappa = 0.42	

b) SA_Log classification results ($\lambda = 0.5$)

(Percent)	Ground Truth			User's accuracy	Error of Omission
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	41.55	3.74	7.94	58.17	58.45
Non-tree	58.45	96.26	92.06	92.93	3.72
Total	100	100	100		
Producer's accuracy	41.55	96.26		Overall Acc = 90.2 %	
Errors of Comission	41.83	7.06		Kappa = 0.43	

c) SA_Log classification results ($\lambda = 1.3$)

(Percent)	Ground Truth			User's accuracy	Error of Omission
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	34.72	2.11	5.74	67.28	65.28
Non-tree	65.28	97.89	94.26	92.3	2.11
Total	100	100	100		
Producer's accuracy	34.72	97.89		Overall Acc = 90.9	
Errors of Comission	32.72	7.70		Kappa = 0.41	

d) SA_Log classification results ($\lambda = 5$)

Table 10. SA_Log Classification results (error matrix)

In the case of SA with exponential cooling schedule following parameters are used: TO=3.0, Tupd = 0.9, 0.95 and 0.99. In the case of SA with logarithmic cooling schedule TO=3.0 is used. All these parameters including their energy values are provided in Appendixes (F) and (H). There no much differences in terms of accuracy assessment using the error matrix between the ICM and two SA algorithms.

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.32	4.47	8.67	54.22	57.68
Non-Tree	57.68	95.53	91.33	92.98	4.47
Total	100	100	100		
Producer's accuracy	42.32	95.53		Overall Acc = 89.6 %	
Errors of Comission	45.78	7.02		Kappa = 0.42	

a) Swap-move classification results ($\lambda = 0.05$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	41.77	3.92	8.12	57.16	58.23
Non-tree	58.23	96.08	91.88	92.97	4.21
Total	100	100	100		
Producer's accuracy	41.77	96.08		Overall Acc = 90.0%	
Errors of Comission	42.84	7.04		Kappa = 0.43	

b) Swap-move classification results ($\lambda = 0.5$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	39.70	3.04	7.12	61.99	60.30
Non-tree	60.30	96.96	92.88	92.93	3.72
Total	100	100	100		
Producer's accuracy	39.70	96.96		Overall Acc = 90.6 %	
Errors of Comission	38.01	7.22		Kappa = 0.43	

c) Swap-move classification results ($\lambda = 1.3$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	22.11	0.94	3.29	74.57	77.89
Non-tree	77.89	99.06	96.71	92.93	3.72
Total	100	100	100		
Producer's accuracy	22.11	96.06		Overall Acc = 90.5 %	
Errors of Comission	25.43	8.95		Kappa = 0.30	

d) Swap-move classification results ($\lambda = 5$)

Table 11. Swap-move Classification results (error matrix)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.32	4.47	8.67	54.22	57.68
Non-Tree	57.68	95.53	91.33	92.98	4.47
Total	100	100	100		
Producer's accuracy	42.32	95.53		Overall Acc = 89.6 %	
Errors of Comission	45.78	7.02		Kappa = 0.42	

a) Expansion-move classification results ($\lambda = 0.05$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	41.77	3.92	8.12	57.16	58.23
Non-tree	58.23	96.08	91.88	92.97	4.21
Total	100	100	100		
Producer's accuracy	41.77	96.08		Overall Acc = 90.0 %	
Errors of Comission	42.84	7.04		Kappa = 0.43	

b) Expansion-move classification results ($\lambda = 0.5$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	39.70	3.04	7.12	61.99	60.30
Non-tree	60.30	96.96	92.88	92.93	3.72
Total	100	100	100		
Producer's accuracy	39.70	96.96		Overall Acc = 90.6 %	
Errors of Comission	38.01	7.22		Kappa = 0.43	

c) Expansion-move classification results ($\lambda = 1.3$)

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	22.11	0.94	3.29	74.57	77.89
Non-tree	77.89	99.06	96.71	92.93	3.72
Total	100	100	100		
Producer's accuracy	22.11	96.06		Overall Acc = 90.5 %	
Errors of Comission	25.43	8.95		Kappa = 0.30	

d) Expansion-move classification results ($\lambda = 5$)

Table 12. Expansion-move Classification results (error matrix)

Table 11-12 shows that swap-move and expansion-move algorithms produces identical results. But the results of these algorithms are comparable with SA and ICM algorithms. In terms of user and producer's accuracy the results are different, although alone from the error matrix sometimes it is not

easy to interprate the results. Visual interpretation helps to interprate and conclude the results. The results are discussed in details in next chapter.

Table 13 provides the results of MLC using the error matrix. In comparison with swap-move and expansion-move algorithms MLC results are comparable; the accuracy measures are relatively less than the graph cuts algorithms and even less than both SA algorithms. The comparison of these algorithms based on visual interpretation is provided in the previous section.

(Percent)	Ground Truth				
Classes	Tree	Non-tree	Total	User's accuracy	Error of Omission
Tree	42.38	4.84	9.01	52.25	57.62
Non-tree	57.62	95.16	90.99	92.96	4.84
Total	100	100	100		
Producer's accuracy	42.38	95.16		Overall Acc = 89.2 %	
Errors of Comission	47.75	7.04		Kappa = 0.40	

Table 13. Maximum Likelihood Classification result

Finally, the computation time of proposed methods are provided in table 14. It is important to mention that the computation time is proportional to number of interactions. In other words, every iteration takes approximately the same amount of time. Table 13 shows the performance in terms of iterations.

Method	Number of iterations			
	$\lambda = 0.05$	$\lambda = 0.5$	$\lambda = 1.3$	$\lambda = 5$
ICM	1	2	4	5
SA_Exp	44	41	36	24
SA_Log	1027	1024	1019	845
Swap-move	2	2	2	2
Expansion-move	2	2	2	2

Table 14. Computation time in terms of iterations

Swap-move and expansion-move algorithms show very interesting results, in all cases the number of iterations is equal to 2, which is considered very fast. ICM shows different results, the number of iterations ranging from 1 to 5 depends on smoothness parameter, with very low smoothness parameter it converges very fast where by increasing the smoothness parameter the number of iterations is increased. SA with exponential cooling schedule performs faster than SA with logarithmic schedule, but still both of these algorithms are very slow in comparison to graph cuts and ICM algorithms.

6.   Discussion

The achieved results are discussed in this chapter. The applicability of selected graph based energy minimization algorithms such as swap-move and expansion-move are discussed in detail. Their results are compared with Simulated Annealing (SA) (both Exponential and Logarithmic cooling schedules) and Interacted Conditional Modes (ICM) and Maximum Likelihood (MLC) algorithms. Performance evaluation of the methods both in term of classification accuracy and computational time are discussed as well. The first part of the chapter discusses the results of energy minimization methods based on visual interpretation, where a reference data is not provided. In addition to visual interpretation of the results, the second part discusses an object based assessment as well as the accuracy assessment based on error matrix.

ICM is a fast energy minimization algorithm and for this particular study it shows interesting results. ICM is applied on binary image classification of Sarez image with eight defined spectral classes. It took 5.64s to classify the entire image. In terms of computational time, ICM outperformed the swap-move and expansion-move algorithms but in terms of energy optimization, it is behind the swap-move and expansion-move algorithms. ICM is applied to test the applicability of adopted methods (Figure12) and the results show that the first three proposed methods (Figure 12 a, b, c) achieved identical results. These results show that the methods based on class separability measures produce identical results. But the inverse tree nodes order (Figure 12 d) produced different result. Based on the achieved results it is observed that the classification results depend on the order of tree nodes. Based on visual interpretation of the results it is observed that a method based on min value produces results with less commission errors. The results are compared with original Aster image (Figure 23). It is important to mention that the classification results depend on smoothness value which is assigned to a particular spectral class. In addition, it is observed that ICM produces results with less unclassified objects than swap-move and expansion-move, which is related to selecting of smoothness value. The comparison study has not carried out in details for Landsat image, because of the limitation of visual interpretation and a lack of the reference data.

Similarly to ICM, swap-move and expansion-move algorithms are applied on binary image classification of Sarez image. These algorithms prove that the defined classification trees leads to different results. Besides binary image classification, these algorithms are applied to find out the optimal smoothness value for each individual spectral cover class (Figures 21 - 22). The optimal range of smoothness values are tested and observed using the swap-move and expansion-move algorithms. Two spectral classes are selected and tested. Class "lake" (Figure 21) represents most separated spectral class, which means that it is easier to differentiate from other classes, while class "landslides" (Figure 22) belongs to the least separated classes. This is concluded based on class separability measures. Figures 21-22 show that each smoothness value has a different impact on these two classes. From Figures 21-22 it is observed that the smoothness parameters affect less the class "lake" rather than class "landslides", in other words, the smoothness values are more sensitive to class "landslides" and less sensitive to class "lake". The objects in the Figure 22 change their shape when smoothness value is increased and in the case of $\lambda = 50$ over smoothness is observed for class "landslides". In the case of class "lake", it preserves well its shape, although some minor changes are observed in the edges of class "lake". This gives sufficient reason to consider the smoothness values for individual spectral class.

In terms of computational the swap-move and expansion-move algorithms show interesting results. In comparison to ICM algorithm, the swap-move and expansion-move algorithms are somewhat slower (Table 6 and 7) but the energy is lower in the majority of cases. The results of swap-move and expansion-move algorithms are similar in terms of energy, because they are also similar in their structure. In terms of computation time, the expansion-move algorithm shows slightly improved results (Table 7). Figure 25 shows the differences between the methods based on visual interpretation of classification results. The classification results show that more unclassified objects are observed in case of swap-move and expansion-move algorithms in comparison to ICM. These unclassified objects appear in transition zones between different spectral classes, which can be related to quality of collected training set. Among the graph cuts algorithms, expansion-move provides less unclassified objects than the swap-move algorithm (Figure 25 b, c). In addition to visual interpretation, an overlap analysis is carried out that to find out whether a single pixel is classified in more than one class. The results showed that there is no overlap in final classification result (Figure 14).

This part of the chapter discusses the results (in case of Quickbird image) of all energy minimization algorithms and compare their performance evaluation in terms of visual image interpretation, object based analysis and the error matrices. The results of swap-move and expansion-move, MLC, ICM and SA with both cooling schedules algorithms are discussed and compared. The results based on error matrices for individual method with different smoothness values are discussed.

The results in Figures 26-27 show how the smoothness values effect classification results, based on visual interpretation. Using the smallest smoothness value ($\lambda = 0.05$) the classification results look noisy, commission errors are observed but at the same time most of the small individual trees are detected. Using $\lambda = 0.05$, 458 out of 653 tree objects are detected; in case of $\lambda = 1.3$, 332 out of 653 tree objects; and in case of $\lambda = 5$ only 72 out of 653 tree objects are detected. The selected range of smoothness values shows that they lead to different classification results and to choose the most optimal smoothness parameter depends upon specific purposes. If we are interested in small individual trees then smaller smoothness value can be used. But the disadvantage of choosing small smoothness value is that the results look noisy, because of commission errors. In the case of large trees the smoothness value should be increased. It is observed that in case of large trees commission errors are not observed. Within the range of defined smoothness values, $\lambda = 1.3$ shows better classification results. But it should be mentioned that some small trees are not detected using this smoothness value. In case of $\lambda = 5$ only the large groups of tree are well represented, the small and medium sizes of trees are not detected. It is also observed that due to similar spectral properties of class "tree" and "grasslands" some trees are not fully detected.

Figures 29-31 shows the results of Maximum Likelihood (MLC), expansion-move and Simulated Annealing with logarithmic schedule algorithms with different smoothness values. The results of MLC classification look noisy. The results of MLC is similar to expansion-move algorithm with $\lambda = 0.05$. In case of $\lambda = 1.3$ the results based on expansion-move algorithm look best for this particular case. If we assign $\lambda = 0$ for expansion-move algorithm that means that it is similar to MLC, because the context is ignored. The results of expansion-move and SA with logarithmic schedule algorithms are similar, although in Figure 30 expansion-move shows more complete results in comparison to SA with logarithmic schedule. Increasing the smoothness value further will lead to removing the objects from an image, or in other words over-smoothness can be observed.

Tables 8-13 show the classification results based on the error matrix using different smoothness parameters. The first three Tables 8-10 show the results based on ICM and SA algorithms and the next three 11-13 show the results of MLC and swap-move and expansion-move algorithms. The results are different in all four cases. There is no significant difference in terms of overall accuracy and kappa coefficient among all results. Some difference is observed in terms of user and producer's accuracy in all case. SA with exponential is considered faster version of the original SA algorithm, because of its faster cooling schedule, which is the advantage of this algorithm but in terms of energy SA with logarithmic schedule produces lower energy values (Appendixes F - G). But once you can not rely on only lower energy values. In practice lower energy does not always produces better classification results. The results in this research of both cooling schedules do not differ much. Some improvement in terms of energy is observed using the logarithmic schedule. From the Tables 10-12 it is observed that SA with logarithmic schedule shows slightly improved results in terms of accuracy over ICM and SA with exponential cooling schedule.

The swap-move and expansion-move algorithms produced identical results in this particular study (Table 11, 12). Their results in terms of classification accuracy are similar to SA with logarithmic schedule. In comparison with MLC, swap-move and expansion-move algorithms produces better classification accuracy, mainly because these algorithms consider spatial context in an image.

The classification results are often depends on the quality of reference data. In the case of tree crown class some trees are covered by shadow in the original Quick Bird image, which means that spectrally trees in shadow are different from other trees. Therefore, classification results are affected by these shadows. Figure 31 (e) shows in details this situation, the dark area is the shadows and the green is the tree crown. This problem can be addressed by adding another node to the classification tree, which represents another spectral class (shadow). Another factor that affects the classification result is the existence of small tree in the reference data. Due to the limitation of spatial resolution of Quick Bird images these small trees are not detected. Considering these two factors, the classification results can be improved.

Table 14 shows the performance in terms of computational time of all energy minimization algorithms. This performance is considered in terms of iterations. It can be mentioned that each iteration take approximately same amount of time, that's why we can say that iteration is proportional to time. Table 14 shows that ICM took a few numbers of iterations. In the case of SA with exponential cooling schedule it took within the range of (24 – 44 iterations) for each smoothness value different results: and for SA with logarithmic schedule this range is from 845 - 1027 iterations. SA with logarithmic schedule shows most expensive computation in this case, because of its slow cooling schedule. Swap-move and expansion-move algorithms show interesting results. In all cases the number of iterations is equal to 2, which is quite stable and fast. This shows that swap-move and expansion-move algorithms converge very fast.

7. Conclusion and Recommendation

7.1. Conclusion

Optimization of energy is a challenging issue in MRF based remote sensing image analysis which is addressed in this research. Before making conclusion it is important mention that all the formulated research questions are properly answered. The results show that the graph based energy optimization methods such as swap-move and expansion-move are applicable in MRF based remote sensing image classification. They are applied both on coarse (Landsat TM) and fine (Quickbird) spatial resolution images and produced sufficient classification results.

In addition, two methods are proposed and tested; construction of classification tree and optimizing the smoothness parameter value. Both of these methods show their impact on classification result. In the case of classification tree, the results show their dependencies on the tree nodes sequence (order). It is concluded that the nodes sequence matters, which is linked to class separability measures. In the case of smoothness values it is concluded that they are more sensitive to spectral classes that are least separated and less sensitive to those that are most separated, based on divergence measures.

The advantages of the adopted methods are following: They show very fast performance in terms of computational time, as within a short period of time they are able to classify an entire image. Their classification results are similar to that of SA with a logarithmic schedule algorithm and they outperformed the MLC algorithm. The main advantage of swap-move and expansion-move algorithms over SA algorithms is a reduced computational time. In comparison to ICM, the computational time is comparable.

The limitations of these methods in extracting tree crown are following: some trees are not fully detected, because of similarity in the spectral properties between the grassland and tree crown classes and similarly some other are not detected in the transition zones between spectral classes.

It is remarkable to conclude that the swap-move and expansion-move algorithms are applicable in MRF based remote sensing image classification with the type of energy function that is considered for binary image classification. These algorithms performed well and produced sufficient classification results. The proposed methods for constructing classification tree and optimising smoothness value showed their impacts on classification results. Finally, the computational problem of energy minimization in MRF based remote sensing image classification is addressed in this research. At the same time, limitation is observed that mainly caused due to the spectral similarity between the classes.

7.2. Recommendation

In order to address the limitations following is recommended. The performance in terms of accuracy of the energy minimization algorithms can be improved by introducing higher order of neighbourhood system. In this case only the first order neighbourhood system is used, second order or higher order of neighbourhood system might lead to better classification results. Furthermore, the problem in the transition zones between the classes needs to be addressed in future.

This is one of the first experiments on testing the applicability the swap-move and expansion-move algorithms in MRF based remote sensing image classification and to study the applicability of these algorithms for other application in remote sensing such as de-noising and texture analysis would be an interesting and useful work.

References

1.   Kirkpatrick, S., *Optimization by simulated annealing: Quantitative studies.* Journal of Statistical Physics, 1984. **34**(5): p. 975-986.

2.   Crama, Y. and M. Schyns, *Simulated annealing for complex portfolio selection problems.* European Journal of Operational Research, 2003. **150**(3): p. 546-571.

3.   Ingber, L., *Simulated Annealing - Practice Versus Theory.* Mathematical and Computer Modelling, 1993. **18**(11): p. 29-57.

4.   Boykov, Y. and G. Funka-Lea, *Graph cuts and efficient N-D image segmentation.* International Journal of Computer Vision, 2006. **70**(2): p. 109-131.

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

6.   Kolmogorov, V. and R. Zabih, *What energy functions can be minimized via graph cuts?* Computer Vision - Eccv 2002 Pt Iii, 2002. **2352**: p. 65-81.

7.   Tso, B. and M. Mather, *Classification methods for remotely sensed data,.* Book: published by Taylor and Francis Ltd (London) and Taylor and Francis Inc (New York), 2001. **332**.

8.   Besag, J., *Spatial Interaction and the Statistical Analysis of Lattice Systems.* Journal of the Royal Statistical Society Series B (Methodological), 1974. **36**(2): p. 192-236.

9.   Geman, S. and D. Geman, *Stochastic relaxation Gibbs distribution and the Bayesian restoration of the image.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984. **PAMI-9(1)**(1): p. 39-55.

10.   Solberg, A.H.S., T. Taxt, and A.K. Jain, *A Markov random field model for classification of multisource satellite imagery.* IEEE Transactions on Geoscience and Remote Sensing, 1996. **34**(1): p. 100-113.

11.   Melgani, F. and S.B. Serpico, *A Markov random field approach to spatio-temporal contextual image classification.* IEEE Transactions on Geoscience and Remote Sensing, 2003. **41**(11): p. 2478-2487.

12.   Tso, B. and R.C. Olsen, *A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process.* Remote Sensing of Environment, 2005. **97**(1): p. 127-136.

13.   Amador, J.J., *Markov random field approach to region extraction using Tabu Search.* Journal of Visual Communication and Image Representation, 2005. **16**(2): p. 134-158.

14.   Kim, H.J., E.Y. Kim, J.W. Kim, and S.H. Park, *MRF model based image segmentation using hierarchical distributed genetic algorithm.* Electronics Letters, 1998. **34**(25): p. 2394-2395.

15.   Arques, P., P. Compan, R. Molina, M. Pujol, and R. Rizo, *Minimization of an energy function with robust features for image segmentation.* Kybernetes, 2003. **32**(9-10): p. 1481-1491.

16.   Barker, S.A. and P.J.W. Rayner, *Unsupervised image segmentation using Markov random field models.* Pattern Recognition, 2000. **33**(4): p. 587-602.

17.   Hu, R.M. and M.M. Fahmy, *Texture Segmentation Based on a Hierarchical Markov Random Field Model.* Signal Processing, 1992. **26**(3): p. 285-305.

18.   Xu, X., D.R. Li, and H. Sun, *Multiscale MRF-based texture segmentation of SAR image.* Chinese Journal of Electronics, 2004. **13**(4): p. 671-675.

19.   Li, M., Y. Wu, and Q. Zhang, *SAR image segmentation based on mixture context and wavelet hidden-class-label Markov random field.* Computers & Mathematics with Applications, 2009. **57**(6): p. 961-969.

20.   Sarkar, A., M.K. Biswas, B. Kartikeyan, V. Kumar, K.L. Majumder, and D.K. Pal, *A MRF model-based segmentation approach to classification for multispectral imagery.* Ieee Transactions on Geoscience and Remote Sensing, 2002. **40**(5): p. 1102-1113.

21.   Li, F., J.X. Peng, and X.J. Zheng, *Object-based and semantic image segmentation using MRF.* Eurasip Journal on Applied Signal Processing, 2004. **2004**(6): p. 833-840.

22.   Aiazzi, B., L. Alparone, S. Baronti, G. Cuozzo, C. D'Elia, and G. Schirinzi, *SAR image classification via Tree-Structured Markov Random Field and Information-Theoretic Heterogenity features.* Inst. of Applied Physics "Nello Carrara", IFAC-CNR, Via Madonna del Piano, 50019 Sesto Fiorentino (Italy). ( ).

23.   Gaetano, R., G. Poggi, and G. Scarpa, *Multispectral image classification automatic building of tree-structured MRF models* Dipartimento di Ingegneria Elettronica e delle

Telecomunicazioni, Universit`a Federico II di Napoli – via Claudio 21, 80125 Napoli, Italy, 2006. ( ): p. .

24. Poggi, G. and A.R.P. Ragozini, *Image segmentation by tree-structured Markov random fields.* IEEE Signal Processing Letters, 1999. **6**(7): p. 155-157.

25. D'Elia, C., G. Poggi, and G. Scarpa, *A tree-structured Markov random field model for Bayesian image segmentation.* IEEE Transactions on Image Processing, 2003. **12**(10): p. 1259-1273.

26. Poggi, G., G. Scarpa, and J.B. Zerubia, *Supervised segmentation of remote sensing images based on a tree-structured MRF model.* IEEE Transactions on Geoscience and Remote Sensing, 2005. **43**(8): p. 1901-1911.

27. Besag, J., *On the Statistical-Analysis of Dirty Pictures.* Journal of the Royal Statistical Society Series B-Methodological, 1986. **48**(3): p. 259-302.

28. Park, J. and L. Kurz, *Image enhancement using the modified ICM method.* IEEE Transactions on Image Processing, 1996. **5**(5): p. 765-771.

29. Magnussen, S., P. Boudewyn, and M. Wulder, *Contextual classification of Landsat TM images to forest inventory cover types.* International Journal of Remote Sensing, 2004. **25**(12): p. 2421-2440.

30. Dunmur, A.P. and D.M. Titterington, *Mean fields and two-dimensional Markov random fields in image analysis.* Pattern Analysis and Applications, 1998. **1**(4): p. 248-260.

31. Fwu, J.K. and P.M. Djuric, *Unsupervised vector image segmentation by a tree structure - ICM algorithm.* IEEE Transactions on Medical Imaging, 1996. **15**(6): p. 871-880.

32. Gray, A.J., J.W. Kay, and D.M. Titterington, *On the Estimation of Noisy Binary Markov Random-Fields.* Pattern Recognition, 1992. **25**(7): p. 749-768.

33. Modestino, J.W. and J. Zhang, *A Markov Random Field Model-Based Approach to Image Interpretation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992. **14**(6): p. 606-615.

34. Lamotte, J.L. and R. Alt, *Comparison of Simulated Annealing Algorithms for Image-Restoration.* Mathematics and Computers in Simulation, 1994. **37**(1): p. 1-15.

35. Nasab, N.M., M. Analoui, and E.J. Delp, *Robust and efficient image segmentation approaches using Markov random field models.* Journal of Electronic Imaging, 2003. **12**(1): p. 50-58.

36. Hurn, M. and C. Jennison, *A Study of Simulated Annealing and a Revised Cascade Algorithm for Image-Reconstruction.* Statistics and Computing, 1995. **5**(3): p. 175-190.

37. Lempitsky, V., C. Rother, and A. Blake, *LogCut - Efficient Graph Cut Optimization for Markov Random Fields.* International Conference on Computer Vision - ICCV 2007.

38. Kohli, P. and P.H.S. Torr, *Dynamic graph cuts for efficient inference in Markov random fields.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007. **29**(12): p. 2079-2088.

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

40. Boykov, Y. and M.P. Jolly, *Interactive organ segmentation using graph cuts.* Medical Image Computing and Computer-Assisted Intervention - Miccai 2000, 2000. **1935**: p. 276-286.

41. Greig, D.M., B.T. Porteous, and A.H. Seheult, *Exact Maximum a-Posteriori Estimation for Binary Images.* Journal of the Royal Statistical Society Series B-Methodological, 1989. **51**(2): p. 271-279.

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

43. Wainwright, M.J., T.S. Jaakkola, and A.S. Willsky, *MAP estimation via agreement on trees: Message-passing and linear programming.* IEEE Transactions on Information Theory, 2005. **51**(11): p. 3697-3717.

44. Li, S.Z., *Markov Random Field Moddeling in Image Analysis.* Vol. XIX. 2001: Springer. 323pp.

45. Varshney, P.K. and M.K. Arora, *Advanced Image processing Technique for Remotely Sensed Hyperspectral Data.* 2003, Germany: Springer. 322.

46. Kasetkasem, T., M.K. Arora, and P.K. Varshney, *Super-resolution land cover mapping using a Markov random field based approach.* Remote Sensing of Environment, 2005. **96**(3-4): p. 302-314.

47. Dutta, A., *Fuzzy c-Means Classification of Multispectral Data Incorporating Spatial Contextual Information by using Markov Random Field.* MSc thesis, 2009.

48. Lu, C.L., S.M. Pizer, and S. Joshi, *A Markov random field approach to multi-scale shape analysis.* Scale Space Methods in Computer Vision, Proceedings, 2003. **2695**: p. 416-431.

49. Li, S.Z., H. Wang, K.L. Chan, and M. Petrou, *Minimization of MRF energy with relaxation labeling.* Journal of Mathematical Imaging and Vision, 1997. **7**(2): p. 149-161.

50. Boykov, Y. and D. Huttenlocher, *A New Bayesian Framework for Object Recognition.* Computer Vision and Pattern Recognition, 1999. **2**: p. 517.

51. Ishikawa, H., *Exact optimization for Markov random fields with convex priors.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. **25**(10): p. 1333-1336.

52. Boykov, Y., O. Veksler, and R. Zabih, *Markov Random Fields with Efficient Approximations.* 1997.

53. Roy, S. and V. Govindu, *MRF Solutions for Probabilistic Optical Flow Formulations.* 15th International Conference on Pattern Recognition (ICPR'00), 2008. **3**.

54. Boykov, Y. and M.P. Jolly, *Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images.* Proceedings of "Internation Conference on Computer Vision", Vancouver, Canada, July 2001, 2001. **1**: p. 105.

55. Ford, L. and D. Fulkerson, *Flows in Networks*. 1962: Princeton University Press. .

56. Wu, C.H. and P.C. Doerschuk, *Tree Approximations to Markov Random-Fields.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995. **17**(4): p. 391-402.

57. Liu, G.Y., Q.Q. Qin, T.C. Mei, W. Xie, and L.G. Wang, *Supervised Image Segmentation Based on Tree-Structured MRF Model in Wavelet Domain.* IEEE Geoscience and Remote Sensing Letters, 2009. **6**(4): p. 850-854.

58. Tolpekin, V.A. and A. Stein, *Quantification of the Effects of Land-Cover-Class Spectral Separability on the Accuracy of Markov-Random-Field-Based Superresolution Mapping.* IEEE Transactions on Geoscience and Remote Sensing, 2009. **47**(9): p. 3283-3297.

Appendix A – Training set of Study area - 1
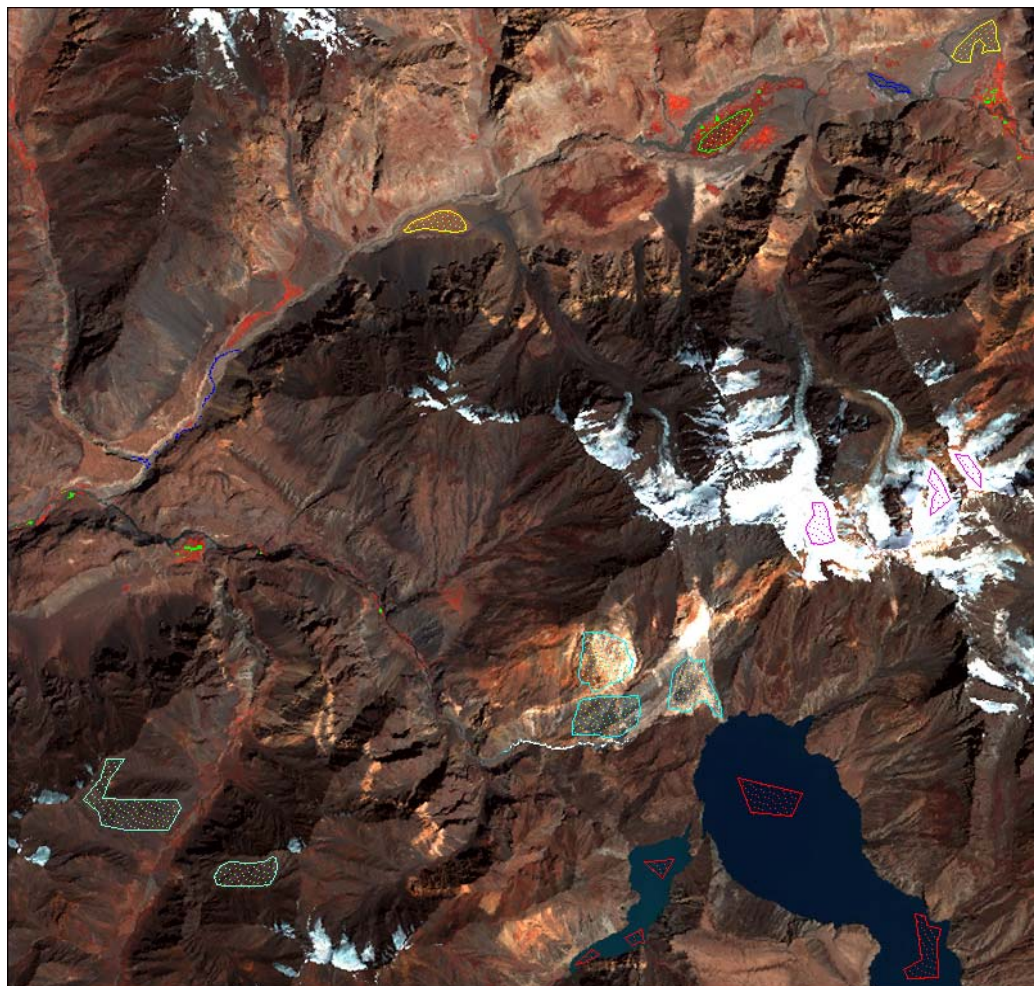
Training set for Landsat TM image, Sarez Lake.

Appendix B – Defined spectral classes for study area - 1

Class name	Color	Number of pixels	Number of polygons	Number of points
Lake	blue	2030	5	0
Snow	black	968	3	0
Landslides	cyan	4036	3	0
Vegetation	green	806	1	184
Baresoil	magenta	1091	2	0
Rock	red	2647	2	0
River 1	coral	232	1	10
River 2	sienna	91	0	91

Appendix C – Mean and covarience values for Study area - 1

Mean values of spectral classes (classes that are defined classes in Appendix B)

Mean_1	Mean_2	Mean_3	Mean_4	Mean_5	Mean_6	Mean_7	Mean_8
55.554680	219.893595	82.796829	50.069479	70.920257	52.993578	66.297414	142.879121
35.098520	206.995868	86.300793	44.578164	68.151237	45.506611	61.172414	123.186813
18.776840	229.503099	106.745045	48.832506	81.477544	49.723461	65.952586	119.945055
9.950246	146.450413	79.859267	73.740695	61.905591	35.94031	41.405172	57.406593
7.407882	16.634298	96.758672	71.637717	74.289643	36.706838	22.077586	26.835165
7.276355	14.103306	82.634787	49.584367	69.558203	33.701171	20.512931	23.868132

Covariance table for class Lake

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
8.997255	13.367208	6.220931	0.843284	0.825394	0.708637
13.367208	25.155395	11.136831	1.524372	1.579312	1.259107
6.220931	11.136831	6.555895	0.70698	0.747546	0.59891
0.843284	1.524372	0.70698	0.685054	0.097189	0.097049
0.825394	1.579312	0.747546	0.097189	0.69703	0.078452
0.708637	1.259107	0.59891	0.097049	0.078452	1.129603

Covariance table for class Snow

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
705.997973	612.772052	504.996711	324.466797	58.077903	40.603558
612.772052	581.398121	483.409527	343.627509	54.514516	38.626074
504.996711	483.409527	449.263693	293.999637	36.328953	25.243733
324.466797	343.627509	293.999637	287.716256	53.066646	38.176793
58.077903	54.514516	36.328953	53.066646	23.280808	16.971634
40.603558	38.626074	25.243733	38.176793	16.971634	13.916928

Covariance table for class Landslides

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
434.90741	548.253941	738.913491	562.790979	665.420845	471.08241
548.253941	727.547172	1010.940648	785.668364	970.543735	687.585717
738.913491	1010.940648	1445.998179	1138.593354	1447.181084	1025.064733
562.790979	785.668364	1138.593354	908.498653	1173.850291	832.067549
665.420845	970.543735	1447.181084	1173.850291	1639.410396	1181.895237
471.08241	687.585717	1025.064733	832.067549	1181.895237	884.224951

Covariance table for class Vegetation

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
20.395167	17.916302	33.704819	-50.985688	47.339488	49.997858
17.916302	22.015622	33.718077	-29.177843	40.767483	41.096503
33.704819	33.718077	65.148308	-90.860877	92.69204	94.093031
-50.985688	-29.177843	-90.860877	296.425845	-185.235673	-199.791139
47.339488	40.767483	92.69204	-185.235673	196.412687	187.325013
49.997858	41.096503	94.093031	-199.791139	187.325013	190.007159

Covariance table for class Baresoil

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
13.174369	7.286383	-0.707756	-8.94882	-30.183304	-33.370125
7.286383	7.519308	5.532298	0.757411	-6.089717	-8.075324
-0.707756	5.532298	19.743303	19.127694	40.626694	41.616675
-8.94882	0.757411	19.127694	28.089244	64.326452	67.094032
-30.183304	-6.089717	40.626694	64.326452	169.549058	179.056521
-33.370125	-8.075324	41.616675	67.094032	179.056521	193.017481

Covariance table for class Rock

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
18.413035	17.795394	21.070786	14.602035	14.264934	12.725215
17.795394	20.442795	24.415283	17.618309	19.205643	16.861198
21.070786	24.415283	33.590164	24.121053	28.622224	24.703801
14.602035	17.618309	24.121053	19.812762	24.141225	20.711181
14.264934	19.205643	28.622224	24.141225	37.815762	31.348114
12.725215	16.861198	24.703801	20.711181	31.348114	28.403865

Covariance table for class River 1

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
1311.365825	1062.557151	933.250707	280.150453	67.53769	48.092504
1062.557151	868.000173	764.248043	233.674009	58.72691	43.523893
933.250707	764.248043	681.168644	212.91489	55.949839	42.902842
280.150453	233.674009	212.91489	78.675692	35.058083	30.192018
67.53769	58.72691	55.949839	35.058083	93.48626	82.838952
48.092504	43.523893	42.902842	30.192018	82.838952	77.447782

Covariance table for class River 2

Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
389.240781	308.489499	287.182173	75.00525	-67.309035	-62.005006
308.489499	259.242491	245.254823	74.056532	-44.068864	-40.297314
287.182173	245.254823	249.341392	86.000366	-34.63138	-31.118437
75.00525	74.056532	86.000366	47.643956	4.21221	3.865324
-67.309035	-44.068864	-34.63138	4.21221	89.805861	78.266911
-62.005006	-40.297314	-31.118437	3.865324	78.266911	72.249084

Appendix D - Mean and covariance values of defined classes for Boothoven area, Enschede

(Class 1 – bright trees, Class 2 – dark trees, Class 3 – grass dark, Class 4 – grass bright, Class 5 – shadow, Class 6 – impervious, Class7 – shadow vegetation, Class 8 – shrubs)

Mean values of defined classes for Boothoven area, Enschede

Mean_1	Mean_2	Mean_3	Mean_4	Mean _5	Mean_6	Mean_7	Mean_8
158.680718	146.935272	163.37015	159.13701	141.689482	259.139674	144.35294	160.615607
208.498829	174.967167	220.77909	220.28031	153.214802	379.895245	165.82353	216.736994
105.4879	81.266417	120.5	103.99685	66.485253	261.32725	75.176471	109.794798
662.28025	363.611632	467.1796	680.58268	103.845854	296.076856	212.70588	584.150289

Covariance table for class 1

Band 1	Band 2	Band 3	Band 4
38.570636	82.138298	54.959024	52.473924
82.138298	301.159569	166.171275	213.703844
54.959024	166.171275	135.812549	86.235034
52.473924	213.703844	86.235034	5201.22218

Covariance table for class 2

Band 1	Band 2	Band 3	Band 4
32.160126	68.362192	47.575007	178.049017
68.362192	218.197043	142.349601	492.052964
47.575007	142.349601	113.681068	360.991828
178.049017	492.052964	360.991828	3825.47532

Covariance table for class 3

Band 1	Band 2	Band 3	Band 4
20.700847	31.362008	44.152321	3.695492
31.362008	83.684491	88.638819	82.905531
44.152321	88.638819	146.462869	30.531224
3.695492	82.905531	30.531224	1592.893457

Covariance table for class 4

Band 1	Band 2	Band 3	Band 4
19.090032	22.994657	21.609265	-93.802355
22.994657	63.520669	38.687004	-93.363906
21.609265	38.687004	51.495258	-317.122768
-93.802355	-93.363906	-317.122768	5372.107901

Covariance table for class 5

Band 1	Band 2	Band 3	Band 4
59.882367	74.685334	63.487623	-72.437647
74.685334	122.413745	93.880675	-50.969098
63.487623	93.880675	91.041681	-64.820482
-72.437647	-50.969098	-64.820482	752.249611

Covariance table for class 6

Band 1	Band 2	Band 3	Band 4
7801.126599	12697.46134	9957.783238	7508.037057
12697.46134	21454.48799	16953.63718	13856.91866
9957.783238	16953.63718	13646.54885	11353.38556
7508.037057	13856.91866	11353.38556	11408.34541
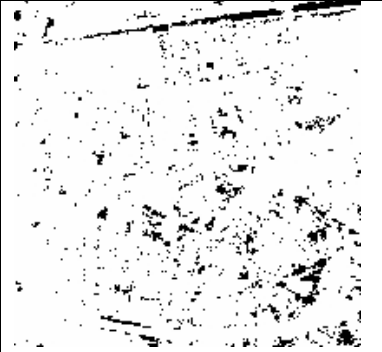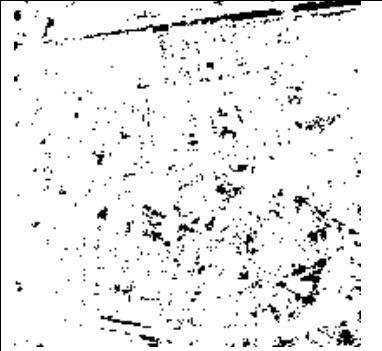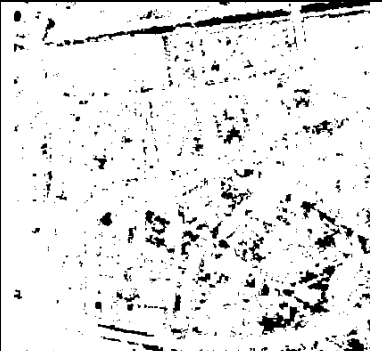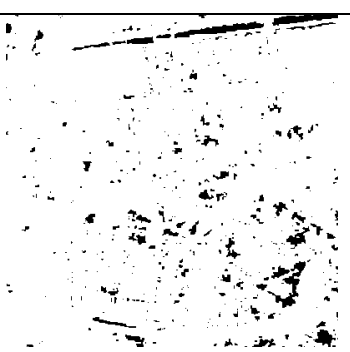
Covariance table for class 7

Band 1	Band 2	Band 3	Band 4
45.992647	40.566176	55.308824	-9.014706
40.566176	99.029412	64.095588	-4.367647
55.308824	64.095588	97.654412	-7.757353
-9.014706	-4.367647	-7.757353	720.345588

Covariance table for class 8

Band 1	Band 2	Band 3	Band 4
22.022828	26.472522	22.642632	95.197068
26.472522	86.629178	54.67341	329.115004
22.642632	54.67341	86.273712	-105.679216
95.197068	329.115004	-105.679216	6545.044014

Appendix E – Influence of smoothness parameter on classification result

Swap-move lambda 0.05	Swap-move lambda 1.3	Swap-move lambda5
Expansion-move 0.05	Expansion-move lambda 1.3	Expansion-move lambda 5
ICM lambda 0.05		
SA log lambda 005	SA log lambda 1.3	SA log lambda 5

Appendix F – Energy values for Simulated Annealing (Exponential schedule)

Simulated Annealing (Exponential schedule)

T0=3.0

	Tupd	Repetition(trial)					Mean	STDEV
		1	2	3	4	5		
Lambda=0.05	0.9	15.965745	15.965739	15.965796	15.965788	15.965733	15.9657602	2.94737E-05
	0.95	15.900328	15.900353	15.900334	15.900353	15.900344	15.9003424	1.12383E-05
	0.99	15.900369	15.900352	15.900347	15.900332	15.900347	15.9003494	1.32778E-05
Lambda=0.5	0.9	15.965745	15.965739	15.965796	15.965788	15.965733	15.9657602	2.94737E-05
	0.95	16.056953	16.056911	16.056993	16.057001	16.057123	16.0569962	7.94305E-05
	0.99	15.965818	15.965737	15.965725	15.965769	15.965749	15.9657596	3.64664E-05
Lambda=1.3	0.9	16.056953	16.056911	16.056993	16.057001	16.057123	16.0569962	7.94305E-05
	0.95	16.057117	16.056881	16.056965	16.056705	16.056507	16.056835	0.000236127
	0.99	16.057112	16.056635	16.057043	16.056671	16.056763	16.0568448	0.000218857
Lambda=5	0.9	16.273563	16.266922	16.265339	16.261799	16.26421	16.2663666	0.004434707
	0.95	16.263857	16.263126	16.261732	16.262936	16.261869	16.262704	0.000894892
	0.99	16.258329	16.258564	16.258165	16.258436	16.258411	16.258381	0.000147287

Appendix H - Energy values for Simulated Annealing (Logarithmic schedule)

Simulated Annealing (Logarithmic schedule, max iter = 1000)

T0=3.0

Lambda	Repetition (trial)					Mean	STDEV
	1	2	3	4	5		
0.05	15.905455	15.906125	15.90576	15.905895	15.905972	15.9058414	0.000253125
0.5	15.970675	15.970982	15.971231	15.971072	15.971217	15.9710354	0.00022656
1.3	16.062046	16.061014	16.062183	16.06155	16.062166	16.0617918	0.000505253

Faster version of SA with Logarithmic schedule

T0=3.0

Lambda	Repetition (trial)					Mean	STDEV
	1	2	3	4	5		
0.05	15.900319	15.900347	15.900328	15.900349	15.900332	15.900335	1.27867E-05
0.5	15.965725	15.96574	15.965783	15.96572	15.9657	15.9657336	3.11014E-05
1.3	16.057106	16.057045	16.056793	16.056875	16.056877	16.0569392	0.000130749
5	16.260073	16.258778	16.259539	16.258297	16.258741	16.2590856	0.000709892