

DETECTION OF BUILT-UP AREAS IN SAR AND ASTER IMAGES USING CONDITIONAL RANDOM FIELDS

BENSON KIPKEMBOI KENDUIYWO
March, 2012

SUPERVISORS:
Dr. V. A. Tolpekin
Prof. Dr. Ir. A. Stein



DETECTION OF BUILT-UP AREAS IN SAR AND ASTER IMAGES USING CONDITIONAL RANDOM FIELDS

BENSON KIPKEMBOI KENDUIYWO
Enschede, The Netherlands, March, 2012

Thesis submitted to the Faculty of Geo-information Science and Earth
Observation of the University of Twente in partial fulfilment of the requirements
for the degree of Master of Science in Geo-information Science and Earth
Observation.
Specialization: Geoinformatics

SUPERVISORS:

Dr. V. A. Tolpekin
Prof. Dr. Ir. A. Stein

THESIS ASSESSMENT BOARD:

Prof. Dr. Ir. A. Stein (chair)
Dr. Ir. L. Kooistra

Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

The current increase in population has resulted in widespread spatial changes, particularly rapid development of built-up areas, in the city and its environs. Up-to-date spatial information is requisite for the effective management and mitigation of the effects of built-up area dynamics. Spectral heterogeneity in built-up areas, however, remains a challenge to existing classification methods. This study developed a method for detecting built-up areas from SAR and ASTER images using conditional random field (CRF) framework. A feature selection approach and a novel data dependent term of CRF was designed and used to classify image blocks. Mean, standard deviation and variogram slope features were used to characterize training areas with slope describing spatial dependencies of classes. The association potential was designed using support vector machines (SVM), which can handle redundant data with any form of distribution while the inverse of transformed euclidean distance was used as a data dependent term of the interaction potential. The latter maintained stable accuracy when subjected to a range of small to large smoothness parameters while preserving class boundaries and aggregating similar labels during classification. This enforced a discontinuity adaptive model that moderated smoothing given data evidence unlike MRF, which penalizes every dissimilar label indiscriminately. The accuracy of detecting built-up areas using CRF exceeded that of markov random field (MRF), SVM and maximum likelihood classification (MLC) by 1.13%, 2.22% and 8.23% respectively. It also had the lowest false positives. The method illustrated that built-up areas increased by 98.9 hectares while 26.7 hectares were converted to non-built-up areas. Thus, it can be used to detect and monitor built-up area expansion; hence provide timely spatial information to urban planners and other relevant professionals.

Keywords

Spatial dependencies, built-up areas, non-built-up areas, Conditional Random Field (CRF), Markov Random Field (MRF), Support Vector Machines (SVM), variogram.

ACKNOWLEDGEMENTS

All glory to GOD my eternal Father, the creator of everything and beholder of all mysteries humans can not fathom, for the gift of life, care and blessing you bring my way daily.

My utmost appreciation to my supervisors, Dr. Valentyn Tolpekin and Prof. Dr. Alfred Stein for scientific advice and guidance throughout the research. Dr. Valentyn for encouraging me to use L^AT_EX an interesting document processing software, providing image analysis advice and making me enjoy using R software. Prof. Stein for the valuable geostatistical advice.

I am indebted to Elizabeth Marcello project manager Center for Sustainable Urban Development (CSUD) for availing me nairobi land-use vector data, Dr. Richard Sliuzas for his advice, Dr. Rolf de By and Dr. Ivana Ivanova for providing L^AT_EX support through discussion board and personal consultation.

I wish to extend my gratitude to all my classmates and friends who made life here in Enchede home away from home. I am especially grateful to Juan Pablo Ardila Lopez, Kennedy Were and Felix Ngetich for their encouragement and advice during the research period.

My appreciation to my parents, James Kenduiywo and Agnes A. Kenduiywo for their love and support throughout my life, my sister Irine and brother Denis who have always encouraged me, and most importantly my lovely wife Terry and son Ethan to whom I dedicate this work.

Finally, I wish to thank the Dutch government who realized my thirst to enrich my skills in the field of geo-information through the Nuffic scholarship programme.

Inspiration:

"When you have eliminated the impossible, whatever remains, however improbable, must be the truth." Sherlock Holmes.

TABLE OF CONTENTS

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 Background	1
1.2 Motivation and problem statement	1
1.3 Research identification	2
1.3.1 Research objectives	2
1.3.2 Research questions	3
1.3.3 Innovation aimed at	3
1.3.4 Research approach	4
1.4 Thesis outline	4
2 Literature review	5
2.1 Theory and concepts	5
2.1.1 Texture or context?	5
2.1.2 MRF	6
2.1.3 Image labelling with CRF	8
2.1.4 CRF theory	9
2.2 Built-up area detection using remote sensing	11
2.3 Built-up area detection studies using spatial context	13
2.4 Summary	14
3 Materials	15
3.1 Study area location	15
3.2 Data	15
3.2.1 Optical images	15
3.2.2 SAR images	15
3.2.3 Subset images	17
3.2.4 Reference data	17
3.3 Software	18
4 Implementation	19
4.1 Data pre-processing	19
4.2 Class definition	20
4.3 Feature selection	21
4.4 Conditional random fields for built-up area detection	22
4.4.1 Association potential (A)	22
4.4.2 Interaction potential (I)	24
4.4.3 Training of designed association potential	25
4.4.4 Classification of built-up areas using CRF	27
4.4.5 Test of CRF interaction potential data dependent models	27
4.5 Application of CRF in detecting built-up areas including temporal changes.	27

4.6	Accuracy assessment	28
5	Results	29
5.1	Feature selection	29
5.2	Designed CRF results	35
5.2.1	Designed association potential	35
5.2.2	Test of CRF interaction potential data dependent models	35
5.3	Application of CRF in detecting built-up areas including temporal changes . . .	37
5.4	Assessment of CRF in detecting built-up areas	37
6	Discussion	41
7	Conclusion and recommendations	45
A	Appendix: Part one	A-1
A.1	Designed CRF potentials: R program code	A-1
A.2	Optimization of CRF using ICM: R program code	A-4
A.3	Change detection: R program code	A-5

LIST OF FIGURES

1.1	An illustration of the fact that remotely sensed images contain strong spatial dependencies rather than being a random collection of independent pixels or blocks. (a) Natural image scene. (b) Image obtained by randomly combining pixel values in (a). (c) Image obtained by randomly combining the original image blocks. Modified from [22].	3
1.2	Research approach.	4
2.1	MRF neighbourhood systems according to [26]. (a) First-order four local neighbours sharing a side with pixel i . (b) Second-order eight local neighbours of pixel i . Higher-order can be extended in a similar fashion.	7
2.2	Figures (a), (b) and (c) show first order neighbourhood system cliques.	8
2.3	Physical patterns characterizing built-up area development. (a) Compact development. (b) Scattered development. (c) Linear strip development. (d) Legend. (e) Leapfrogging development. Modified from [13].	12
3.1	Study area location.	16
3.2	Part of Nairobi city as seen from: (a) ASTER false colour composite (3,2,1) image. (b) SAR image. (c) Aerial photo.	17
3.3	(a) ASTER false colour composite (3,2,1) image and (b) SAR intensity image.	17
4.1	Methodological framework.	19
4.2	Slope (θ) computation using first three points of lag and semi-variance from the variogram.	22
4.3	Sample variogram with a linear model fit on the first three points.	23
4.4	Training of association potential to determine most optimal values for cost (C) and sigma (σ) parameters. (a) Scale I and II. (b) Scale III.	25
4.5	ASTER false colour composite (3,2,1) image showing study area training samples collected using scale II blocks.	26
5.1	Mean and standard deviation image features showing homogeneity and variability of classes respectively using scale II. (a) Mean of amplitude features. (b) Standard deviation of amplitude features. (c) Mean of intensity features. (d) Standard deviation of intensity features.	30
5.2	Amplitude and intensity training blocks feature space plots.	31
5.3	Sample variograms of built-up and non-built-up super-classes as defined in Section 4.2. (a) Sample variograms of scale I blocks using ASTER intensity. (b) Sample variograms of scale I blocks using SAR amplitude.	33
5.4	Scale I block samples from high density built-up and forest classes using amplitude and intensity images; amplitude blocks gave zero slope.	34
5.5	Association potential classification results and misclassification error distribution. (a) Scale I. (b) Scale II.	35
5.6	Results of interaction potential designs in comparison to MRFs.	36
5.7	CRF classification and misclassification error distribution. (a) Scale I. (b) Scale I.	37
5.8	CRF classification of built-up areas. (a) 2006 classification. (b) 2011 classification.	38

5.9	Built-up area changes from 2006–2011 and reference data. (a) Land-cover changes. (b) Study area reference data for 2011.	38
5.10	2011 class probability images computed using association potential. (a) Probability of built-up areas. (b) Probability of non-built up areas.	39

LIST OF TABLES

4.1	Image geo-referencing accuracy summary.	20
5.1	Summary of evaluation of suitable combination of features from hue, amplitude and intensity.	30
5.2	Summary of evaluation of detection ability of amplitude and intensity features. .	32
5.3	Summary of assessment of variogram's slope features detection ability using scale I blocks.	32
5.4	CRF accuracy assessment.	39
5.5	MRF accuracy assessment.	39
5.6	SVM accuracy assessment.	39
5.7	MLC accuracy assessment.	39

Chapter 1

Introduction

1.1 BACKGROUND

Technological development in remote sensing has increased availability of data with rich spectral, spatial and temporal resolutions. Consequently, demand of remotely sensed data for urban area applications such as, monitoring urban growth, study of population density, road network planning and land-use planning has increased. Remote sensing provides a fast and cost-effective means of automatically obtaining up-to-date information, temporal datasets over large areas, fast digital image processing and analysis amongst others. Technological advancement is often faced with new challenges some of which are foreseen while others are unforeseen. One emerging demand is the need for new efficient image analysis methods to extract information from the ever increasing data. This is because existing classification methods are no longer effective in extracting information from outcomes of daily acquired geospatial data.

1.2 MOTIVATION AND PROBLEM STATEMENT

Developed and developing countries are currently characterized by rapid urbanization process and accelerated population growth [40]. Cities are expanding to accommodate these processes and the subsequent challenge is growth of built-up areas. Nairobi city is expanding rapidly and built-up areas have become common [32]. The dynamic growth demands regular update of land-use information in order to ensure correspondence with changes in built-up area extent and location. Urban planners lack efficient tools because existing data capture methods like, surveying or digitizing of aerial photos, have technical and economic constraints which hamper frequent and wide area updates [12]. Ground surveys, for instance, are often limited by logistical constraints and are confined to small areas. Synoptic view of remote sensing sensors makes data acquisition over wide areas efficient. This offers a timely and cost-effective way of gathering information on built-up areas if appropriate an classification method is developed.

Precise detection of built-up areas using remote sensing techniques is a challenge to ordinary pixel based classification methods. Built-up areas are composed of a complex mixture of land-cover classes (i.e. impervious surfaces, grass, scattered trees and small gardens) all which occur within a small area [29, 48]. The mixed land-cover structure causes substantial intra-class and inter-class spectral variability. This is a problem to ordinary per-pixel classifiers which perform well in spectrally homogenous areas as opposed to areas with high spectral variability [48]. In spectrally heterogenous areas, pixel-based methods like maximum likelihood classification (MLC) produce noisy classified images with "salt and pepper" like appearance [18, 30]. MLC is based on the assumption of normal distribution which is violated when used in built-up areas where data deviate from normal distribution. Thus, built-up area inter-pixel and intra-pixel spectral changes, hampers the ability of pixel-based methods to resolve inter-class confusion. A method that can delineate built-up areas accurately from remote sensing images is requisite. A multi-source strategy integrating synthetic aperture radar and optical images was adopted as one of the

solutions as demonstrated by [14].

Choice and performance of a classification method depends on the resolution of images used [30]. The study used 15 m resolution ASTER¹ and ALOS-PALSAR² images. In such resolution individual objects, i.e. buildings, trees and roads, except large structures, in urban areas are merged into a single class "built-up area" which consist of heterogenous spectral mixture [18]. Though the images are of medium resolution, considerable spectral variability still exist in the optical image. In addition, SAR images are accompanied by speckle which degrades accuracy of per pixel analysis. This can be minimized by analyzing each pixel within its neighbourhood as noted by [14]. A new approach that can overcome pixel variability — optical image spectral variability and speckle noise in SAR image — during classification is needed.

Object oriented classification approach is designed to deal with the challenge of pixel variability. The technique merges pixels into blocks known as "objects" using image segmentation. Classification is then conducted using "objects" instead of an individual pixel. This improves classification accuracy because spectral variability and noise in pixels is averaged. Despite improved accuracy, the method still ignores contextual information. Context accounts for spatial dependencies among pixels. It determines probability of a pixel or a group of pixels occurring at a given location based on the nature of other pixels in the image [42]. Goodchild [15] defines spatial dependence as "the propensity for nearby locations to influence each other and to possess similar attributes." Neighbouring locations in built-up areas indeed have similar spatial attributes. Therefore, integration of spatial dependencies using contextual classification methods can enhance detection.

Markov Random Field (MRF) is one of the commonly used contextual classification methods. It is a probabilistic approach that models spatial dependencies of labels in a classified image [26]. Despite the popularity of MRF, its underlying assumption of conditional independence of pixels adopted for computational tractability, neglects spatial dependencies in the observed data [49]. In reality, images exhibit a coherent scene because pixels that represent objects in it have strong spatial dependencies [22]. This is illustrated by Figure 1.1, which proves that images contain strong spatial dependencies rather than being a random collection of independent pixels. Modelling this spatial dependency can improve classification accuracy [23].

A method that models spatial dependencies in both image data and labels can improve classification [18]. Conditional Random Field (CRF) framework was adopted as it includes spatial dependencies of both class labels and data in a statistical manner [49]. Their structure incorporates data-dependent interaction in image classification unlike conventional MRF method [23]. This facilitated design of a discontinuity adaptive smoothness model used in detection of built-up areas. The framework was developed to model features³ useful for built-up area detection. Detection of built-up areas was done using blocks of merged pixels in a 8×8 window size. Use of image blocks minimized spectral variability common in built-up areas [see 48, chap. 3].

1.3 RESEARCH IDENTIFICATION

1.3.1 Research objectives

The main objective of this research is to develop and apply a method based on CRF to detect built-up areas from SAR and ASTER images. This is achieved through the following specific objectives:

¹Advanced Spaceborne Thermal Emission and Reflection Radiometer

²Advanced Land Observing Satellite Phased Array type L-band Synthetic Aperture Radar

³Features refer to characteristics or attributes that provide a quantitative measure of a class, i.e. mean or standard deviation.

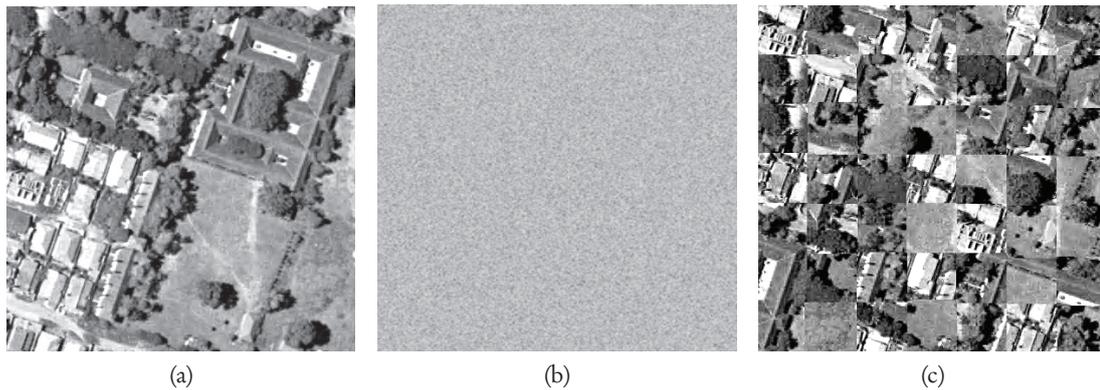


Figure 1.1: An illustration of the fact that remotely sensed images contain strong spatial dependencies rather than being a random collection of independent pixels or blocks. (a) Natural image scene. (b) Image obtained by randomly combining pixel values in (a). (c) Image obtained by randomly combining the original image blocks. Modified from [22].

1. To identify, select and incorporate features characterizing built-up areas from SAR and ASTER images.
2. To design an algorithm that model spatial dependencies of features using CRF.
3. To apply the designed algorithm in detection of built-up areas including their temporal changes.
4. To evaluate performance of the designed method compared to maximum likelihood classification, support vector machines and markov random field.

1.3.2 Research questions

1. Which land-cover classes and features are suitable for detection of built-up areas?
2. How can spatial dependencies in built-up areas be modelled using CRF?
3. Which scale/block size is suitable for detecting built-up area and their temporal changes?
4. What is the performance of the designed method compared to maximum likelihood classification, support vector machines and markov random field?

1.3.3 Innovation aimed at

This research aimed at developing a multi-source classification approach for detecting built-up areas using spatial dependencies of features. Optimal features characterizing built-up areas were selected and used for classification. Variogram slope, mean and standard deviation image features were used. CRF unary potential was designed using Support Vector Machines (SVM) non-linear decision boundary instead of the ordinary logistic classifier. A new data dependent term was developed for CRF interaction potential using inverse of transformed euclidean distance. Slope features derived from the variogram were used in the data dependent term to model spatial dependencies of built-up areas.

1.3.4 Research approach

In order to attain the stated objectives a sequence of activities were adopted. To begin with, a review of contextual classification methods was done with much emphasis on limitations of conventional MRF and how CRF addresses the limitations. An approach of feature selection and modelling spatial dependencies using CRF was then designed. A novel data dependent term based on inverse of transformed euclidean distance was designed and used for the first time. The designed CRF was then used to classify built-up areas. Figure 1.2 illustrates the main research approach. A detailed methodological approach is summarized in Figure 4.1.

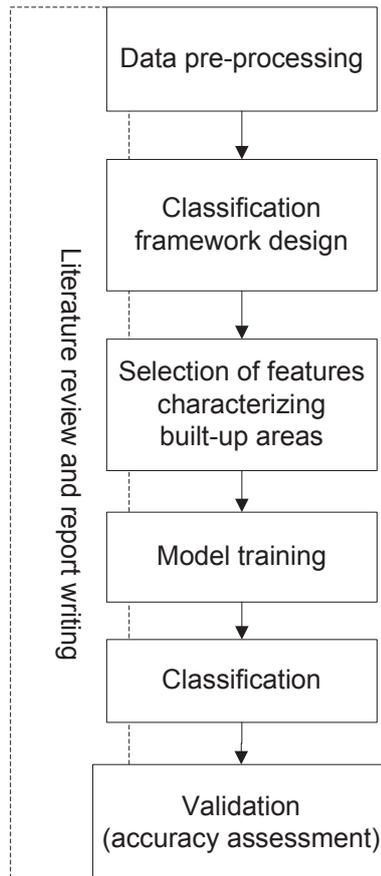


Figure 1.2: Research approach.

1.4 THESIS OUTLINE

This thesis consist of seven chapters. Chapter one sets the research background, problem, objectives and research questions. A review of theoretical concepts of contextual classification methods and related works is addressed in the second chapter. The third chapter describes the study area chosen, data and software used to execute the research. Chapter four sets out a detailed explanation of research implementation including the details of the designed method. Results are presented in chapter five followed by the discussion in chapter six. The last chapter presents conclusions from the study and makes suggestions for further work.

Chapter 2

Literature review

This chapter provides a theoretical background to the research content. The use of spatial dependency in image classification as well as the relevant methods for modelling spatial dependency in images have been discussed. Some key urban area terminologies and physical characteristics are defined while addressing ways of measuring them using remote sensing techniques. Finally, a review of urban area detection studies using contextual classification techniques is provided.

2.1 THEORY AND CONCEPTS

2.1.1 Texture or context?

Humans naturally use spectral, textural and contextual features to interpret information in images. Features refer to characteristics or attributes that provide a quantitative measure of a class, i.e. forest, from an image. Such attributes may be spectral reflectance or emittance values from optical image or secondary measurements derived from an image like texture [42]. An overview of these features is addressed in this paragraph with reference to [16]. Spectral features describe the average tonal variations over various bands of an image. Textural features bear information on the spatial distribution of tonal variations within an image. The concept of tone is derived from different shades of gray scale of image pixels while texture describes the spatial distribution of gray tones. Contextual features contain information derived from a pixel based on the nature of pixels in the neighbouring and/or remainder scene. Spatial context determines the probability of a pixel occurring at a given location in the image based on the attributes of the surrounding pixels; hence spatial dependencies.

Land-use/land-cover classification aims to delineate homogeneous areas in an image as opposed to separate objects such as tree or building. The homogeneous areas are referred to as classes. Built-up areas consist of heterogeneous spectral features which are related in a given neighbourhood. These neighbourhoods form a homogeneous area as opposed to being independent pixels. Probability of pixels occurring within a given neighbourhood is described by spatial context while their spectral features account for their differences. Textural features contain information about the structural arrangement of features and their neighbourhood [16]. Similarly, spatial context accounts for the relationship between a pixel being analyzed and those in the remainder of the scene [see 42, pg. 221]. Generally, the two describe spatial structure of an object in an image [see 42, pg. 63]. This depicts an inextricable relationship between texture and spatial context.

Context is used to model different spatial dependencies in images. In [22], two categories of context are defined viz: local and global. Local context models local smoothness of pixels or dependencies among different parts of an object while global context describes dependencies among bigger objects and classes in images. Modelling spatial dependencies in images can improve classification results. Commonly used methods in modelling spatial dependencies are conventional MRF and recent CRF used in the study. Theoretical frameworks of the methods are reviewed in Sections 2.1.2 and 2.1.4 based on the works of [see 42, chap. 8] and [see 26, chap. 1 and chap. 2].

2.1.2 MRF

Markov random fields provide a convenient and consistent framework for modelling context-dependent entities by characterizing their mutual influences using local conditional probabilities. Assuming that $W = \{W_1, \dots, W_m\}$ is a family of random variables defined on the set S , of pixels values, where each random variable W_i is assigned a value w_i in a set of labels L , then W is called a random field. For consistency with notations w is used in this study to refer to random field W . Thus, w is a MRF with respect to a defined neighbourhood system N if its probability density function fulfills three properties namely:

1. Positivity: $P(w) > 0 \forall$ possible configurations of w ,
2. Markovianity: $P(w_i|w_{S-i}) = P(w_i|w_{N_i})$, and
3. Homogeneity : $P(w_i|w_{N_i})$ is the same \forall sites i .

where $S - i$ is the set of all pixels in S excluding i , w_{S-i} denotes the set of labels at sites¹ $S - i$ and N_i are the neighbours of site i . Markovianity property shows that labelling of a site i is dependent on its neighbours (local neighbourhood property). Homogeneity property defines the likelihood for a label at site i given its neighbourhood regardless of the relative position of i in S . Figure 2.1 shows commonly used neighbourhood systems.

MRF is used in a generative probabilistic framework where joint (prior and the conditional) probability distribution of the observed data and the corresponding labels is modelled. It is a generative model because it uses Bayes rule to predict the posterior probability of a label and then pick the most likely label. For instance, if d is observed data (image) where $d = \{d_i\}_{i \in S}$ and d_i is data from a given site i , S the set of all pixels in the image and $w = \{w_i\}_{i \in S}$ the corresponding labels, the posterior probability $P(w|d)$ over the labels, given the observed data, is expressed using Bayes' rule:

$$P(w|d) = \frac{P(d|w)P(w)}{P(d)} \quad (2.1)$$

A pixel at a site i can then be allocated to a class w_k which has the highest value of the term $P(w|d)$, that is, the Maximum A Posterior (MAP) solution:

$$w_k = \arg \max_w \{P(d|w)P(w)\} \quad (2.2)$$

where $\arg \max$ denotes maximum value of the argument. MAP probability is a popularly used optimization statistical criteria chosen for MRF modeling. MRF and MAP concepts together form a MAP-MRF classification framework that determines the joint probability of class labels given a neighbourhood system. The objective function, MAP in Equation 2.2, is then adopted for classification where labelling is performed by minimizing the posterior energy. Minimum energy in this case is equivalent to maximum probability of a label.

Neighbourhood system

Spatial dependency of a set of sites S in an image is defined by a neighbourhood system. A guiding principle of MRF is that the information contained in the local neighbourhood of a site i is sufficient to obtain a good global image representation. This is attributed to its equivalence with Gibbs Random Fields (GRF). GRF describes the global properties of an image in terms of joint distribution of labels in all sites. This property allows MRF model to be defined in terms

¹Site refer to a pixel at agiven location in the image.

of GRF formulation which makes it easier to deal with spatial dependency. The GRF considers a neighbourhood system based on cliques. A clique is a subset in which all pair of sites are mutual neighbours as illustrated in Figure 2.2.

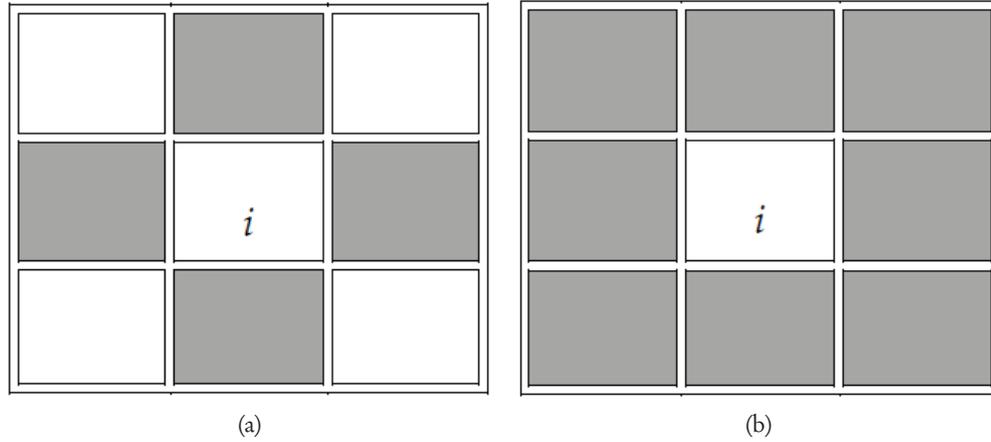


Figure 2.1: MRF neighbourhood systems according to [26]. (a) First-order four local neighbours sharing a side with pixel i . (b) Second-order eight local neighbours of pixel i . Higher-order can be extended in a similar fashion.

Energy functions

Markov random fields model spatial dependency by optimizing local conditional distributions. The MRF-GRF equivalence allows the use of GRF to model spatial dependency during classification. A GRF provides a global model for an image by modelling the probability distribution function (p.d.f) as:

$$P(w) = \frac{1}{Z(d)} \exp \left[-\frac{U(w)}{T} \right] \quad (2.3)$$

where T is a constant termed *temperature*, $Z(d)$ is a data normalizing constant referred to as the partition function and $U(w)$ is referred to as energy function. The role of the energy function is twofold. First, it acts as a quantitative measure of global quality of a solution. Second, it acts as guide in the search of a minimal solution.

$$Z(d) = \sum_{\forall \text{ configurations of } w} \exp \left[-\frac{U(w)}{T} \right] \quad (2.4)$$

Maximizing $P(w)$ in Equation 2.3 is equivalent to minimizing the energy function:

$$U(w) = \sum_{c \in C} V_c(w) \quad (2.5)$$

where C is known as a clique and $V_c(w)$ is the potential function with respect to clique type C . Cliques of first order neighbourhood system with respect to a site i are illustrated in Figure 2.2.

MAP-MRF labelling

The popularity of MRF in labelling can be attributed to its equivalence with GRF as proved in [4]. A MRF is defined by local property whereas GRF describes global properties of an image in terms of joint distribution of classes of an image. This provides a means of dealing with

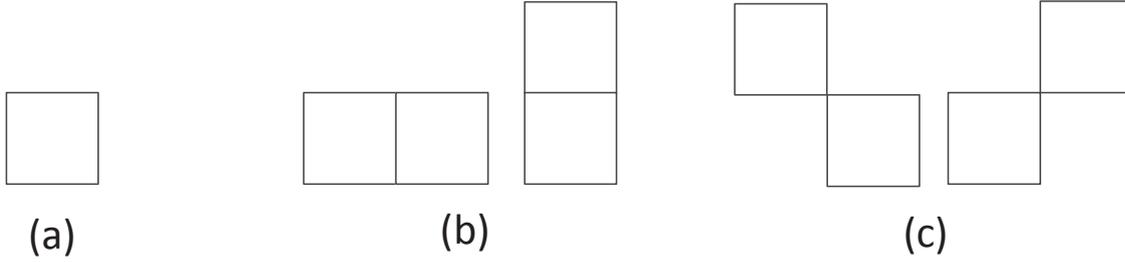


Figure 2.2: Figures (a), (b) and (c) show first order neighbourhood system cliques.

MRF-based spatial dependency and reduces the complexity of the model as it can be expressed in terms of GRF formulation. The MAP-MRF framework has two roles. One is to derive posterior distribution using Bayesian formulation and to determine the parameters in it. Second is to design an optimization algorithm to find the maximum posterior distribution. An optimal solution for Equation 2.2 is computationally intractable for any practical situation. Computational tractability is attained by simplifying the likelihood model to:

$$P(d|w) = \prod_{i \in S} P(d_i|w_i) \quad (2.6)$$

Consequently, this implies computational complexity is minimized by factorizing global optimization with a collection of local optimizations. This simplifies the MAP estimate into a minimization of a sum of local energy functions:

$$\hat{w} = \arg \min_w U(w|d) \quad (2.7)$$

Where $U(w|d)$ is the energy function being minimized [see 26, chap. 1]. In binary classification, the prior $P(w)$ is assumed to be homogeneous and isotropic with only pairwise non-zero potentials. The model often used for classification is,

$$P(w|d) = \frac{1}{Z(d)} \exp \left\{ \sum_{i \in S} \log P(s_i(d_i)|w_i) + \frac{1}{2} \sum_{i \in S} \sum_{j \in N_i} \beta w_i w_j \right\} \quad (2.8)$$

where β is the interaction parameter of MRF and $s_i(d_i)$ is a single site feature vector acting as an association parameter [23]. Multiplying the interaction parameter by half accounts for double counting of labels that occurs in a given neighbourhood system. Optimization methods are used to determine the solution of Equation 2.7 by maximizing Equation 2.8. Iterative algorithms such as Iterated Conditional Modes (ICM), Simulated Annealing (SA) and Maximizer of Posterior Marginals (MPM) are usually adopted [see 42, chap. 8.4].

2.1.3 Image labelling with CRF

A good classification approach involves a method capable of learning diverse dependencies in an image automatically in a single consistent framework from the training data [23]. To achieve this two types of spatial dependencies can be modelled. First, remotely sensed images exhibit spatial smoothness in labels of the same class. Accordingly neighbouring sites tend to bear similar labels except at class boundaries. Second, complex spatial dependencies exist in the observed data. CRF

proposed by [25] offer this capability by relaxing the assumption of conditional independence in the observed data.

As depicted in Equation 2.6, for computational tractability, the likelihood model of MRF is assumed to be a fully factorized form. The assumption is too restrictive as it ignores spatial dependencies inherent in remotely sensed images when assigning labels to classes [18, 49]. For instance, a class that contains man-made structures like built-up areas is highly dependent on its neighbours. This is because in built-up areas, the lines or edges at spatially adjoining sites follow underlying rules other than being random [23]. Heterogenous appearance of built-up areas in images also show local inter-related patterns which should be modelled [18]. This exhibits a spatial order in images where an observation at a particular site is correlated with those of the surrounding sites. In addition, when the observations are conditioned on the labels, they are not independent from each other as assumed by MRF [33]. While it is essential to have a model that attains a tractable inference, it is also desirable that it represents the data without making unwarranted independence assumption. To fulfil both requirements, the conditional distribution over the labels given the observed data is modelled instead of the joint probability distribution over both labels and the observations [38]. This avoids modelling a complicated probability distribution function over the data, $P(d)$, that can lead to intractable models. This is the discriminative approach adopted by CRF. A review of CRF theory is addressed in section 2.1.4 with reference to [18, 25, 26].

2.1.4 CRF theory

For consistency with notations, let $d = d_1 \dots d_m$ be a family of random variables over the observed data and $w = w_1 \dots w_m$ be a family of random variables over the corresponding labels. Lafferty et al. [25], defines a label set w conditioned on d to be a CRF if every w_i satisfies the markovianity property with positivity assumed, where w and d are random fields. Therefore, a CRF can be viewed as a MRF globally conditioned on the observations d .

$$P(w_i|d, w_{S-i}) = P(w_i|d, w_{N_i}) \quad (2.9)$$

Markov-Gibbs equivalence provides a global model for observed data by specifying a probability distribution function as:

$$P(w|d) = \frac{1}{Z(d)} \exp \left\{ -\frac{1}{T} U(w|d) \right\} \quad (2.10)$$

Where $U(w|d)$ is the energy function and T is temperature which controls the width of the distribution. MRF generative classification approach determines class labels by maximizing the posterior, $P(w|d)$, over the class labels given the data. In contrast, CRF discriminative framework model the posterior probability, $P(w|d)$, directly as an MRF without modeling the prior and likelihood individually [26]. This reduces the complexity encountered by MRF in modelling the prior and likelihood individually. The conditional distribution over the labels is expressed as:

$$P(w|d) = \frac{1}{Z(d)} \exp \left\{ \sum_{i \in S} A_i(w_i|d) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(w_i, w_j|d) \right\} \quad (2.11)$$

Determination of Z is computationally intractable and approximate methods are used to compute parameters and optimize the solution in Equation 2.11. The association potential A_i links the class label w_i at site i to the observed data d . Interaction potential I_{ij} models the dependencies between the labels w_i and w_j of neighbouring sites i and j and the data. The association and interaction potentials offer two advantageous differences compared to MRF. First, the association potential of site i is a function of all the data as well as the label of that site w_i . Therefore, data

from neighbouring sites N_i are no longer conditionally independent as in MRF where the association potential is a function of data only at that site, i.e., d_i . Second, MRF models dependencies in labels only making the interaction term, adopted from prior probability $P(w)$, to act as a smoothness term over the labels. The interaction potential in CRF models spatial dependencies between the labels and the data.

CRF was proposed by [25] in the context of labelling and segmenting one dimensional (1-D) text sequences by directly modelling the posterior as Gibbs field. Image classification is a binary problem which involves assigning labels to pixels, i.e. $w_i \in \{-1, 1\}$, on a two dimensional (2-D) regular grid. Assuming the random field in Equation 2.11 to be homogenous and isotropic, that is A_i and I_{ij} are independent of locations i and j , CRF model simplifies to:

$$P(w|d) = \frac{1}{Z(d)} \exp \left\{ \sum_{i \in S} A(w_i|d) + \sum_{i \in S} \sum_{j \in N_i} I(w_i, w_j|d) \right\} \quad (2.12)$$

The association (unary) and interaction (pairwise) potentials can be regraded as arbitrary local classifiers. This property enables use of domain-specific discriminative classifiers in structured data rather than restricting the potentials to a certain form [50]. Sections 4.4.1 and 4.4.2 defines CRF functional model of A and I .

Association potential (A)

The association potential is a measure of how likely a site i takes a label w_i given the observed data d_i without influence of all other sites. Kumar et al. [24] propose use of local discriminative classifiers to determine A . To achieve this, Generalized Logistic Models (GLM) are used to determine local class posteriors (conditional probability of a class w_i at site i given the observed image d) of the A . Thus, a logistic function is used to fit the local class posteriors directly as:

$$P'(w_i|d) = \frac{1}{1 + e^{-(x_0 + x_1^T f(d_i))}} = \sigma(x_0 + x_1^T f(d_i)) \quad (2.13)$$

where $x = \{x_0, x_1\}$ are model parameters, $\sigma(w) = \frac{1}{1+e^{-w}}$ and $f(d_i)$ is a site-wise (unary) feature vector computed for site i but may also depend on the entire image d [23]. The form of logistic model in Equation 2.13 yields a linear decision boundary spanned by feature vectors $f(d_i)$. The logistic model can also be extended to introduce a non-linear decision boundary by adopting a transformed feature vector in 2.14 at each site i .

$$h(d_i) = [1, \phi_1(f(d_i)), \dots, \phi_N(f(d_i))]^T \quad (2.14)$$

where ϕ_1, \dots, ϕ_N are arbitrary non-linear functions and $N + 1$ is the dimension of the transformed feature space. The first coefficient of $h(d_i)$ is fixed to one in order to accommodate the bias parameter x_0 . Equation 2.13 becomes:

$$P'(w_i|d) = \frac{1}{1 + e^{-(x_0 + x_1^T h(d_i))}} = \sigma(x_0 + x_1^T h(d_i)) \quad (2.15)$$

where the vector $x = [x_1, x_2, \dots, x_N, \alpha_1]^T$ contains the weights/parameters of the features in $h(d_i)$ that are tuned during the training process, and α_1 is a bias term [27]. Basically, the logistic model adopted for A does not include interaction among neighbouring image sites. This simplifies the model of A into an expression of local log-likelihood of an individual image site:

$$P'(w_i|d) = \log \sigma(w_i x^T h(d_i)) \quad (2.16)$$

The transformation of A in Equation 2.16 ensures that the CRF model (Equation 2.12) simplifies to a logistic classifier when I is set to zero.

Interaction potential (I)

The interaction potential acts as a measure for the influence of image data d and the neighbouring label w_j on the label w_i of site i . It imposes spatial interaction and can be seen as a data dependent smoothing function. The homogenous and isotropic Ising model, $I = \beta w_i w_j$, in MRF framework does not permit data dependent interaction due to the assumption of conditional independence in observed data. This ignores existing spatial dependencies in images. The MRF I model penalizes each dissimilar pair of labels by a constant smoothness parameter (β). Such model gives preference to piecewise constant smoothing without explicitly accounting for discontinuities in the data [24]. In contrast, I in CRF is a function of all the observed data d . It can be linked to the conditional probability, $P''(w_i = w_j|d)$, of the existence of similar labels at sites i and j given the observed image d :

$$I(w_i, w_j, d) = \log P''(w_i = w_j|d) \quad (2.17)$$

Thus, the model for I is:

$$I(w_i, w_j, d) = w_i w_j v^T \Psi(d_{ij}) \quad (2.18)$$

where $\Psi(d_{ij})$ is a pairwise feature vector obtained by concatenation or passing the two vectors $h(d_i)$ and $h(d_j)$ through a distance function i.e. by subtraction ($\Psi(d_{ij}) = |h(d_i) - h(d_j)|$) [23]. The first component of $\Psi(d_{ij})$ is fixed to 1 to accommodate the bias term in v^T . Vector $v = [v_1, v_2, \dots, v_N, \alpha_2]^T$ contains the weights of features where α_2 is a bias term. The model in Equation 2.18 was proposed by [24, 23] as it is simple and makes parameter learning a convex problem. Substituting the defined models of A and I , the CRF model can be expressed as:

$$P(w|d) = \frac{1}{Z(d)} \exp \left\{ \sum_{i \in S} \log \sigma(x^T h(d_i)) + \sum_{i \in S} \sum_{j \in N_i} w_i w_j v^T \Psi(d_{ij}) \right\} \quad (2.19)$$

The CRF model is normally used for classification after determining interaction and association potentials. This requires determination of posterior probability $P(w|d)$ in Equation 2.19. Computation of posterior probability involves evaluation of partition function $Z(d)$ which is computationally intractable. A solution can be obtained by using sampling techniques or approximations such as pseudo-likelihood (PL), mean field or Loopy Belief Propagation (LBP) to estimate the parameters. After the partition function is determined, image labelling is performed by determining optimal class label w given the extracted features. This is done by inference methods. A MAP estimate is normally adopted as the solution to the inference problem [22].

2.2 BUILT-UP AREA DETECTION USING REMOTE SENSING

Urban area or built-up area? Bhatta [5] defines an urban area as "towns and cities — an urban landscape". The author notes that definition of urban varies country to country. For instance, a country may define a farmland surrounded by a residential area as part of an urban area while in another country is defined as non-urban. Thus, determination of land-use classes is often subjective. Bhatta proposes eliminating such confusion by differentiating urban land-cover and urban land-use. Land-cover refers to objects on land surface natural or man-made while land-use refers to activities on land. This study considers built-up area as a land-cover class composed of impervious surfaces², grass, scattered trees and small gardens surrounding buildings: urban cover. Non-built

²Impervious surfaces are mainly constructed surfaces — rooftops, sidewalks, roads, runways, and parking lot — covered by impenetrable materials such as asphalt, concrete, and stone [5].

area land-cover class includes: vegetation, cropland, bare-soil and water. The terms built-up area and urban area are therefore used synonymous in this study.

It is important to consider patterns of built-up areas occurring in an urban setting, for instance urban sprawl. Similar to urban area, urban sprawl has several definitions. Rashed et al. [see 36, chap. 2] adopts a physicalist definition due to difficulties inherent in measuring and characterizing urban sprawl. They define urban sprawl as the rapid and uncoordinated growth of urban settlements at urban-rural frontier, associated with modest population growth and sustained economic growth. Galster et al. [13] classified patterns characterizing urban sprawl with respect to: *density*, *continuity*, *concentration*, *clustering*, *centrality*, *nuclearity*, *land-use mix* and *proximity* as shown in Figure 2.3. These patterns depicts a correlated nature of development that is a property of built-up areas. Modelling such spatial dependencies can enhance detection of built-up areas.

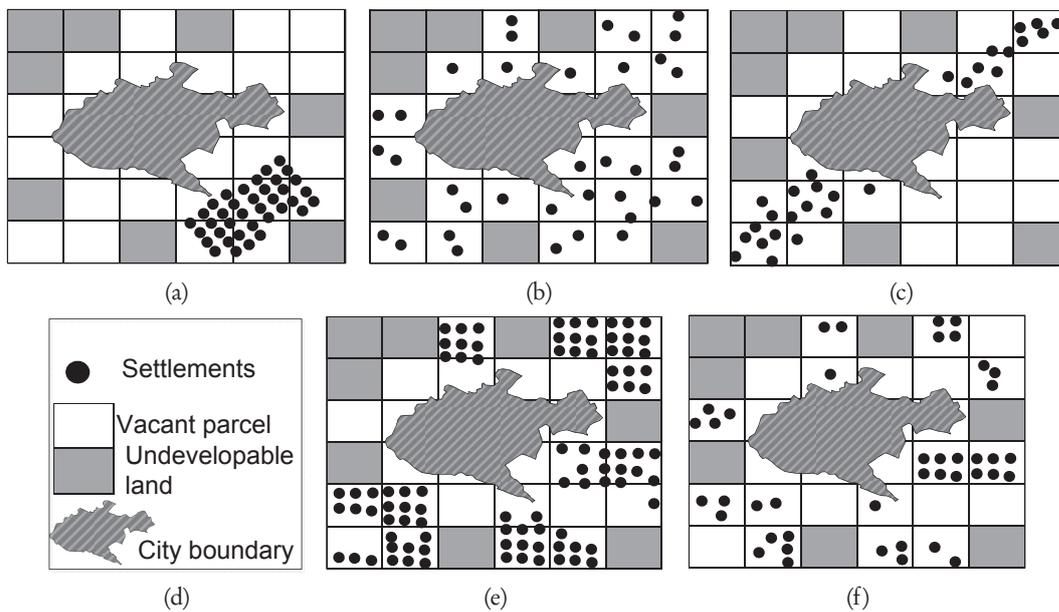


Figure 2.3: Physical patterns characterizing built-up area development. (a) Compact development. (b) Scattered development. (c) Linear strip development. (d) Legend. (e) Leapfrogging development. Modified from [13].

Built-up areas have high spectral variability due to heterogeneous land-cover. This presents a challenge to conventional image processing algorithms and techniques [see 48, chap. 1]. Parametric approaches like MLC perform well in homogeneous areas. In heterogeneous areas, MLC produces noisy results with "salt and pepper" like appearance [18]. This can be attributed to MLC limitation caused by frequency distribution assumptions [see 42, pg. 61]. Better classification can be achieved when within-class variation is less than between-class variation. In built-up areas, within-class variance is high due to high spectral variability of land-cover classes hence parametric classifiers fail. While built-up area heterogeneity is a challenge to basic classifiers, the nature of such areas can be exploited for optimal classification. For instance, [31] assessed the performance of pixel based, segmentation based and spectral-spatial combined classification approaches in heterogeneous urban landscapes. Pixel based method performed poorly while the inclusion of spatial information improved land-cover classification. Thus, making use of spatial information (spectral and spatial), through incorporation of non-spectral information [42, pg. 62] like spatial dependencies in built-up areas, can improve detection. Moreover, use of complementary information from images of different sensors, i.e. SAR and optical sensors, has indicated improved classification

results.

SAR and optical sensors deliver contrasting yet complementary information. Optical sensors record reflected and thermal radiation of the earth surface objects. Similarly, SAR sensors detect emitted and reflected microwave radiation from natural and man-made objects. Use of SAR and optical images enhances exploitation of complementary information [43]. Optical images are easier to interpret visually while geometric perturbations in built-up areas and speckle complicates interpretation of SAR images. Nonetheless, SAR sensors overcome some major drawbacks of optical sensors: their signals can penetrate clouds and are independent of daylight [14, 43]. SAR images convey information about geometric configurations (structure), texture and dielectric properties of man-made features in urban areas [10, 2]. These attributes emphasize objects appearing with low contrast in the optical image counterpart like man-made objects in urban areas [43]. This is echoed by [1] whose results illustrate that SAR data enhanced discrimination of urban areas. SAR is typically characterized by speckle, layover, foreshortening and shadows [43]. These effects give additional information not available in the optical images. For instance, in [17] layover and shadows from SAR were used to extract buildings. The work of [2] demonstrated that multi-source information and a good classification method improved urban land-cover classification. The authors also recommend the use of spatial properties of classes in places such as built-up areas which have similar spectral properties. SAR data holds a potential for analysis of built-up area structural characteristics, however, some effects like speckle and shadows hamper automated classification. Integration of spatial dependencies of built-up areas can be used to overcome this challenge [11].

2.3 BUILT-UP AREA DETECTION STUDIES USING SPATIAL CONTEXT

Classification techniques using contextual information for image segmentation and classification have recently gained popularity [42]. Conventional MRF has been used in many studies but recent CRF has become more popular. Several studies have used CRF for urban related mapping using remotely sensed images. Hoberg et al. [18], designed a simple binary CRF model used to classify settlements in a rural area using IKONOS image. Their method outperformed the standard MLC approach but was unable to distinguish between forest and settlements. They recommend extension of CRF framework to incorporate: prior segmentation results, spatial as well as temporal context and more than two classes during classification. The study motivated extension of CRF by a time-dependent parameter to incorporate temporal differences in land cover classification [19].

Zhong et al. [49] developed a multiple CRF (MCRF) ensemble learning classifier that utilized five groups of texture features from Quickbird and SPOT to detect urban areas. Texture features were fused by multiplicatively combining their conditional distribution with MRF assigning labels. In an earlier study [50], they integrated MRF with CRF to learn from multilevel structural information using gradient magnitude, gradient orientation and line length features to detect urban areas. A challenge of redundant features was considered in the work of [27] where a feature selection strategy was used in detection of urban areas. Another study by [7], utilized Gaussian MRF (GMRF) to incorporate texture in delineating urban areas using complementary features from SAR and optical sensors. The study proved suitable for urban growth monitoring which is essential in obtaining new information on built-up areas. This inspired another study to evaluate the performance of multi-parameter SAR data for delineating urban areas using a similar method [6]. Similar GMRF texture method by [28] proposed a parameter estimation approach for extracting urban areas from images of different resolutions.

Some studies have been done using CRF to detect buildings in urban areas. Wegner et al. [46],

applied CRF to detect buildings from one orthophoto and Interferometric SAR (InSAR) data using orthophoto colour features and SAR texture features. The outcome demonstrated that CRF performed well compared to standard MLC and MRF. Combined use of InSAR and optical features was also significant. This inspired another study where use of irregular and regular image grid structure for building detection was evaluated [47]. The irregular grid structure reduced computation time significantly and produced better results. In the work of [17], CRF was used to extract buildings from Polarimetric SAR data. Layover and shadow of buildings proved promising for building analysis application.

2.4 SUMMARY

It is evident that levels of development in built-up areas are mutually dependent. Most urban area studies have demonstrated that non-spectral information can improve classification accuracy. Spatial information and use of multi-sensor data are among proposed approaches. Spatial information integrate spatial dependencies of features while multi-source information exploits complementary properties of different images during classification. Spatial dependencies can be incorporated using contextual classification methods. Classical MRF integrate spatial dependencies in labels only ignoring existing dependencies in images. In contrast, CRF theory demonstrates its robustness to handle diverse dependencies in images and class labels. Therefore, exploiting the proposed CRF framework can improve classification results.

Chapter 3

Materials

This chapter provides a description of selected study area, data used for analysis including reference data and software.

3.1 STUDY AREA LOCATION

Study area selected is located in Nairobi province, Kenya. It lies between $36^{\circ}39'46''$ E– $37^{\circ}06'07''$ E and $1^{\circ}09'19''$ S– $1^{\circ}26'27''$ S. The area covers a fast growing city engulfed in up coming built-up areas commonly referred to as "satellite towns". An area extending to the western part of the city was selected as illustrated in Figure 3.1. It contains a complex mixture of heterogeneous land-cover materials encompassing a transition between built-up and non-built-up areas which challenges conventional classification. The area proved suitable for developing and evaluating the performance of CRF method in built-up area detection.

3.2 DATA

Data used for urban studies must meet certain spectral, spatial, radiometric and temporal characteristics. An evaluation by [48] demonstrated that improved spectral and radiometric resolution are helpful in achieving better classification accuracies as compared to improved spatial resolution. Increased spatial resolution not only increases interclass variability but also intra-class variability which challenges classification. In this study medium resolution images were chosen for built-up area detection.

3.2.1 Optical images

ASTER level 1B images were used. The study used bands in the visible and near infrared (VNIR) part of the spectrum with 15 m resolution and 8-bit unsigned integer data type. VNIR consist of three bands namely: Band 1 (green), Band 2 (red) and Band 3 (near infrared). Band 3 consist of two bands: backward-scanning band labeled *Band 3B* which creates parallax and at nadir scanning band labelled *Band 3N*. Band 3B is used in production of stereo view images of the earth useful in developing elevation information. The band was not used in any analysis or classification. Nadir looking band (Band 3N) was used in classification. A total of two images of 2011 and 2006 both acquired in January with 0% cloud cover were used. Figure 3.2a shows part Nairobi city on ASTER image of 2011.

3.2.2 SAR images

ALOS-PALSAR level 1.5 fine beam single horizontal-horizontal (HH) polarized data were used. Level 1.5 products are geo-referenced and can be obtained from a resolution of 6.25 m. SAR images of two time periods acquired in January 2011 and December 2006–January 2007 with spatial

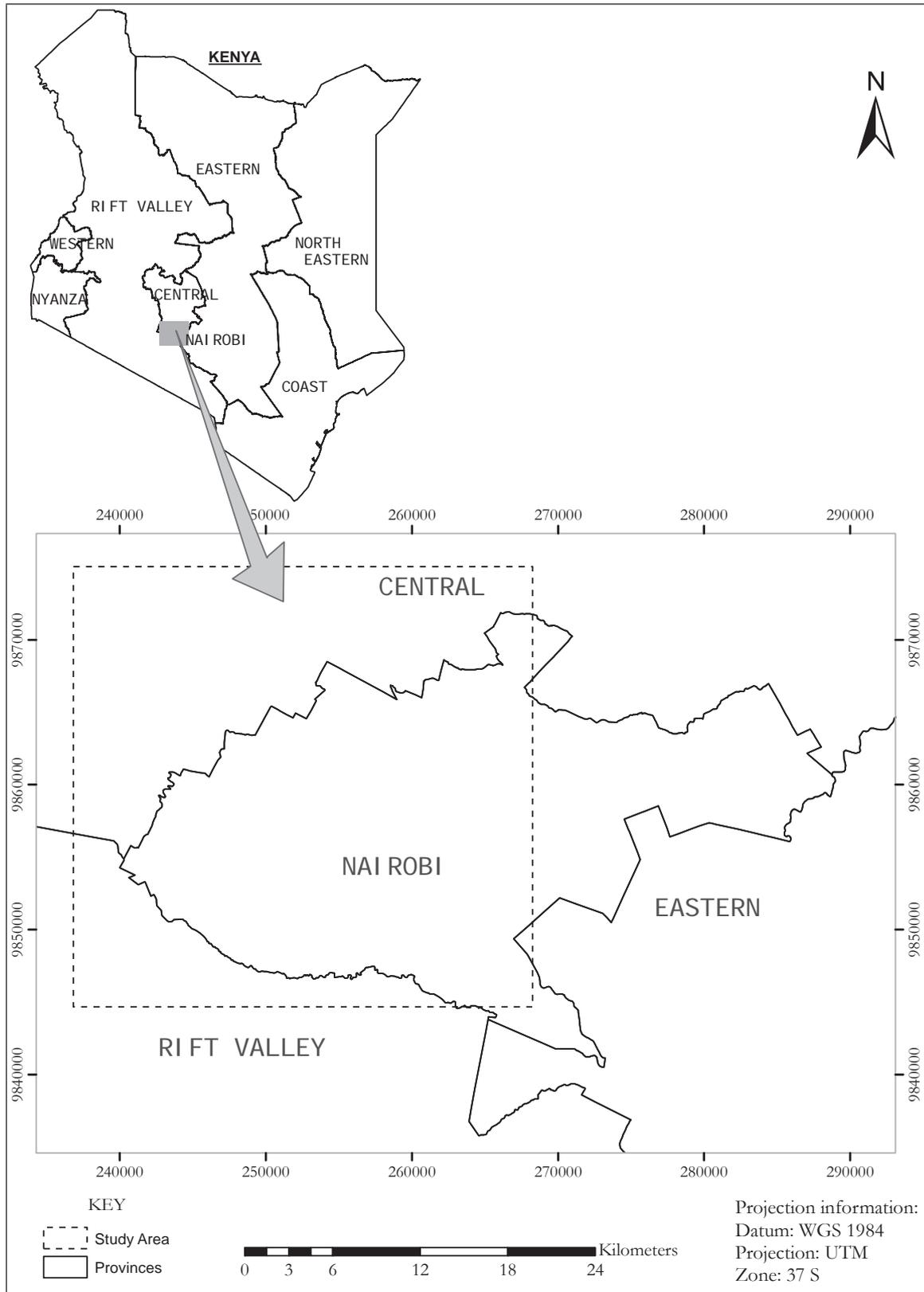


Figure 3.1: Study area location.

resolution of 6.25 m and 12.5 m were used respectively. Figure 3.2b illustrates part of Nairobi city as seen from SAR image.

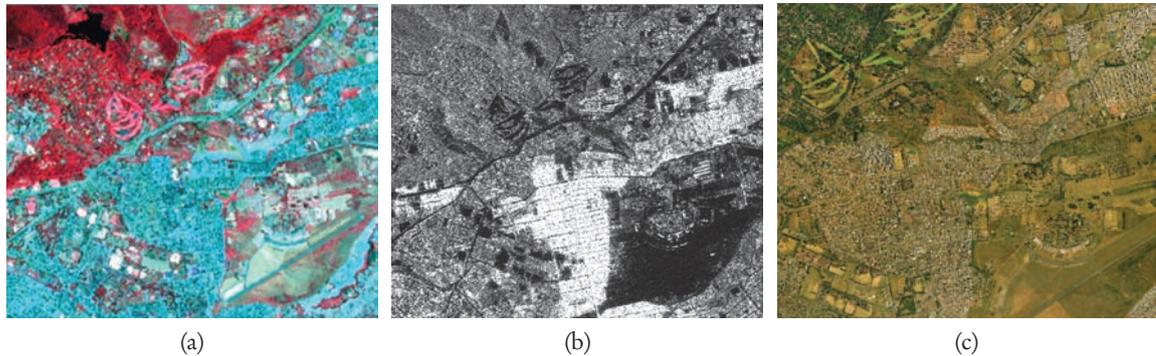


Figure 3.2: Part of Nairobi city as seen from: (a) ASTER false colour composite (3,2,1) image. (b) SAR image. (c) Aerial photo.

3.2.3 Subset images

Figure 3.3 shows subset images used to test designed CRF potentials. The area consist of a complex mixture of impervious surface materials, forest, grass and bare-soil land-cover classes.

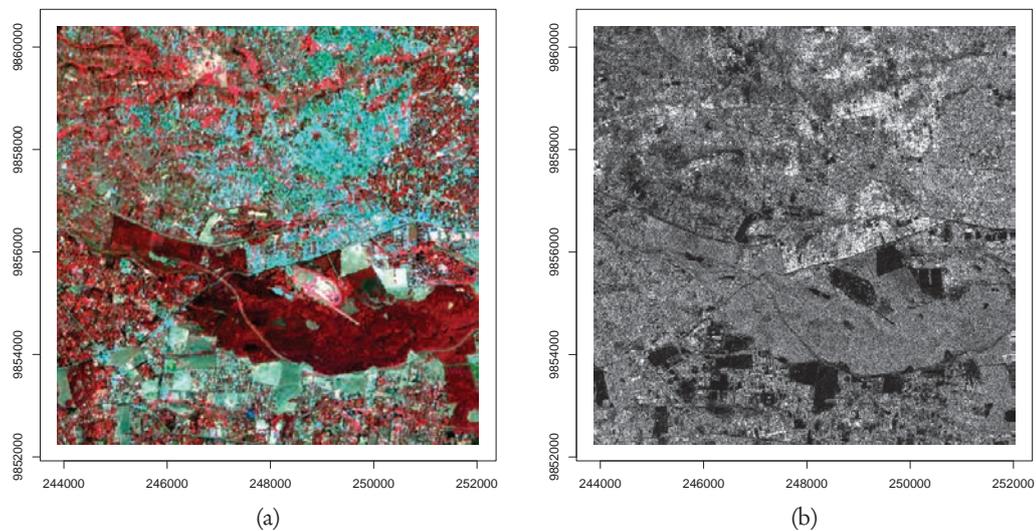


Figure 3.3: (a) ASTER false colour composite (3,2,1) image and (b) SAR intensity image.

3.2.4 Reference data

Vector data, GeoEye 5 m resolution and aerial photography 30 cm resolution data were used as reference data. Nairobi land-use vector data made available in 2010 was used during feature extraction and validation. The data is based on a 2003 base map updated in 2010 by the University of Nairobi, Department of Urban and Regional Planning (DURP). The original dataset was produced by a collaboration between the Center for Sustainable Urban Development (CSUD) and

the Spatial Information Design Lab (SIDL) at Columbia University, generously funded by the Volvo Research and Educational Foundations (VREF). More information on Nairobi land-use vector data can be obtained at [8]. The GeoEye image of 2011 was used to update the vector land-use data. Aerial photographs acquired in 2003 as shown in Figure 3.2c were used for georeferencing the satellite images.

3.3 SOFTWARE

The following software were used in the study:

- ERDAS Imagine 2011: this software was used for preprocessing task which include: image mosaic, geo-referencing, geo-coding and sub-setting the study area.
- ENVI 4.8: this software was used for training area selection and computing general class statistics.
- R programming software version 2.14.0. Most analysis was done in this software including classification with CRF see Appendix A.1 and A.3 for implemented code. Packages used include:
 1. kernlab
 2. rgdal
 3. GEOMap
 4. geoR
- ArcGIS 10: used to prepare the land-use reference data for use in validation.

Chapter 4

Implementation

This chapter gives a sequential explanation of research execution as summarized in Figure 4.1. Details of how CRF framework was extended and used for built-up area detection are presented in Section 4.4.

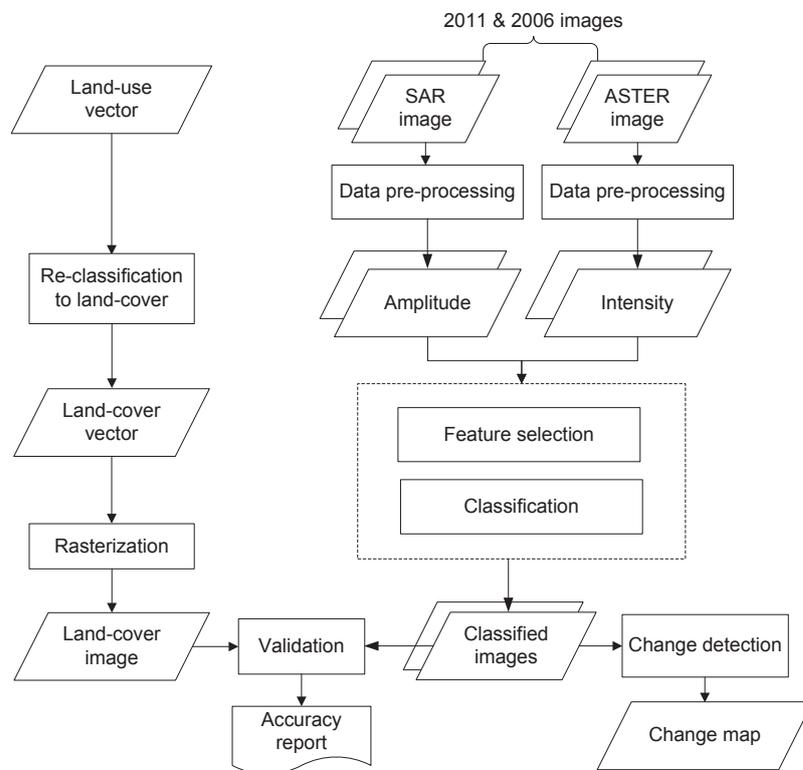


Figure 4.1: Methodological framework.

4.1 DATA PRE-PROCESSING

ASTER data was acquired in TIFF¹. Each band was stored in a separate TIFF file with accompanying metadata in a text file. Bands 1, 2 and 3 (green, red and near infra red) in VNIR² spectrum of ASTER were stacked to produce a false colour composite (3,2,1) image of the area. The false colour composite image was used for visual analysis, which facilitated definition of classes and identification of suitable features to discriminate them. ALOS-PALSAR images acquired in

¹tagged image file format

²visible and near infrared

CEOS³ format were imported into TIFF and mosaicked using ERDAS Imagine 2011. Study area extent was then defined and used to subset ASTER and SAR images.

Pixel coordinates in SAR, ASTER and GeoEye images did not correspond to ground locations. This is normally attributed to effects such as sensor geometry distortions, platform instabilities and earth rotation. Some of these distortions are corrected by data providers. However, pixel locations in the acquired images did not match. The images were geo-referenced to a high resolution aerial photograph to ensure co-registration. Geo-referencing was done using ERDAS Imagine 2011. An affine transformation with evenly distributed Ground Control Points (GCPs) in the study area was used. Pixels of SAR images were re-sampled to 15 m resolution using nearest neighbour method. This allowed direct comparison of ASTER and SAR data on a pixel by pixel basis. Table 4.1 gives a summary of Root Mean Square Error (RMSE) in pixels obtained and the number of GCPs used in geo-referencing the images.

Table 4.1 Image geo-referencing accuracy summary.

Image	Resolution (m)	Year	RMSE (pixels)	Number of GCPs
ASTER	15.0 m	2011	0.092	21
SAR	6.25 m	2011	0.063	21
SAR	12.5 m	Dec 2006–Jan 2007	0.066	21
GeoEye	5.0 m	2011	0.076	10

ASTER intensity was computed using three bands: $\text{intensity} = (NIR + R + G) / (3 \times 255)$. A square root transformation was used to transform SAR intensity to amplitude. Features computed from the data were standardized by dividing the difference of the features from the mean with their standard deviation.

The last phase of data pre-processing involved preparation of reference data. First, land-use classes in the vector reference data were reclassified into built-up and non-built-up land-cover classes. This was followed by an update of the reclassified vector data using GeoEye image. Reference images at relevant scales were then created through a vector to raster conversion. Polygon to raster option in conversion tools of ArcMAP 10 was used. Pixels whose cell centers and maximum areas overlapped the polygons inherited their attributes.

4.2 CLASS DEFINITION

A visual analysis was done on the pre-processed images to identify built-up area land-cover classes. Two main super-classes: built-up and non-built-up areas were used. The following sub-classes were defined in each super-class:

- **Built-up area**- composed of three categories:
 1. High density built-up: high proportions of impervious surfaces with little or no vegetation.
 2. Medium density built-up: mixed proportions of impervious surface materials and vegetation.
 3. Low density built-up: high proportions of vegetation with little impervious surfaces.

³Committee on Earth Observation Satellites

4. Tarmac and pavements.
 - **Non-built-up area-** composed of four categories:
 1. Vegetation: forest, grass and shrubs.
 2. Cropland.
 3. Bare-soil.
 4. Water.

The study assumes that built-up areas are composed of impervious surfaces and proportions of vegetation. Ambiguities exist with such definition, for instance, as to whether there is a park or swimming pool in a built-up area. However, the aim of the study is to delineate land-cover and not land-use as discussed in Section 2.2.

The classes were defined using blocks of merged pixels in a defined window size. Three window sizes: scale I, II and III of pixel size 16×16 , 8×8 and 4×4 respectively were tested in an experiment area. Scale II was adopted in order to conform to a minimum requirement of one hectare in land-use/land-cover mapping [3]. Dividing the image into blocks reduced noise due to spectral variability and speckle. Use of image blocks also introduces a human concept in image classification. Human interpreters normally identify classes in an image using observation windows rather than individual pixels [20].

4.3 FEATURE SELECTION

Several combinations of features were identified, tested and optimal ones selected through experimental classification. The features formed part of CRF framework; the unary feature vectors $f(d_i)$ used in both A and I . Hue, intensity and amplitude images were chosen and discriminative ability of their features tested. Mean, standard deviation and variogram (slope) features were computed from the image blocks.

Variability of pixel values in the images was influenced by their spatial locations. A pair of neighbouring pixels were more likely to have similar spectral values as opposed to a pixel elsewhere in the image. This spatial dependency was well described using a variogram. The variogram related variance/semi-variance (γ) to distance/lag (h) which illustrated variability of pixels at given spatial scales. It was used to extract information on local variability of pixels used for characterizing spatial dependencies. Variogram features, viz: range (r), sill (σ^2), slope at origin (θ) and nugget (n), have been proposed for training set characterization [see 9, chap. 6]. The study adopted slope features computed from ASTER intensity only. Slope features were computed as follows:

1. The intensity images (2006 and 2011) were divided into blocks of pixels.
2. Coordinates and spectral values corresponding to each pixel in a block were defined.
3. An omnidirectional variogram was computed for each block using Cressie's robust estimation method.
4. Slope was then computed by fitting a linear regression model on the first three points of semi-variance and lags as shown in Figure 4.2 and Equation 4.1.

Three points were chosen for two reasons. First, they represent a better linear model fit for approximating slope at origin of a block. Second, some sample variograms reached a partial sill on the third point as shown in Figure 4.3.

In total, three features: mean $f_{\mu i}(d)$, standard deviation $f_{\sigma i}(d)$ and slope at origin $f_{\theta i}(d)$, were used for training set characterization. Mean and standard deviation features were used in both the association and interaction potentials while slope features were included in the interaction potential only. Slope features were exempted in scale III as it is recommended to have at least 50 points in order to compute a sample variogram [see 45, sec. 6.1.2]. Use of slope in the interaction potential incorporated spatial dependencies of built-up areas during classification.

$$\gamma = \theta h + n \quad (4.1)$$

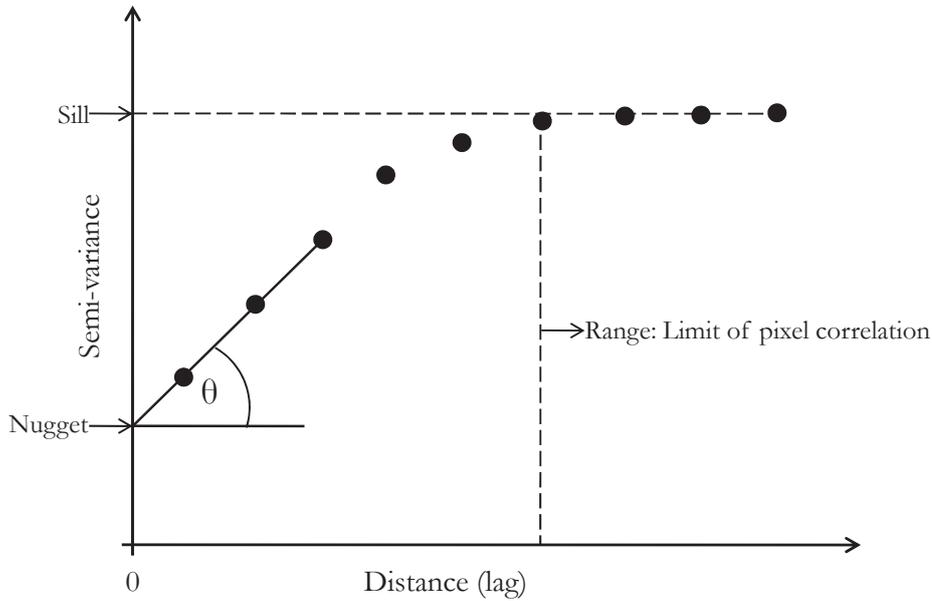


Figure 4.2: Slope (θ) computation using first three points of lag and semi-variance from the variogram.

4.4 CONDITIONAL RANDOM FIELDS FOR BUILT-UP AREA DETECTION

Conditional random fields classification framework was designed in order to model spatial dependencies. This involved design of association and interaction potentials. The framework was then used in a supervised classification approach to predict class labels of data instances in the training set given their features.

4.4.1 Association potential (A)

Conditional random fields framework offers much flexibility because local discriminative classifiers can be used to determine the A and I . Most studies, [18, 50, 49], have used logistic classifiers to determine local class probabilities. Logistic classifiers are prone to over-fitting when subjected to large number of features and are not capable of handling redundant data which is a characteristic of built-up areas. Li et al. [27] used SVM recursive feature elimination for feature selection in an attempt to reduce redundancy. In the study, the association potential was designed using SVM. SVM makes no assumptions on class distribution and performs well when subjected to large amount of features. SVM performs classification by building an optimal separating margin

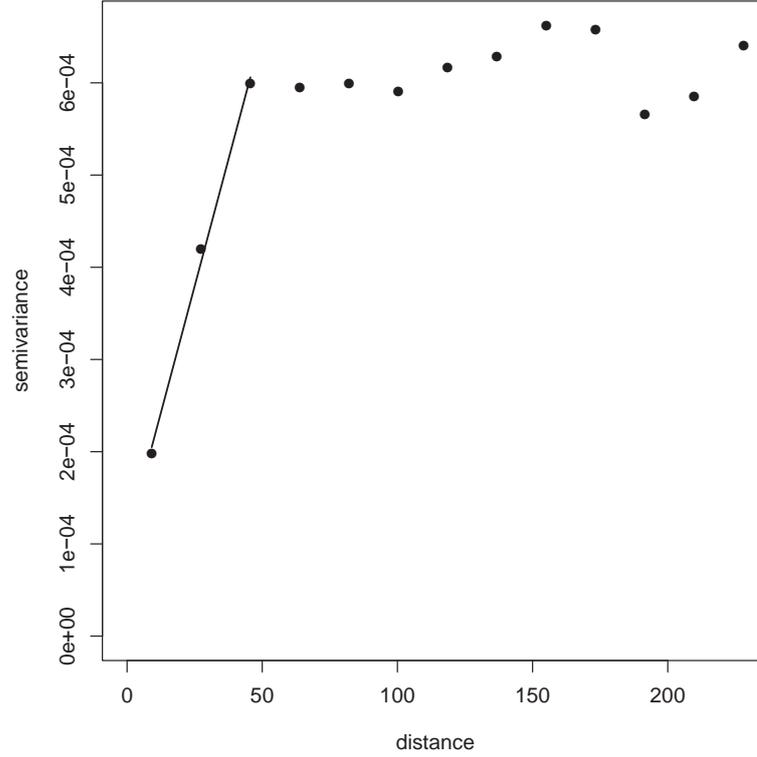


Figure 4.3: Sample variogram with a linear model fit on the first three points.

in the feature space of provided classes [42]. Therefore, it separated built-up and non-built-up classes from training data using a hyperplane corresponding to the decision function:

$$g(f(d)) = m(f(d)) + b \quad (4.2)$$

where m is the normal to the hyperplane, b denotes the bias, $f(d)$ is 4 dimensional site-wise feature vector of: mean and standard deviation features from intensity and amplitude. The hyperplane constructed by SVM is a decision boundary that separated classes in the feature space.

SVM offers linear and non-linear decision boundaries. A non-linear decision boundary was constructed using Gaussian Radial Basis Function (RBF) kernel. The boundary considered data not separable using a linear hyperplane. It used a function Φ to map the feature vector $f(d)$ into $\Phi(f(d))$, a high dimensional feature space (H) [see 42, chap. 4]. This spread the distribution of the training features in a way that facilitated fitting of a linear hyperplane. The decision function in Equation 4.2 can be expressed as:

$$g(f(d)) = \mathbf{sign} \left(\sum_{i=1}^k \alpha_i w_i K(f(d), f(d_i)) + b \right) \quad (4.3)$$

where α_i are positive Lagrangian multipliers, w_i are class labels, k is the total number of features computed for each site i , $K(f(d), f(d_i))$ is the RBF kernel function and \mathbf{sign} is a discriminant function defined as $\{1, -1\}$ corresponding to built-up and non-built areas respectively.

The non-linear RBF kernel has two parameters that were determined during training: penalty (C) and sigma (σ). Class posterior probabilities of A , $P(w|d)$, were then generated from decision values using the function in Equation 4.4 proposed by [34].

$$P(w = 1|g) = \frac{1}{1 + e^{Ag+B}} \quad (4.4)$$

where g are decision values generated from SVM's decision function in 4.2, A and B are parameters that were estimated by fitting the negative log-likelihood function.

4.4.2 Interaction potential (I)

Designing I using a data-dependent term and a constant smoothing parameter (β) of the Ising model has been proposed by [23] among other studies. The design incorporates a distance function to model the data dependent term. In the study, the data dependent smoothness parameter is modelled as a function of inverse of transformed euclidean distance (TED) of a pair of block features in a first order neighbourhood system. The concept is extended from [41] where transformed divergence, as a class separability measure, was used in determining an optimal smoothness parameter in MRF based super resolution mapping. The design generates two scenarios: First, when separability value is low, then a pair of neighbouring image blocks belong to the same class. The smoothing parameter is increased so that the blocks are aggregated into the same class. Second, when separability value is large, then the pair of image blocks belong to different classes. In this case the smoothing parameter is decreased to maintain the blocks in different classes. This design was used as a data dependent discontinuity adaptive model that moderated smoothing when data from neighbouring blocks were "dissimilar". The indiscriminate smoothing nature common with MRF was thus controlled using the designed model. Equation 2.17 was extended as:

$$I(w_i, w_j, d) \equiv \beta t^{-1} \delta(w_i, w_j) \text{ where:} \quad (4.5)$$

$$\delta(w_i, w_j) = \begin{cases} 0 & \text{if } w_i = w_j \\ 1 & \text{if } w_i \neq w_j \end{cases}$$

where β is a constant smoothness parameter and t (transformed euclidean distance) is defined as:

$$t = 1 - \exp \left\{ - \sum_{i=1}^{i=N_i} \left((\Psi(d_{ij}))^2 \right) \right\} \quad (4.6)$$

where N_i are neighbours of site i and $\Psi(d_{ij})$ is a 5 dimensional pairwise feature vector of: intensity slope features, mean and standard deviation features of intensity and amplitude. $\Psi(d_{ij})$ is obtained by subtracting two feature vectors ($f(d_i)$ and $f(d_j)$) of a pair of neighbouring blocks i and j as shown below:

$$\Psi(d_{ij}) = f(d_i) - f(d_j) \quad (4.7)$$

The term $\delta(w_i, w_j)$ in Equation 4.5 compared a pair of neighbouring labels w_i and w_j . Dissimilar labels were penalized while preserving corresponding labels and edges (class boundaries). The magnitude of penalization was controlled by the discontinuity adaptive model in Equation 4.6. Therefore, by substituting the designed models of A and I , the adopted CRF model can be expressed as:

$$P(w|d) = \frac{1}{Z(d)} \exp \left\{ - \sum_{i \in S} \log \frac{1}{1 + e^{Ag+B}} - \frac{1}{2} \sum_{i \in S} \sum_{j \in N_i} I(w_i, w_j, t^{-1}) \right\} \quad (4.8)$$

Multiplying the interaction potential by half accounts for double counting of labels in the neighbourhood N_i of a block.

Performance of the designed data dependent model (t^{-1}) was compared to: euclidean distance e and absolute difference a as illustrated below:

$$e = \sum_{i=1}^{i=N_i} (\Psi(d_{ij}))^2 \quad (4.9)$$

$$a = \sum_{i=1}^{i=N_i} |\Psi(d_{ij})| \quad (4.10)$$

4.4.3 Training of designed association potential

Training CRF model designed and presented in Equation 4.8 involved determination of (C, σ) RBF kernel parameters from training samples for use in classification. Optimal parameters that minimized the overall training error were determined. Parameter values (2,2), (2,2) and (10,10) were used during experiments with scales I, II and III respectively. In application of CRF in the study area, parameter values (1,1) were used in classification of 2011 and 2006 images. Figure 4.4 illustrates the training error attained with corresponding parameters values.

The training process was done simultaneously with visual assessment of classified image. Not all combination of parameters with minimum training error gave accurate results. Therefore, the quality of classified results influenced the final choice of parameters.

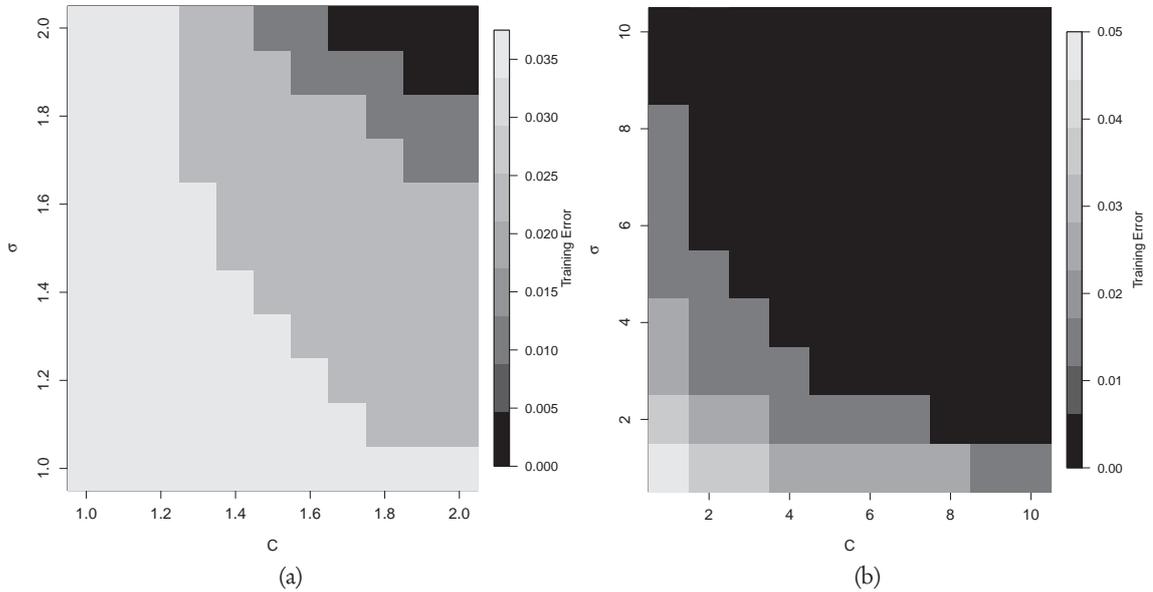


Figure 4.4: Training of association potential to determine most optimal values for cost (C) and sigma (σ) parameters. (a) Scale I and II. (b) Scale III.

A total of 80 training blocks, 40 each for built-up and non-built-up areas, were used. The training set blocks were collected from a larger scene image and used in classifying study area and subset images. Figure 4.5 shows 40 training blocks, 20 each in built-up and non-built-up class, falling inside the study area. The remaining 40 blocks were collected outside the study area. This was done to avoid model over-fitting due to biased collection of training samples.

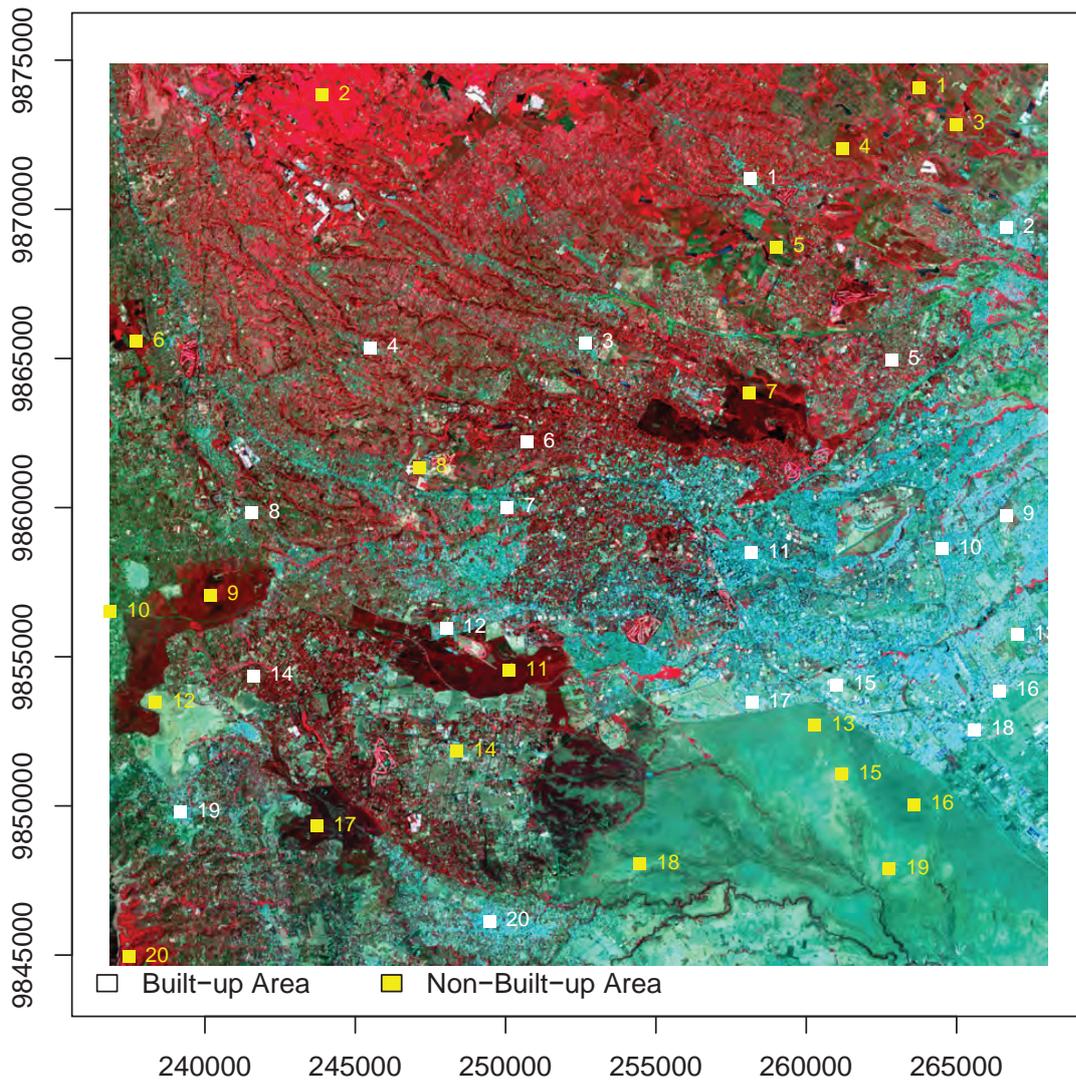


Figure 4.5: ASTER false colour composite (3,2,1) image showing study area training samples collected using scale II blocks.

4.4.4 Classification of built-up areas using CRF

Classification task involved determination of optimal class labels w from input image blocks given training features. This was done by determining MAP estimate that maximized the posterior probability $P(w|d)$ in Equation 4.8 which, is equivalent to minimizing the posterior energy:

$$\hat{w}_i = \arg \min_{w_i} U(w_i|w_{N_i}, d) \equiv \left(U(w_i|f(d_i)) + \frac{1}{2}U(w_i = w_j|t^{-1}, N_i) \right) \quad (4.11)$$

The solution of Equation 4.11 required a specifically designed estimation approach. This was done iteratively using ICM. ICM is a local optimization method which converges at local minimum of the defined energy function in Equation 4.11. The minimum energy defined the maximum probability of a given class label. The procedure used in optimization is shown below:

Optimization using ICM:

1. Number of iterations N_{iter} and first order neighbourhood system N_i using cliques of the energy function in Equation 4.11 were defined.
2. Block labels were initialized from association potential solution.
3. For each block i , the energy for every class (built-up and non-built-up) w'_i was estimated, and w_i updated by the minimum energy w'_i .
4. Step 3 was repeated N_{iter} times.

The implementation of designed CRF potentials and optimization using ICM are provided in the Appendix Section A.1 and A.2 respectively.

4.4.5 Test of CRF interaction potential data dependent models

Classification experiments were conducted using the subset images in Figure 3.3 to determine a suitable data dependent model of CRF. Overall classification accuracy of the models were compared against values β in a logarithmic scale. The experiment was conducted using the three block scales (I, II and III) with 2011 images. Inverse of transformed euclidean distance performed better than normal transformed divergence, euclidean distance (Equation 4.9) and absolute difference (Equation 4.10) of site-wise features. A threshold was used to exempt zero values in the denominator of t^{-1} that would have otherwise led to infinite values.

4.5 APPLICATION OF CRF IN DETECTING BUILT-UP AREAS INCLUDING TEMPORAL CHANGES.

The designed CRF model using inverse of transformed euclidean distance as a data interaction term was used to classify built-up areas using 2011 and 2006 images. Inverse of transformed euclidean distance was selected based on insights from the experiments conducted in a complex area described in Section 3.2.3. Scale II was used in order to conform to approximately one hectare land-cover classification requirement. A constant smoothness parameter $\beta = 1$ was used. Cost and sigma parameters (1,1) were used in the association potential to classify 2011 and 2006 images respectively. Reference information available corresponded to date 2011 and thus 2006 classification was not validated. However, a visual assessment was done on the classified image.

In order to detect temporal changes, a change detection using 2011 and 2006 classified images was done. Image-differencing method was used where the later image was subtracted from the earlier image. Built-up area class was re-assigned label 1 and non-built-up area class label 2 from the original binary labels (1, -1). After image differencing, blocks with labels 0, 1 and -1 referred to no change, non-built-up to built-up and built-up to non-built-up changes respectively. Implementation code is provided in Appendix A.3.

4.6 ACCURACY ASSESSMENT

A quality assessment technique was essential in evaluating the performance of the designed classification method. The outcome of this assessment enabled a degree of confidence to be attached to the results. A commonly used accuracy assessment method in remote sensing, confusion matrix/error matrix, was employed for this purpose. The error matrix shows the proportions of correctly classified (overall accuracy) and misclassified image blocks in a table. Therefore, several accuracy measures were derived from it [see 21, chap 6.2.4].

Error matrices were generated by comparing each classified pixel against the reference image. Overall accuracy, false positives, false negatives and kappa statistics were used as quality measures. False positives and false negatives are synonymous to type I and type II errors which indicate the proportions of image blocks omitted and incorrectly classified respectively. They are a consequence of producer and user accuracies. Kappa statistics was useful in evaluating different remote sensing methods since it accounted for the degree of accuracy that were attained when labels were assigned at random.

Chapter 5

Results

Results of research objectives implementation are presented sequentially in this chapter.

5.1 FEATURE SELECTION

Standard deviation features clearly outline areas of high spectral variability such as built-up areas. Mean features delineate homogenous regions such as forest, grassland and cropland. These appearance is the same in all experiment scales. The images in Figure 5.1 depicts the appearance of mean and standard deviation features derived from intensity and amplitude images using scale II. These features capture homogeneity and variability of land-cover classes respectively. This implies, mean features accounts for inter-class spectral variability while standard deviation features describes intra-class spectral variability; hence defining dissimilarity and similarity of classes respectively.

Mean and standard deviation feature space plots using scales I, II and III are illustrated in Figure 5.2. The plots show that the built-up and non-built-up classes are not completely separable. Separability of features derived from intensity increases with reduction in block size as shown in Figures 5.2a, 5.2c and 5.2e. In contrast, separability of features derived from SAR amplitude decreases with reduction in block size as shown in Figures 5.2b, 5.2d and 5.2f. Decreasing separability of amplitude derived features is a consequence of speckle noise whose influence increases on reduction of block size; see Figure 5.4.

Table 5.1 shows kappa and overall accuracy obtained by including and exempting hue (colour) information. Combining hue, intensity and amplitude reduces classification quality as demonstrated by kappa and overall accuracy percentages. Decreasing block size further degrades the classification quality. In contrast, classification quality remains high in all scales when amplitude and intensity features are used. There is a slight decrease in accuracy when block size is reduced though, classification quality remains high compared to inclusion of hue features.

Contribution of SAR amplitude and ASTER intensity in the overall classification was also evaluated as shown in Table 5.2. Amplitude gave high overall accuracy and kappa values in scale I. The classification quality decreases with reduction in block size due to increase in speckle noise. On the contrary, reduction in block size increases classification quality when only intensity image is used.

Table 5.3 summarizes the outcome of slope features discriminative ability assessment. On average, built-up areas have higher slope values compared to non-built-up areas. The slope of high density built-up areas, water and forest derived from amplitude is zero. The shape of the variogram of these classes, Figure 5.3, illustrate a nonzero intercept termed as "nugget effect" which has zero slope value. Such variogram shape suggest that the data lack a spatial structure as a result of random noise. Intensity image gives positive slope in all the subclasses. Figure 5.4 show images of densely built-up area and forest used to compute slope. Amplitude image blocks show noisy pixels with little or no spatial structure compared to intensity which show a spatial structure.

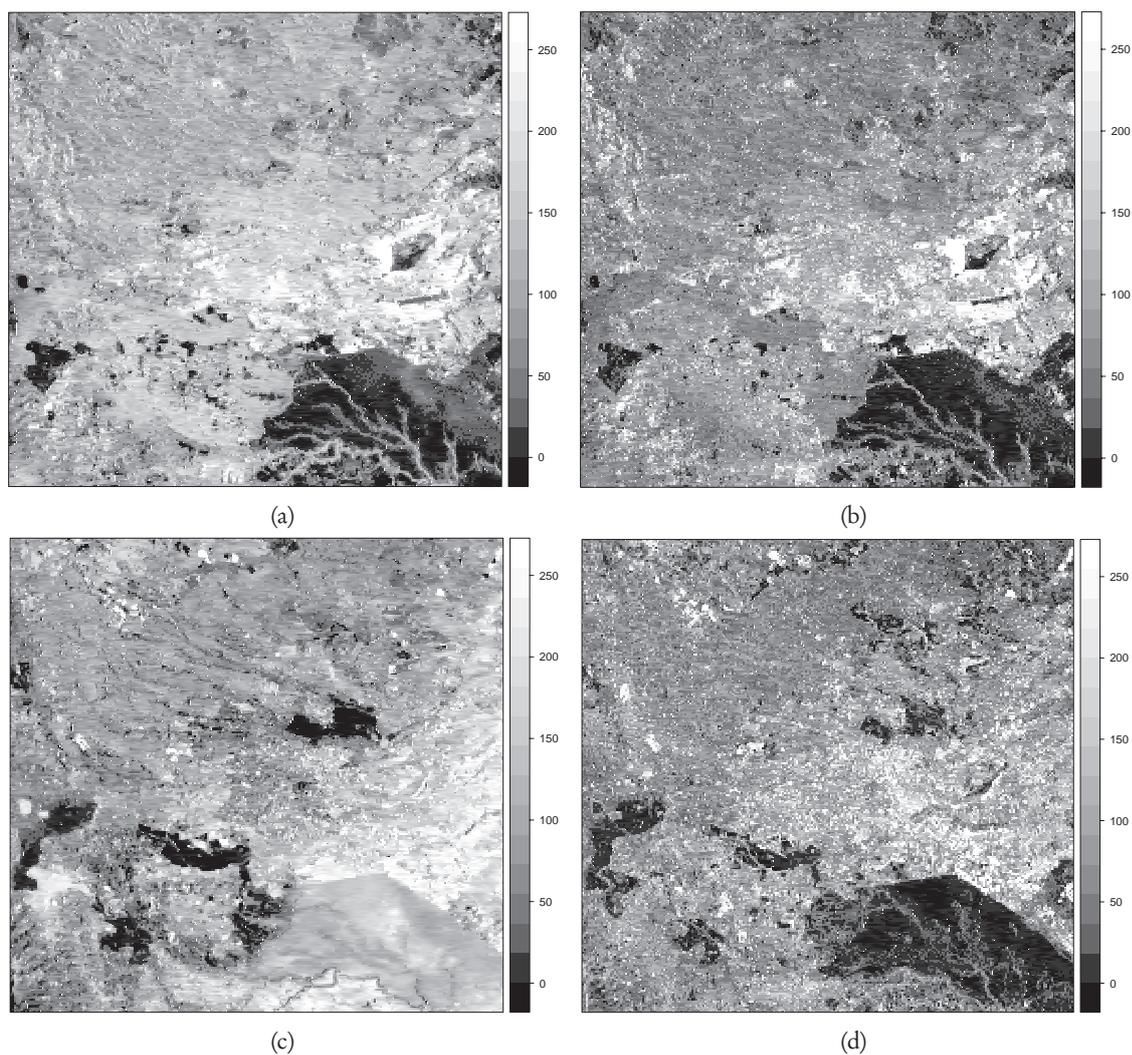
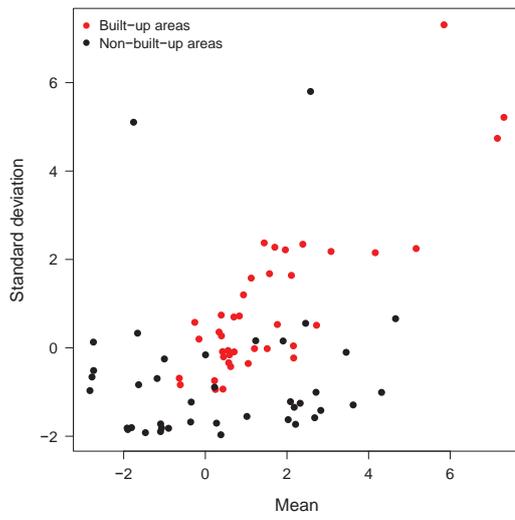


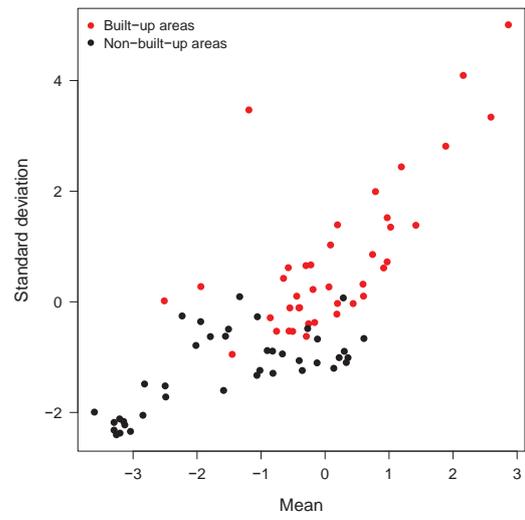
Figure 5.1: Mean and standard deviation image features showing homogeneity and variability of classes respectively using scale II. (a) Mean of amplitude features. (b) Standard deviation of amplitude features. (c) Mean of intensity features. (d) Standard deviation of intensity features.

Table 5.1 Summary of evaluation of suitable combination of features from hue, amplitude and intensity.

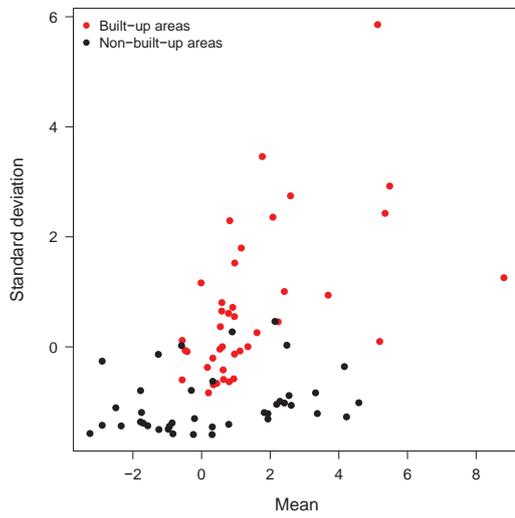
	Hue, intensity & amplitude	Intensity & amplitude
SCALE I		
OA ¹	79.7%	86.2%
Kappa	0.454	0.665
SCALE II		
OA	78.3%	82.2%
Kappa	0.428	0.570
SCALE III		
OA	66.2%	77.8%
Kappa	0.043	0.453



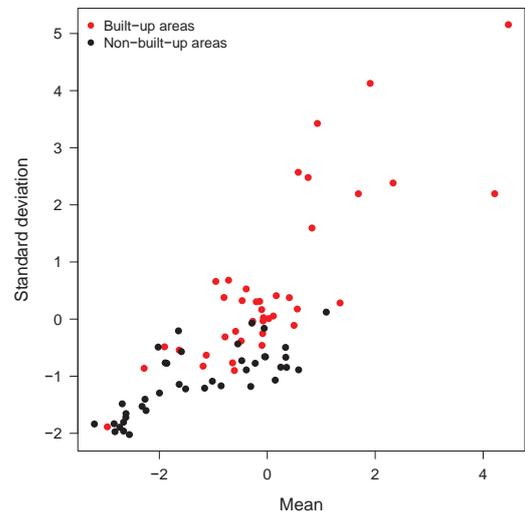
(a) Scale I feature space using ASTER intensity.



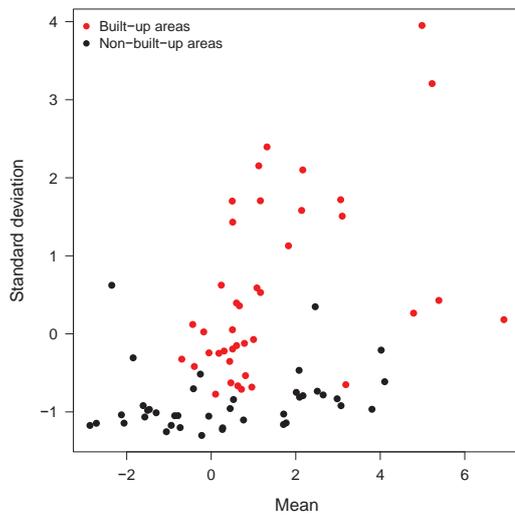
(b) Scale I feature space using SAR amplitude.



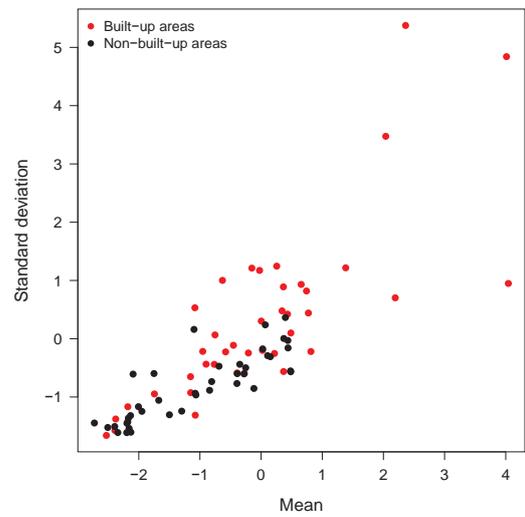
(c) Scale II feature space using ASTER intensity.



(d) Scale II feature space using SAR amplitude.



(e) Scale III feature space using ASTER intensity.



(f) Scale III feature space using SAR amplitude.

Figure 5.2: Amplitude and intensity training blocks feature space plots.

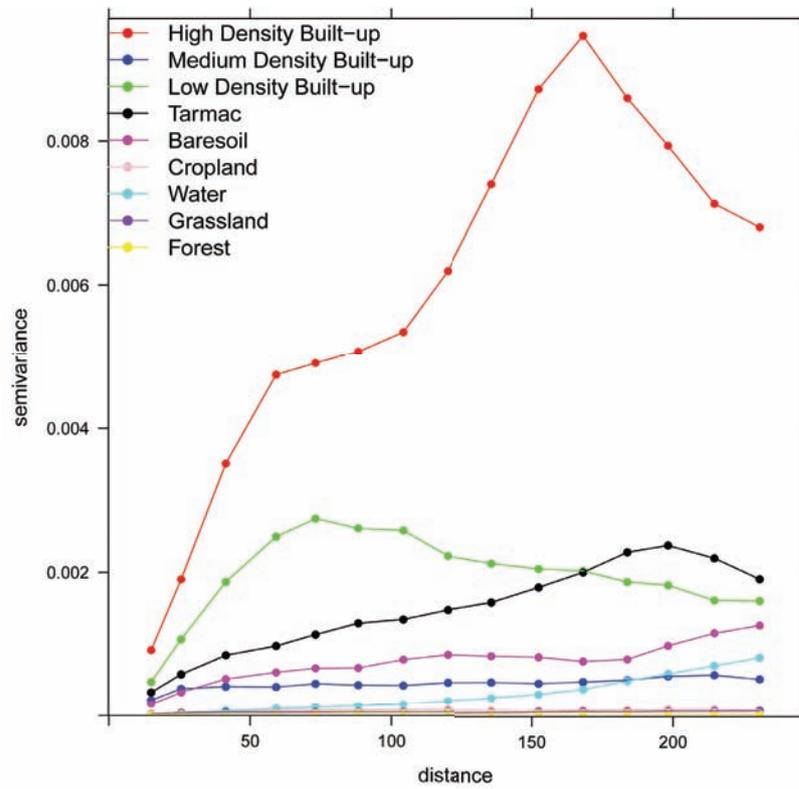
Table 5.2 Summary of evaluation of detection ability of amplitude and intensity features.

	SAR Amplitude	ASTER intensity
SCALE I		
OA	83.1%	72.2%
Kappa	0.622	0.400
SCALE II		
OA	77.6%	74.0%
Kappa	0.503	0.393
SCALE III		
OA	64.6%	78.4%
Kappa	0.268	0.490

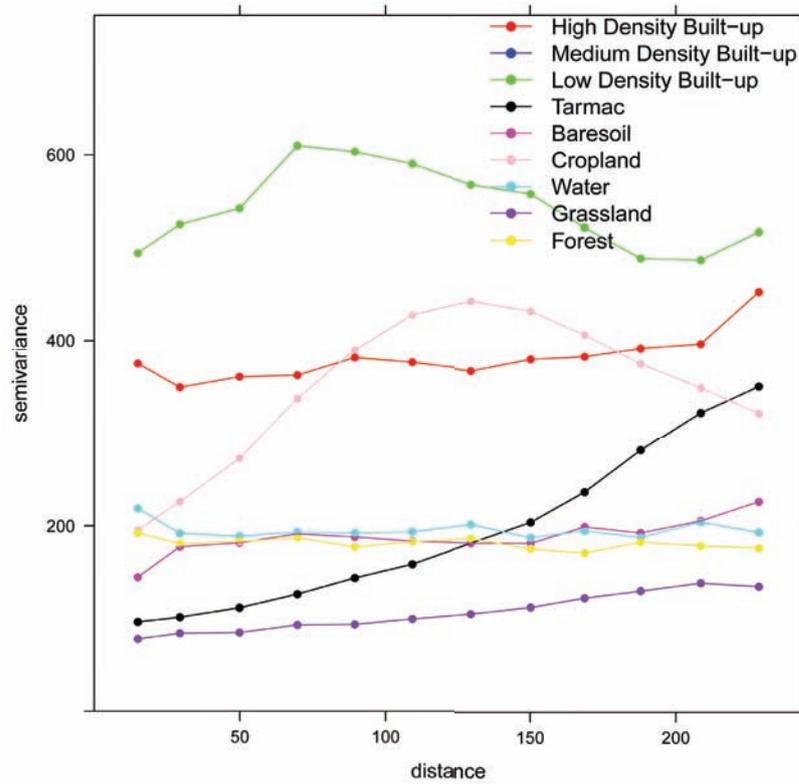
Table 5.3 Summary of assessment of variogram's slope features detection ability using scale I blocks.

Built-up areas		
Class	Intensity slope	Amplitude slope
High density built-up	1.10×10^{-05}	0
Medium density built-up	4.68×10^{-06}	16.9
Low density built-up	4.73×10^{-05}	1.52
Tarmac	1.52×10^{-05}	0.38
Average	1.95×10^{-05}	4.70
Non-Built-up areas		
Baresoil	1.07×10^{-05}	0.93
Cropland	1.15×10^{-06}	1.89
Water	1.67×10^{-06}	0
Grassland	8.25×10^{-07}	0.19
Forest	4.68×10^{-07}	0
Average	2.97×10^{-06}	0.60

¹Overall accuracy (OA).



(a)



(b)

Figure 5.3: Sample variograms of built-up and non-built-up super-classes as defined in Section 4.2. (a) Sample variograms of scale I blocks using ASTER intensity. (b) Sample variograms of scale I blocks using SAR amplitude.

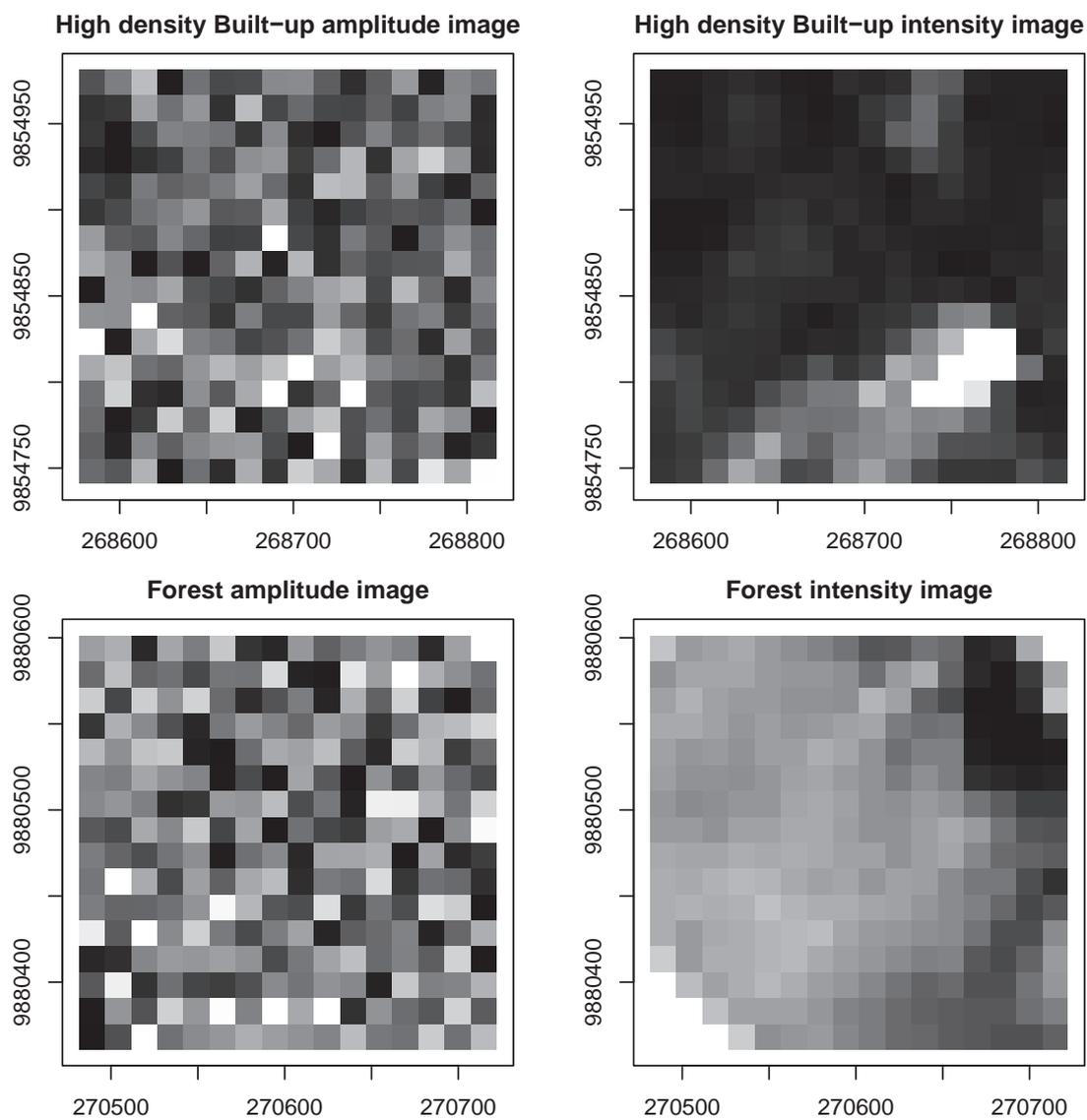


Figure 5.4: Scale I block samples from high density built-up and forest classes using amplitude and intensity images; amplitude blocks gave zero slope.

5.2 DESIGNED CRF RESULTS

5.2.1 Designed association potential

Figures 5.5a and 5.5b show association potential classification results and misclassification error distribution, where white blocks over black background and black blocks over white background represent false positives and false negatives respectively with reference to built-up area detection. Most of the areas misclassified occur at the margins of the two classes. Homogenous areas with white and black colours were correctly classified as built-up and non-built classes: they correspond to the reference.

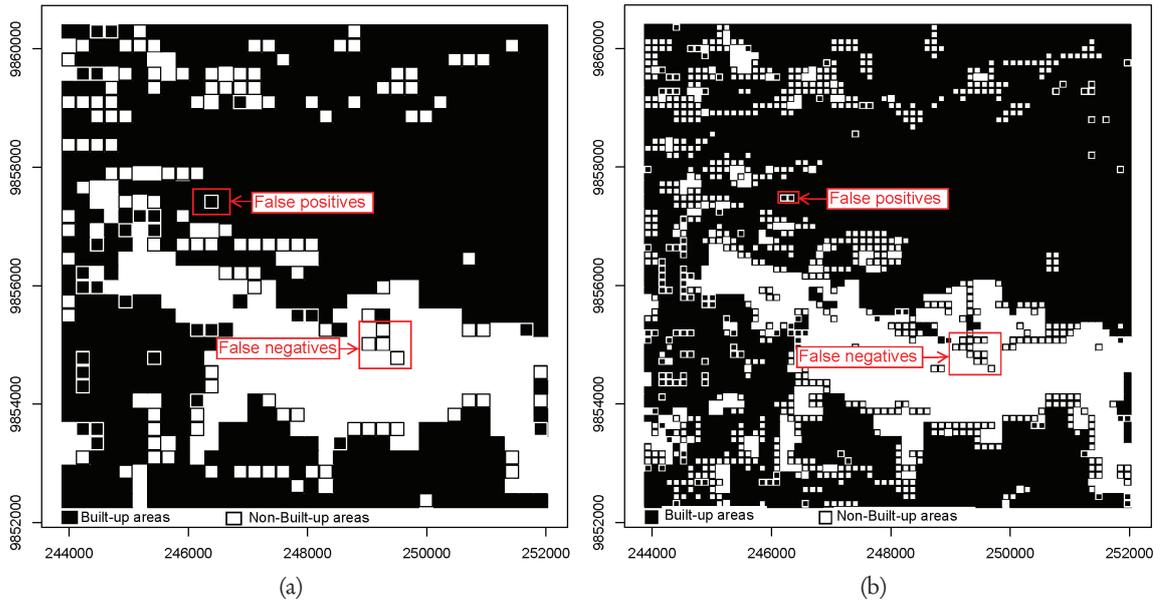


Figure 5.5: Association potential classification results and misclassification error distribution. (a) Scale I. (b) Scale II.

5.2.2 Test of CRF interaction potential data dependent models

Experiment results of four data dependent models and MRF constant smoothing term are presented in Figure 5.6. The Figure shows overall accuracy of each model against β parameter values using a logarithmic scale. At scale II and III inverse of transformed euclidean distance model gives the highest overall accuracy when subjected to a range of low to high values of β . This is followed closely by MRF model. In scale I, overall accuracy of the two models gradually decreases below that of association potential for $\beta > 100$. The overall accuracy of transformed euclidean distance, euclidean distance and absolute difference models, Equations 4.6, 4.9 and 4.10, decline gradually going below that of association potential for $\beta > 1$. The trend is the same in all three scales.

CRF false detection, using the model in Equation 4.5 as a data dependent term, is shown in Figures 5.7a and 5.7b. Classified blocks are displayed over reference image. White blocks over black background depict false positives (areas misclassified as non-built-up) and black blocks over white background shows false negatives (areas misclassified as built-up). CRF minimized the misclassifications errors compared to association potential (Figures 5.5a and 5.5b), however, errors at class boundaries remained.

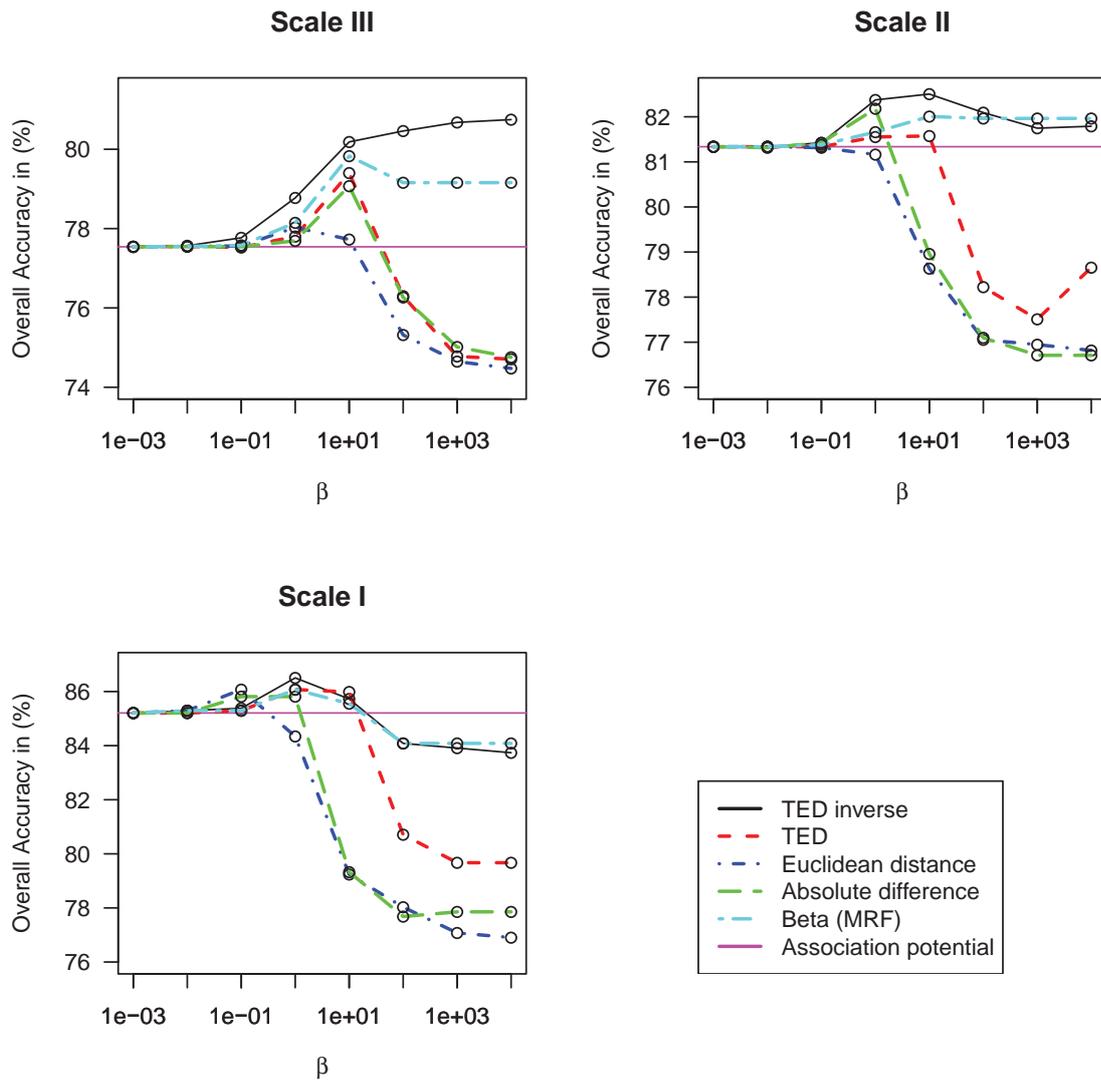


Figure 5.6: Results of interaction potential designs in comparison to MRFs.

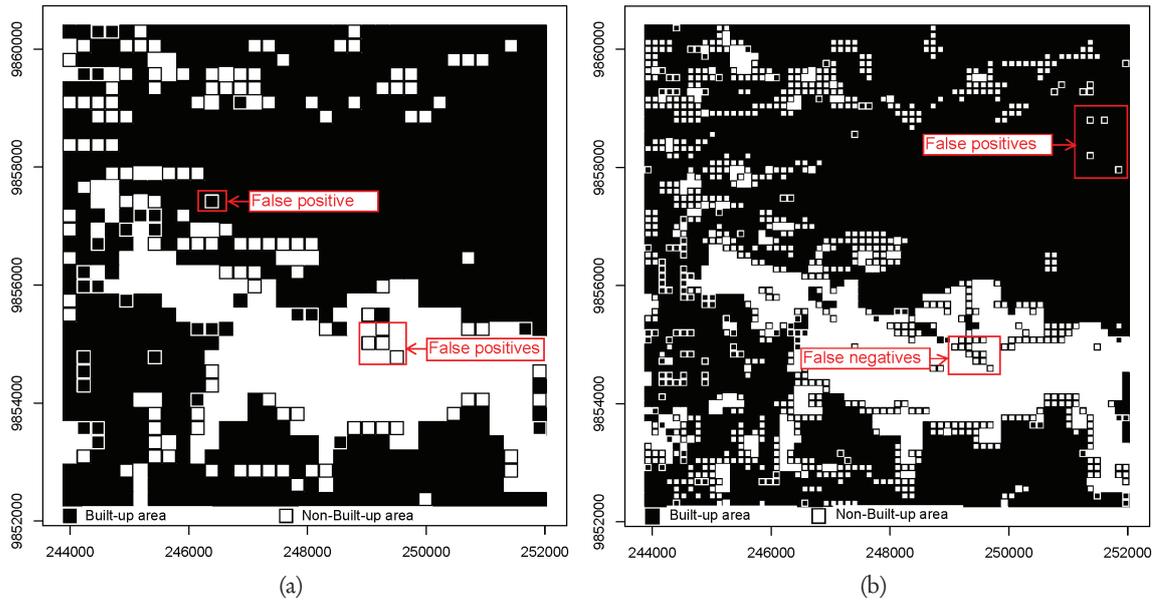


Figure 5.7: CRF classification and misclassification error distribution. (a) Scale I. (b) Scale I.

5.3 APPLICATION OF CRF IN DETECTING BUILT-UP AREAS INCLUDING TEMPORAL CHANGES

Figure 5.10 shows class probabilities in 2011 generated from the association potential. Probability of built-up area class is shown in Figure 5.10a while that of non-built area class in Figure 5.10b respectively. The class probabilities and data features were used in the interaction potential to incorporate spatial dependencies in classification of built-up areas.

Figure 5.8 shows classification results of application of CRF to detect built-up areas in 2006 and 2011 dates respectively. It is evident from 2011 classification that built-up areas have spatially expanded compared to 2006. Major changes have occurred outside the city boundary where there is space for development. The development mainly follows a linear strip pattern as depicted by the change detection results in Figure 5.9. Such development is constrained along major roads.

5.4 ASSESSMENT OF CRF IN DETECTING BUILT-UP AREAS

Validation results of CRF application in detection of built-up areas as compared CRF, MRF, SVM and MLC are presented in Tables 5.4, 5.5, 5.6 and 5.7 respectively. CRF method has the highest overall accuracy and kappa values followed by MRF, SVM and MLC. MLC has the lowest and highest false positives in non-built-up and built-up area classes as compared to the other three methods. CRF maintains the lowest false positives with respect to built-up areas with the least false negatives in both classes. The false positives in non-built-up area class is higher compared to MLC which has the lowest percentage.

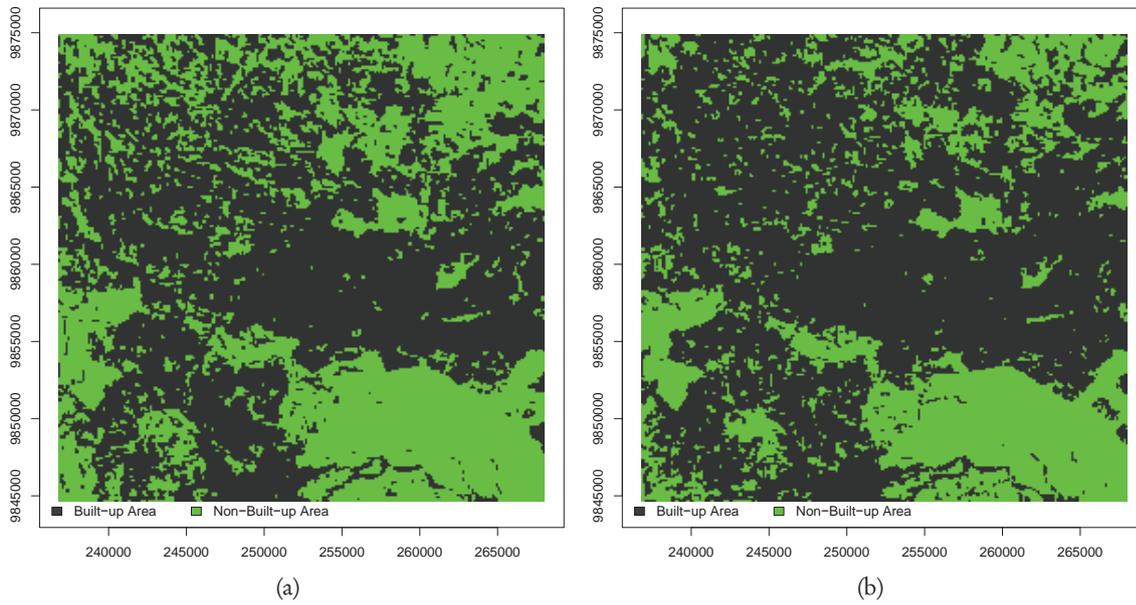


Figure 5.8: CRF classification of built-up areas. (a) 2006 classification. (b) 2011 classification.

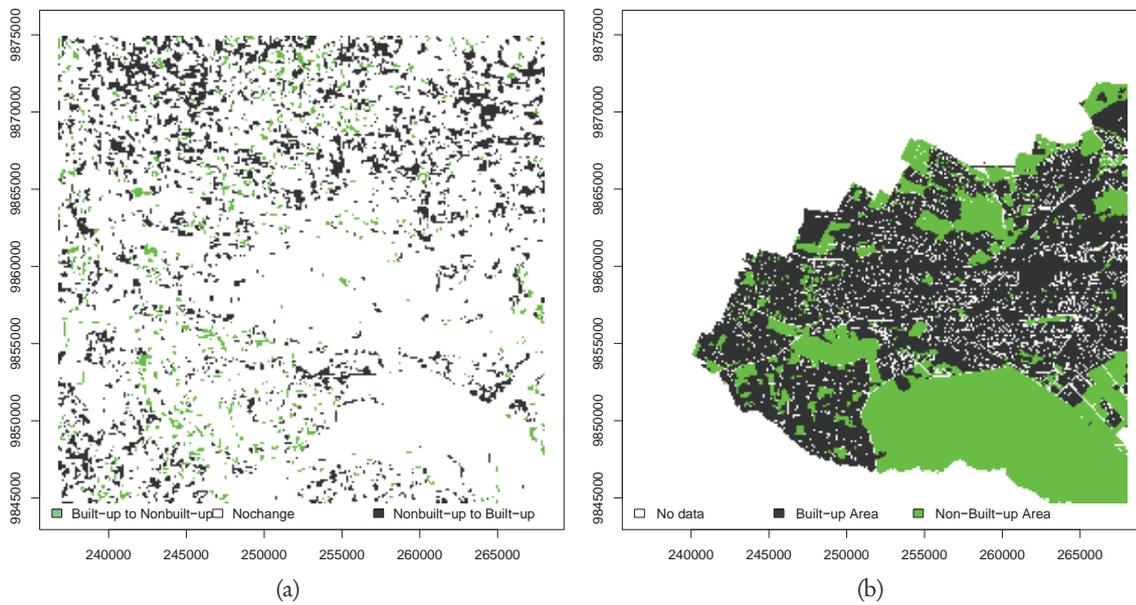


Figure 5.9: Built-up area changes from 2006–2011 and reference data. (a) Land-cover changes. (b) Study area reference data for 2011.

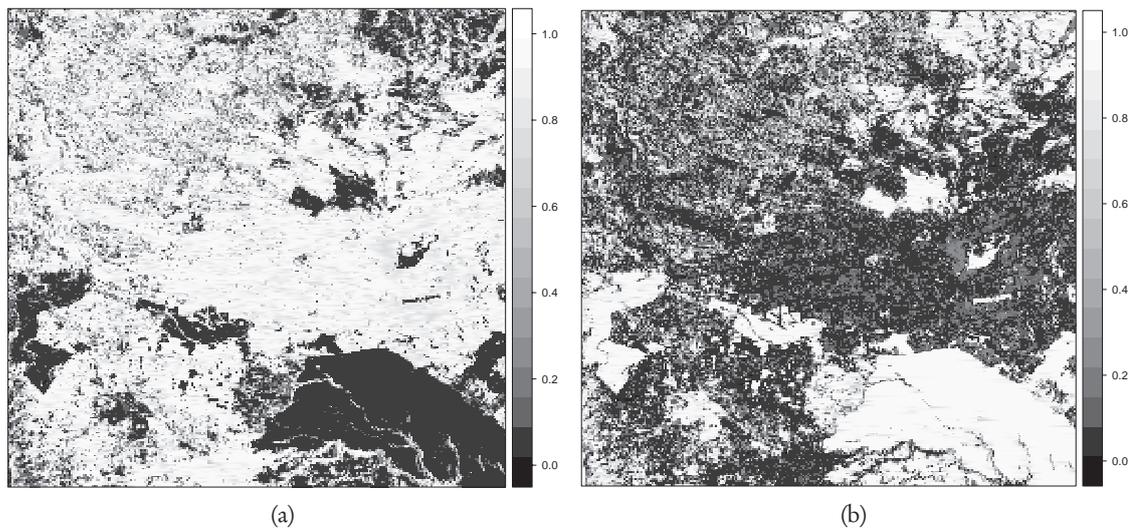


Figure 5.10: 2011 class probability images computed using association potential. (a) Probability of built-up areas. (b) Probability of non-built up areas.

	BU ²	NBU	FN ³
BU	16935	2828	14.3%
NBU ⁴	1274	9754	11.6%
FP ⁵	7.00%	22.5%	
OA	86.7%		
Kappa	0.720		

Table 5.4: CRF accuracy assessment.

	BU	NBU	FN
BU	16486	2939	15.1%
NBU	1723	9643	15.2%
FP	9.46%	23.4%	
OA	84.9%		
Kappa	0.682		

Table 5.6: SVM accuracy assessment.

	BU	NBU	FN
BU	16697	2936	14.9%
NBU	1512	9646	13.5%
FP	8.30%	23.3%	
OA	85.5%		
Kappa	0.696		

Table 5.5: MRF accuracy assessment.

	BU	NBU	FN
BU	15591	2828	15.3%
NBU	4565	11318	28.7%
FP	22.7%	20.0%	
OA	78.4%		
Kappa	0.560		

Table 5.7: MLC accuracy assessment.

²Built-up area

³False negatives

⁴Non-built-up area

⁵False positives

Chapter 6

Discussion

Colour information did not improve classification results when integrated with amplitude and intensity. This occurred because discriminating built-up and non-built-up areas based on colour information is not sufficient. Built-up areas are composed of heterogeneous land-cover materials some of which also occur in non-built-up class. For example, bare-soil and built-up areas with tiled roofs have similar spectral reflectance which is difficult to discriminate by colour information. The outcome corresponds to that of [18] whose work indicates that addition of mean of hue features does not improve classification of settlement areas.

Amplitude features gave high detection quality at scale I, but decreased on reduction of block size. SAR images are accompanied by speckle which introduces sudden variation in pixel values due to noise. The effect of noise is averaged out when features are computed from a larger block size. Variability due to noise increases on reduction of block size. Consequently, class separability reduces, as illustrated by feature spaces, degrading classification accuracy. Gomez-Chova et al. [14] noted the inaccuracies due to speckle and concluded that each pixel be analyzed with respect to its neighbours. How many pixels should be considered as neighbours is still an open question. As demonstrated in the three scales, SAR image block size considerably affected classification accuracy. The study proposes 8×8 window size (scale II) for built-up area related analysis using SAR. This is not a general rule because such choices depend on the minimum size of a class, i.e. building or park, to be detected. In contrast, intensity features improved detection quality when block size was reduced. The number of spectrally different pixels reduced enhancing homogeneity within a block. Thus, separability of intensity features in the feature space increased improving classification quality. Therefore, scale II was selected for classification in a bid to optimize the synergy of amplitude and intensity features.

Slope features were computed by fitting a linear regression model using first three points of the variogram. Experiments conducted proved that this approach is better than model based approach which, automatically fits a model to a variogram. In some blocks of a given class, the chosen variogram models were inaccurate and slope values computed using sill, nugget and range of the fitted model were misleading. This problem is also confirmed by [35].

Amplitude slope features could not enhance discrimination of water, forest and dense built-up area blocks. The variogram of these block have zero slope. This is an indication that scale I and II blocks lack a meaningful spatial structure mainly as a result of random noise due to speckle. Water, forest and dense built-up areas scatter SAR signals differently which contribute to pixels with random brightness variation that introduce noise [see 37, chap. 5]. These land-cover classes could not be delineated using amplitude slope features because they have the same slope value. Therefore, mean and standard deviation features derived from both amplitude and intensity including intensity slope features were selected for classification. Mean, standard deviation and slope incorporated information on homogeneity, variability and spatial dependencies of classes respectively. The feature selection approach is better than subjecting a classifier to multiple features introducing redundancy which, increases computational complexity and degrades classification accuracy [27, 50].

The designed CRF consist of association and interaction potential considered as local site-wise and context classifiers. Association potential was trained such that training error was as low as possible. Despite low training error, misclassifications were still evident. A visual assessment with reference to image in Figure 3.3a showed that bare-soil blocks were misclassified. Thus, association potential alone was not sufficient to detect built areas. Spatial dependency was incorporated with the inclusion of slope features in the data dependent term of the interaction potential. This enhanced detection when the smoothness parameter was properly selected.

Experiments were conducted to evaluate data interaction models as shown in Figure 5.6. Studies like [17, 19] and [27, 49, 50] used a classical approach of modelling the data dependent term by adopting absolute difference and concatenation of features respectively. The study introduced a novel data interaction term using the inverse of the transformed euclidean distance of site-wise features. This design maintains stable overall accuracy when subjected to different smoothness parameters unlike transformed euclidean distance, euclidean distance and absolute difference models whose accuracies decline when high values are used. The stable trend in accuracy is an attribute of the model's scaled values between 0 and 1. The model controls smoothness parameter imposed based on similarity of blocks. Similar blocks are subjected to high smoothing and regulated in areas with dissimilar blocks. This allows the interaction potential to partition space of the relational features in Equation 4.7 into pairs hypothesized to have similar labels and dissimilar labels as concluded by [23]. In contrast, transformed euclidean distance, euclidean distance and absolute difference models degrade the accuracy when subjected to high β values. These models penalize dissimilar blocks thereby destroying class boundaries and so, produce over-smoothed classified images. The same is exhibited by MRF models. Therefore, the inverse of the transformed euclidean distance was considered a better model and applied in detection of built-up areas.

The designed CRF detection accuracy exceeds that of MRF, SVM and MLC by 1.13%, 2.22% and 8.23% respectively. It has the lowest false negatives in both classes and false positives with respect to built-up areas. The percentage difference in accuracy are substantial given that reference data used only covered main city core and other areas were exempted during validation as illustrated in Figure 5.9b. Built-up areas are detected better than non-built-up areas as demonstrated by false positive percentages. This is because CRF is a discriminative model and requires training examples of all possible combinations especially in non-built-up areas, where land-cover structure is complex [44, 49, 50]. Inner areas of built-up class have stable and homogenous spatial dependencies hence CRF performs well. The challenge prompted [50] to integrate MRF and CRF; the former can handle all class compositionality present in the image [44]. CRF also slightly decreased misclassification errors because of the optimization method (ICM) used. ICM is a local optimization method and results from it depend very much on the performance of initial estimator (association potential) [26]. Therefore, CRF slightly enhanced classification, specifically along the margins of built-up areas and bare-soil, because association potential failed in those areas. MLC and SVM have high false negative percentage compared to MRF and CRF because they do not integrate spatial dependencies during classification.

Developed CRF attained an overall accuracy of 86.7% with over 77.5% producer and user accuracies per class. There are no specific standardized accuracy thresholds for land-cover classification. Nevertheless, [3, 39] agree on a target of an overall accuracy of 85% with [39] setting a threshold of not less than 75% accuracy per class. The designed method exceeds the established accuracy targets. Moreover, whether the accuracy values are acceptable or not lies solely on user requirement. The study recommends applying the method for built-up area detection. During the five year period built-up areas increased by approximately 98.9 hectares while 26.7 hectares were converted to non-built-up areas. Such information is useful in urban growth studies, population density estimation and general urban planning needs.

In retrospect, it cannot go without mention that assumptions were made in the study for practical implementation. First, a binary classification was used. Classes with diverse spectral values were merged into either built-up or non-built super-classes. This reduced class separability which degraded classification quality of the association potential. Future work should extend the method to handle multi-class classification. Second, the features used in the data dependent term were computed using pixels in a fixed block size. Variability of pixel values was averaged to one value. Computational efficiency and accuracy increased. However, modeling of spatial dependencies was constrained to neighbouring blocks within a first order neighbourhood system. The trade off can be addressed by including features from different block sizes using a multi-scale approach. Third, the designed data dependent term controlled smoothing using values directly computed by the inverse of the transformed euclidean distance. In reality, spectral values of a given land-cover class have intra-class variability. Thresholds can be introduced in the data dependent term to accommodate intra-class and inter-class variation.

Chapter 7

Conclusion and recommendations

The aim of the research was to develop a method for detecting built-up areas. Blocks of pixels as opposed to individual pixels were used during classification. This reduced noise due to spectral variability and speckle, increased computational efficiency and classification accuracy. The approach introduced a human concept into classification, where classes in an image are identified using observation windows as opposed to individual pixels which enhanced detection of built-up areas. A feature selection approach was used rather than entrusting the classifier the task of optimizing all set of possible features. Slope features incorporated spatial dependencies of classes which improved classification. The use of variogram's slope for classification is promising. Overall, integration of mean, standard deviation and slope features from SAR and ASTER images enhanced detection of built-up areas.

Based on several experiments conducted, the inverse of the transformed euclidean distance emerged the best model of integrating spatial dependencies. The model presents two advantages:

1. it maintains stable accuracy when subjected to a range of small to large constant smoothness parameter (β)
2. it regulates preserves edges/class boundaries and allows aggregation of similar labels during classification thus playing the role of a discontinuity adaptive model.

The notion of the inverse of the transformed euclidean distance was to support the hypothesis of retaining "similar" and "dissimilar" labels given the evidence from the data. This was realized with implementation provided.

The designed CRF model is promising for detection of built-up areas and their temporal changes. Scale II enhanced discrimination of built-up areas compared to scale I and III. It is evident that to overcome spectral variability and speckle noise, spatial dependencies in the image and labels should be included in classification. The developed CRF method addresses this challenge compared to MLC, SVM and MRF. MLC and SVM completely ignore spatial dependencies while MRF considers spatial dependencies in labels only ignoring existing spatial dependencies in data.

The use of variogram for training set characterization is promising. Further work should consider investigating how it can be exploited to efficiently characterize features from SAR. In addition, integration of multi-scale features and a multi-class approach is expected to improve classification results.

Bibliography

- [1] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini. Landsat ETM+ and SAR image fusion based on generalized intensity modulation. *IEEE Transactions on Geoscience and Remote Sensing*, 42(12):2832–2839, 2004.
- [2] D. Amarsaikhan, H. H. Blotevogel, J. L. van Genderen, M. Ganzorig, R. Gantuya, and B. Nergui. Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification. *International Journal of Image and Data Fusion*, 1(1):83–97, 2010.
- [3] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer. A land use and land cover classification system for use with remote sensor data. *Development*, 2001(964):28, 1976.
- [4] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.
- [5] B. Bhatta. *Analysis of urban growth and sprawl from remote sensing data*. Advances in Geographic Information Science. Springer-Verlag, Berlin, Heidelberg, 2010.
- [6] C. Corbane, N. Baghdadi, X. Descombes, G. Wilson, N. Villeneuve, and M. Petit. Comparative study on the performance of multiparameter SAR data for operational urban areas extraction using textural features. *IEEE Geoscience and Remote Sensing Letters*, 6(4):728–732, 2009.
- [7] C. Corbane, J. F. Faure, N. Baghdadi, N. Villeneuve, and M. Petit. Rapid urban mapping using SAR/optical imagery synergy. *Sensors*, 8(11):7125–7143, 2008.
- [8] CSUD. nairobiGISmaps. <http://nairobiGISmaps.wikischolars.columbia.edu>, 2010. [Online: accessed 1-November-2011].
- [9] De Jong, S.M. and van der Meer, F.D. *Remote sensing image analysis : including the spatial domain*, volume 5 of *Remote sensing and digital image processing*. Kluwer Academic, Dordrecht etc., 2004.
- [10] F. Dell’Acqua and P. Gamba. Texture-based characterization of urban environments on satellite SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 41(1):153–159, 2003.
- [11] T. Esch, A. Roth, and S. Dech. Robust approach towards an automated detection of built-up areas from high resolution radar imagery. In *3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas*, Tempe, Arizona, USA, 2005.
- [12] T Esch, Michael Thiel, Andreas Schenk, A Roth, A Muller, and Stefan Dech. Delineation of urban footprints from TerraSAR-X data by analyzing speckle characteristics and intensity information. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):905–916, 2010.

- [13] G. Galster, R. Hanson, M. R. Ratcliffe, H. Wolman, S. Coleman, and J. Freihage. Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing Policy Debate*, 12(4):681–717, 2001.
- [14] L. Gomez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila, and G. Camps-Valls. Urban monitoring using multi-temporal SAR and multi-spectral data. *Pattern Recognition Letters*, 27(4):234–243, 2006.
- [15] M. F. Goodchild. Geographical information science. *International Journal of Geographical Information Systems*, 6(1):31–45, 1992.
- [16] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.
- [17] W. He, M. Jaeger, A. Reigber, and O. Hellwich. Building extraction from polarimetric SAR data using mean shift and conditional random fields. In *7th European Conference on Synthetic Aperture Radar (EUSAR)*, 2008.
- [18] T. Hoberg and F. Rottensteiner. Classification of settlement areas in remote sensing imagery using conditional random fields. In *ISPRS TC VII Symposium-100 Years ISPRS*, 2010.
- [19] T. Hoberg, F. Rottensteiner, and C. Heipke. Classification of multitemporal remote sensing data using conditional random fields. In *Workshop on Pattern Recognition in Remote Sensing (PRRS)*, pages 1–4, Istanbul, 2010.
- [20] M.E. Hodgson. What size window for image classification? A cognitive perspective. *Photogrammetric Engineering and Remote Sensing*, 64(8):797–807, 1998.
- [21] ITC. *GI Science and Earth Observation : a process - based approach*. ITC Educational Textbook Series. University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede, 2010.
- [22] S. Kumar. *Models for learning spatial interactions in natural images for context-based classification*. PhD thesis, Carnegie Mellon University Institute of Robotics, 2005.
- [23] S. Kumar. Discriminative random fields. *International Journal of Computer Vision*, 68(2):179–201, 2006.
- [24] S. Kumar and M. Hebert. Discriminative fields for modeling spatial dependencies in natural images. In *Advances in Neural Information Processing Systems*, 2004.
- [25] J. D. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001.
- [26] S. Z. Li. *Markov random field modeling in image analysis*. Advances in pattern recognition. Springer, 2009.
- [27] Z. Li, J. Ma, R. Zhang, and X. Chen. Urban areas detection based on conditional random field and multiscale textural features. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2009.
- [28] A. Lorette, X. Descombes, and J. Zerubia. Texture analysis through markov random fields: urban areas extraction. In *International Conference on Image Processing*, 1999.

- [29] D. Lu, S. Hetrick, E. Moran, and G. Li. Detection of urban expansion in an urban-rural landscape with multitemporal quickbird images. *Journal of Applied Remote Sensing*, 4(1):041880–041897, 2010.
- [30] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, 2007.
- [31] D. S. Lu, S. Hetrick, and E. Moran. Land cover classification in a complex urban-rural landscape with QuickBird imagery. *Photogrammetric Engineering and Remote Sensing*, 76(10):1159–1168, 2010.
- [32] M Omwenga. Urban growth and sprawl: case study of Nairobi, Kenya. In *World Urban Forum 4*, Nanjing, China, 2008.
- [33] D. Parikh and D. Batra. CRFs for image classification. Technical report, Carnegie Mellon University, 2008.
- [34] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.
- [35] Atkinson P.M. and Lewis P. Geostatistical classification for remote sensing: an introduction. *Computers & Geosciences*, 26(4):361–371, 2000.
- [36] T. Rashed and C. Jürgens. *Remote sensing of urban and suburban areas*, volume 10. Springer, 2010.
- [37] J. A. Richards. *Remote sensing with imaging radar*. Springer series on signals and communication technology. Springer, Heidelberg, 2009.
- [38] C. Sutton and A. McCallum. An introduction to conditional random fields for relational learning. In *Introduction to statistical relational learning*. The MIT Press, 2006.
- [39] J. R. Thomlinson, P. V. Bolstad, and W. B. Cohen. Coordinating methodologies for scaling landcover classifications from site-specific to global: steps toward validating global map products. *Remote Sensing of Environment*, 70(1):16–28, 1999.
- [40] H. Thunig, N. Wolf, S. Naumann, A. Siegmund, C. Jürgens, C. Uysal, and D. Maktav. Land use/ land cover classification for applied urban planning - the challenge of automation. In *Joint Urban Remote Sensing Event (JURSE)*, 2011.
- [41] V.A. Tolpekin 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*, 47(9):3283–3297, 2009.
- [42] B. Tso and P. M. Mather. *Classification methods for remotely sensed data*. CRC Press, Boca Raton, 2nd edition, 2009.
- [43] F. Tupin. Fusion of optical and SAR images. In *Radar Remote Sensing of Urban Areas*. Springer Netherlands, 2010.
- [44] I. Ulusoy and C. M. Bishop. Generative versus discriminative methods for object recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 258–265, 2005.

- [45] R. Webster and M.A. Oliver. *Geostatistics for environmental scientists*. Wiley & Sons, 2nd edition edition, 2008.
- [46] J. D. Wegner, R. Hansch, A. Thiele, and U. Soergel. Building detection from one orthophoto and high-resolution InSAR data using conditional random fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1):83–91, 2011.
- [47] J. D. Wegner, U. Soergel, and B. Rosenhahn. Segment-based building detection with conditional random fields. In *Joint Urban Remote Sensing Event (JURSE)*, Munich, 2011.
- [48] X. Yang. *Urban remote sensing: monitoring, synthesis and modeling in the urban environment*. Wiley-Blackwell, Chichester, 2011.
- [49] P. Zhong and R. Wang. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3978–3988, 2007.
- [50] P. Zhong and R Wang. Using combination of statistical models and multilevel structural information for detecting urban areas from a single gray-level image. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1469 –1482, 2007.

Appendix A

Appendix: Part one

A.1 DESIGNED CRF POTENTIALS: R PROGRAM CODE

```
#####  
# VARIABLE DEFINITION  
#####  
S <- 8 #Block size  
beta0 <- 1 #CRF penalty Parameter  
Nneighb <- 4 #Neighbourhood system definition  
# Image dimensions  
d <- ASTER@grid@cells.dim #original Image  
psize <- ASTER@grid@cellsize #pixel size  
M <- d[1] #Number of rows in the original image  
N <- d[2] #Number of columns in the original image  
Mdeg <- floor(M/S) #Number of rows in the block image  
Ndeg <- floor(N/S) #Number of columns in the block image  
#####  
# DEFINITION OF CRF POTENTIALS  
#####  
#Block Neighbourhood definition  
#  
Nns <- array(0, Mdeg, Ndeg)  
Bns <- array(0, c(Mdeg, Ndeg), 4)  
for(bn in 1:(Mdeg * Ndeg))  
{  
  j <- floor((bn-1)/Mdeg)+1  
  i <- bn-(j-1) * Mdeg  
  Nns[bn] <- 0  
  k <- 1  
  if(i > 1)  
  {  
    Bns[bn, k] <- bn-1  
    k <- k+1  
    Nns[bn] <- Nns[bn]+1  
  }  
  if(i < Mdeg)  
  {  
    Bns[bn, k] <- bn+1  
    k <- k+1  
    Nns[bn] <- Nns[bn]+1  
  }  
}
```

```
}
if (j > 1)
{
    Bns[bn, k] <- bn - Mdeg
    k <- k + 1
    Nns[bn] <- Nns[bn] + 1
}
if (j < Ndeg)
{
    Bns[bn, k] <- bn + Mdeg
    k <- k + 1
    Nns[bn] <- Nns[bn] + 1
}
}
blocknr <- function (i, j)
{
    return (i + (j - 1) * Mdeg)
}
# Definition of features used in Interaction potential
#-----
Nf <- 5
norms <- array (0, c(Nf, 2))
D <- array (0, c(Mdeg, Ndeg, Nf))
D[, 1] <- as.matrix (Iblock@data[, 1]) #ASTER mean
D[, 2] <- as.matrix (Iblock@data[, 2]) #ASTER sd
D[, 3] <- as.matrix (Iblock@data[, 3]) #ASTER Slope
D[, 4] <- as.matrix (Iblock@data[, 4]) #SAR mean
D[, 5] <- as.matrix (Iblock@data[, 5]) #SAR mean
# Standardizing block features by mean and sd
#-----
norms <- array (0, c(Nf, 2))
for (k in 1:Nf)
{
    norms[k, 1] <- mean (D[, k])
    norms[k, 2] <- sd (D[, k])
    D[, k] <- (D[, k] - norms[k, 1]) / norms[k, 2]
}
F <- array (0, Mdeg, Ndeg)
F <- as.vector (SVM@data$class)
Uc <- array (0, c(Mdeg, Ndeg, 2))
Uc[, 1] <- SVM@data$Ucond1
Uc[, 2] <- SVM@data$Ucond2

Initial <- F # Initial class probabilities from SVM
I <- function (x, y) {
    val <- 1
    if (x == y) val <- 0
    return (val)
}
```

```
}
#Data dependent smoothing term
#-----
Bf <- function(v1,v2){
  E <- (v1-v2)
  val <- 1/(1-exp(-sum(E^2))) #Transformed Euclidean distance
  #Threshold to exempt zero in the denominator
  if(val==0)
  {
    val=0.001
  }
  return(val)
}
#Interaction potential
#-----
Uinter <- function(bn){
  val <- 0
  F0 <- F[bn]
  D0 <- D[bn,]
  for(k in 1:Nns[bn])
  {
    bn1<-Bns[bn,k]
    F1 <- F[Bns[bn,k]]
    beta <- beta0 Bf(D0,D[bn1,])
    val <- val + beta I(F0,F1)
  }
  return(0.5 val/Nneighb)
}
#Association potential
#-----
Uasso <- function(bn)
{
  val <- Uc[bn,F[bn]]
  return(val)
}
#CRF
#-----
U <- function(bn)
{
  val <- Uinter(bn) + Uasso(bn)
  return(val)
}
#Total energy of a block
#-----
TotalEnergy<-function(F)
{
  val <- 0
  for(bn in 1:(Mdeg Ndeg))
```

```

    {
      val <- val + U(bn)
    }
    return(val/((Mdeg) (Ndeg)))
  }

```

```

#=====
#END OF DEFINITION OF CRF POTENTIALS
#=====

```

A.2 OPTIMIZATION OF CRF USING ICM: R PROGRAM CODE

```

#=====
#OPTIMIZATION USING ICM
#=====
Niter <- 10# Maximum allowed number of iterations
Etotal <- array(rep(0, Niter), 1)#Total energy array
windows()
par(mfrow=c(2,2))
Fdisp <- SVM
Fdisp@data <-data.frame(F)
for(iter in 1:Niter)
{
  upd_count <- 0
  for(bn in 1:(Mdeg Ndeg))
  {
    F_update <- 3-F[bn]
    Ft <- F[bn]
    u1 <- U(bn)
    F[bn] <- F_update
    u2 <- U(bn)
    u1 <- u2-u1
    if(u1>0)
    {
      F[bn] <- Ft
    }
    else upd_count<-upd_count+1
  }
  Etotal[iter] <- TotalEnergy(F)
  if(upd_count==0) break
}
Fdisp@data <-data.frame(F) #Saving optimized image/classes
names(Fdisp)<-"class"
image(Fdisp,col=c("sienna4","seagreen1"),xlab="",ylab="")
title("CRF") #Displays CRF classification
if(WithRef)
{
  image(Ref,col=c("sienna4","seagreen1"),xlab="",ylab="")
  title("Reference") #Displays reference image
}

```

```
plot(1:iter, Etotal[1:iter], type = 'l', main = 'Energy_
  minimisation', xlab = 'Iteration', ylab = 'Etotal')
iter          # Number of iterations
Etotal[iter] # The obtained energy value:
#=====
#END OF OPTIMIZATION USING ICM
#=====
```

A.3 CHANGE DETECTION: R PROGRAM CODE

```
#=====
#CHANGE DETECTION BY IMAGE DIFFERENCING
#=====
rm(list=ls(all=TRUE))
require(rgdal)
S          <- 8 # Pixel block size
Path      <- 'D:\\Benson_CRF\\ChangeDetection'

#Reading 2006 classified image
#-----
setwd(Path)
image.fn<- paste('2006_CRF',S,'BY',S,','','.TIF',sep='')
# --- GDAL Definition
fn.region<-image.fn
work.region.tif <- new("GDALReadOnlyDataset", fn.region)
CRF2006 <- work.region.tif[,1]
# --- close image
GDAL.close(work.region.tif)
names(CRF2006)<- "class"
#Displaying 2006 classified Image
windows()
par(mar = c(2.2, 2.2, 2.2, 2.2))
image(CRF2006, col=c("sienna4", "seagreen1"), xlab="", ylab="", axes
  =TRUE)
classes.Palette <- colorRampPalette(c("sienna4", "seagreen1"))
legend("bottomleft", c('Built-up_Area', 'Non-Built-up_Area'), fill=
  classes.Palette(2), bty='n', horiz=TRUE)
title('2006_CRF_Classification')

#Reading 2011 classified image
#-----
image.fn<- paste('2011_CRF',S,'BY',S,','','.TIF',sep='')
# --- GDAL Definition
fn.region<-image.fn
work.region.tif <- new("GDALReadOnlyDataset", fn.region)
CRF2011 <- work.region.tif[,1]
# --- close image
GDAL.close(work.region.tif)
names(CRF2011)<- "class"
```

```
#Displaying 2011 classified Image
windows ()
par(mar = c(2.2, 2.2, 2.2, 2.2))
image(CRF2011, col=c("sienna4", "seagreen1"), xlab="", ylab="", axes
      =TRUE)
classes.Palette <- colorRampPalette(c("sienna4", "seagreen1"))
legend("bottomleft", c('Built-up_Area', 'Non-Built-up_Area'), fill=
      classes.Palette(2), bty='n', horiz=TRUE)
title('2011_CRF_Classification')
```

```
#Change detection by Image differencing
#
CRF2011$change <- as.matrix(CRF2006$class)-as.matrix(CRF2011$
  class)
Disp <- CRF2011
Disp@data <- data.frame(CRF2011$change)
windows ()
par(mar = c(2.2, 2.2, 2.2, 2.2))
image(Disp, col=c("seagreen1", 'white', "sienna4"), xlab="", ylab=""
      , axes=TRUE)
classes.Palette <- colorRampPalette(c("seagreen1", 'white', "
  sienna4"))
legend("bottomleft", c('Built-up_to_Nonbuilt-up', 'Unchanged', '
  Nonbuilt-up_to_Built-up'), fill=classes.Palette(3), horiz=TRUE,
      bty='n', x.intersp=0.8)
title('Built-up_area_changes_2006-2011')
```

```
#Computing changes in hectares
#
#Built-up area increase in ha:
Built_up <- (sum(CRF2011$change==1) S 15)/10000
Built_up
#Non-Built-up area increase ha:
Non_Built <- (sum(CRF2011$change==-1) S 15)/10000
Non_Built
#
#=====#
#END OF CHANGE DETECTION BY IMAGE DIFFERENCING
#=====#
```