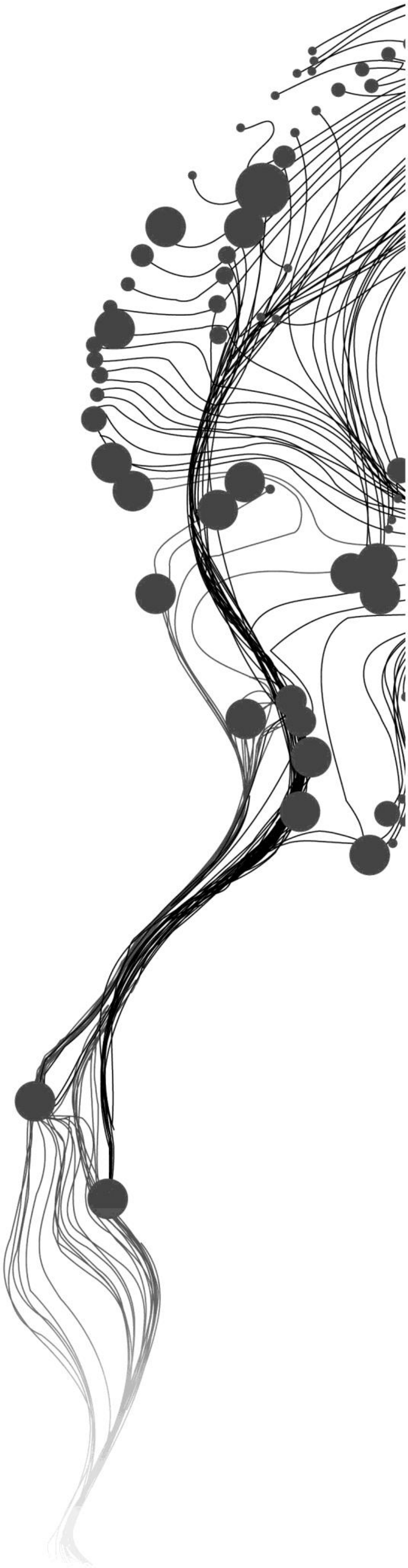
USE OF PRIOR INFORMATION ON OBJECTS IN CONTEXTUAL CLASSIFICATION WITH MARKOV RANDOM FIELDS

OBIKO VICTOR ONYIMBO CALEB
March 2012
1

SUPERVISORS:
Dr. V.A. Tolpekin
Dr. Ir. W. Bijker



USE OF PRIOR INFORMATION ON OBJECTS IN CONTEXTUAL CLASSIFICATION WITH MARKOV RANDOM FIELDS

OBIKO VICTOR ONYIMBO CALEB
Enschede, The Netherlands, February 2012
1

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
Dr. Ir. W. Bijker

THESIS ASSESSMENT BOARD:
Prof. Dr. A. Stein (Chair)
Dr. Ir L. Kooistra (External Examiner, Wageningen UR)

ABSTRACT

Classification of VHR remote sensing satellite imagery is an area that requires extensive research in the improvement of classification on objects. Classification results on objects are less accurate with class separability that is poor or excellent. This study, therefore, is focused in the introduction of prior information on objects in contextual classification with Markov Random Fields (MRF) aiming at improving classification. The prior information incorporated in the classification (in this research topic) is shape and size, tree crowns being the subject. Tree crowns have a round geometric shape with definite area which is modeled in the prior term of the MRF objective function.

A tuning subset of the multispectral image that has a tree crown with grass, bare soil and shadow as the background is used in the estimation of MRF parameters. This study explores optimisation of the MRF parameters: Lambda, lambda segments, temperature, temperature update and lambda shape. Simulated annealing energy minimization algorithm is used to establish the optimal values of temperature and temperature update. Kappa values, producer, user and overall accuracy determine the optimal parameters for lambda and lambda segments. The optimal MRF parameters obtained are applied on the tuning subset; the results produced are more accurate and reproducible.

The methodology of this study is a Hierarchical Markov Random Field (HMRF) approach. In level one, the MRF pixel based level, the energy of each class label is minimized, increasing the probability of that pixel being assigned a particular class by penalizing adjacent pixels. This MRF method is iterative and converges when the energy is minimized to zero. In level two, the MRF object based level, shape and area as spatial context are modeled in the prior term of the MRF energy function by the concept of smoothness prior.

Accuracy assessment is done by use of a reference panchromatic image subset. The HMRF results of the tuning subset are analysed in the confusion matrix to determine the tree crown pixels that have been correctly classified and those misclassified (Errors of omission and commission). Implementation is done on two different subsets: 1) A well separated three tree crown area, 2) Tree crowns close to each other. The results obtained show that the HMRF method improves classification on tree crowns that are separated and the classification accuracy is low for interlocked ones.
In conclusion, the HMRF method developed by integrating shape and area outperforms the MLH in the classification of separated tree crowns.

Key words: - Hierarchical Markov Random Fields (HMRF), Simulated Annealing (SA), Iterative Conditional Modes (ICM), Maximum Likelihood Classification (MLC), Tree crown.

ACKNOWLEDGEMENTS

First and foremost, I wish to thank and appreciate all those who contributed in my studies at ITC, particularly the Netherlands Fellowship Programme (NFP) in the provision of scholarship.

Secondly, I extend my sincere thanks and gratitude to my supervisors, Dr. Valentyn Tolpekin and Dr. Wietske Bijker, who guided me in this study. I appreciate their constructive criticism, comments, suggestions, advice and a fore sight. Thank you.

Thirdly, I give my deepest thanks to my mum Ruth Kemunto and dad Caleb Obiko, who time and again encouraged me to study, aim higher and never give up in life.

My dearest bow to my lovely wife Irene Mwende, my one and only daughter Medyline Mueni Rosa, my son Alfaxard Onchieku, and the last born son Yuvinalis Nyangotto who was born when at ITC. I appreciate you as my family. God shower you with blessings.

I would like to thank the Government of Kenya, particularly the ministry of lands for offering me a study leave to seek further education overseas. Vote of thanks.

Last but not least, I wrap all the thanks and appreciation in the name of the Almighty God.

Ahsanteni.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1. Motivation and problem statement

Remote sensing is the data and information acquisition of a scene on the ground surface by instrument based techniques without physical contact with that scene. This is done by extracting the data and information from the scene. The scene does have typical land cover classes: Grass, bare soil, buildings, trees, roads. These should be classified by a classification process; assigning pixels the respective class labels where one class label can be assigned per pixel or multiple classes might be assigned a single pixel.

In high resolution remotely sensed imagery, land cover classification is a problem in the detection of tree crowns. This is due to the spectral confusion between tree crowns and other land cover classes: Grass, bare soil, shrubs, flowers. Traditional classifiers perform classification of an image scene either in pixel-based or object-based technique. Features such as roads, buildings, tree crowns, have a geometric extent that can be spatially modelled to assist improve classification accuracy. In this context, we wish to establish a technique that can be applied in classification of tree crowns to improve the accuracy of classification.

Since high resolution satellite imagery results in classification inaccuracies when classifying tree crowns, it is challenging to clearly establish the boundary of tree crowns in a remotely sensed scene. Many factors hence determine classification output, for instance, the class separability, spatial and spectral resolution of the satellite imagery, the classification algorithm in use and texture analysis of the scene.

In a high resolution multispectral image, pixels vary in spectral information. When there is a large within-class variation, this causes an overlap with other classes to the extent that there is confusion spectrally between the classes. In addition, adjacent pixels with the same spectral information are usually grouped to form segments in a process termed segmentation. These segments in the image are a representation of objects on the ground. Objects in the image, tree crowns for instance, portray a transitional fuzziness at their boundary pixels making it not clear to establish the tree crown extent, for these pixels are mixed. These mixed pixels are one of the causes of classification inaccuracy.

Contextual classification with Markov Random Fields (MRF) was researched in the paper (Tolpekin and Stein 2009). They explored and tuned the smoothness parameter values, class separability and the scale factor between the prior and likelihood terms in the posterior energy function. The accuracy of the resulting land-cover-map is assessed by means of the kappa statistic at the fine-resolution scale. The study shows that SRM is now applicable to a larger set of images with class separability ranging from poor to excellent.

Arachchige (2011) researched to try and solve classification inaccuracies in a remotely sensed scene albeit partially by focusing on the integration of information from VHR multispectral and panchromatic imagery, to be able to determine the building footprints. In the first place, he did not take into account the spatial context of the objects- buildings -in his research. The basic cognitive characteristics of buildings such as shape, size, pattern, proximity to related features, were not taken into concern in the study. However, his classification technique gives a smooth result that is easy to form segments but not possible to impose classification constraints in a rule based approach on the segment characteristics.

This research topic is based on the assumption that classification in high resolution satellite imagery is inaccurate. The research targets a contextual classification technique that incorporates prior information on tree crowns with MRF to result in improved classified tree crowns. Tree crowns are considered a good case study for they have spatial context that can be incorporated in contextual classification to improve classification. Tree crowns are also round in geometry and their size can be computed.

1.2. Research identification

According to the above motivation and problem statement, mixed pixels at tree crown boundary in high resolution images is an area that require extensive research. Hence, we identify this as a potential research topic.The overall objective of this research is to improve a contextual classification algorithm by taking into account prior information on tree crowns.

1.2.1. Research objectives

The **main objective** of this study is to extend a classification technique by including prior knowledge on tree crowns with MRF, to classify and detect tree crowns from a high resolution multispectral image of Naivasha-Kenya. The extended resultant technique shall be a Hierarchical Markov Random Fields (HMRF) contextual classification.

1.2.2. Research questions

The **research questions** that build on the objectives are:

1. Is shape and size of tree crown an important spatial characteristic in classification?
2. What quantifiable accuracy measures are established by the improved algorithm in comparison with maximum likelihood classification (MLC)?
3. What is the efficiency of this classification algorithm in relation to various determining factors: fitness for purpose, computational time and contextual classification accuracy.

1.2.3. Innovation aimed at

Inclusion of shape and size with MRF in the classification process.

1.3. Thesis structure

This thesis document is structured in six chapters. Chapter one is the introduction, motivation and problem statement. Chapter two is the literature review which contains related work, the MRF theory, supervised classification, maximum likelihood classifier, class separability, iterated conditional modes, circularity in Imagine Objective, Shape Index in E-cognition, Single Feature Probability (SFP) and shape classification. The third chapter details the study area and data preparation for this research. Chapter four focuses on research methods whereas chapter five describes results obtained. Chapter six elaborates results and their discussion. The last and final chapter draws conclusions and recommendations for further research.

2. Literature review

2.1. Related work

A number of research studies have been carried out in the classification of objects and recommended several approaches to tackle the problem of accuracy in image classification.

Pixel-based and object-based classification in high resolution satellite imagery were carried out in the paper (Moskal et al., 2011). They did compare the two methods and established that the object-based classification produces better results than the pixel-based approach. This study was carried out for trees in an urban area. They did find that spectral information and object characteristics; texture, shape, are indeed important in improving classification accuracy.

The paper by Kamal and Phinn (2011) also elaborate that object-based classification methods present higher overall accuracy than Pixel-based techniques. The study was carried out on mangrove species in a high spectral resolution satellite image. Derin and Elliott (1987) present an approach using random field models for noisy and textured images. This was based on a hierarchical Gibbs distribution where statistical information that includes size, shape, orientation, and frequency were considered. Maximum a posterior (MAP) estimation and a hierarchical segmentation scheme were applied. This method has a shortfall for it neither considers parameter estimation techniques in the segmentation algorithms nor the MRF.

A MRF based method that uses contextual information and multiscale fuzzy line process in classification is studied in the paper (Tso and Olsen 2005). A natural reserve park was taken as the study area where multiscale line features are merged through a fuzzy fusion process and integrated into the MRF model. Classification was performed on a multispectral image which produced results that have reduced bias contributed by boundary pixels, and the over smoothness classification patterns are also restricted. In the paper by Khedama (2004), maximum a posterior (MAP) with MRF was employed in contextual classification. Prior and class conditional probability density functions were modelled from contextual information and observed data respectively. Posterior energy models and their parameters are then determined and labelling of pixels done using MAP estimation that equals the minimum energy function. This paper recommends further research in modelling context at the object level to achieve higher classification accuracy. This as per the paper is useful in obtaining a good interpretation of the whole remote sensing scene.

Temporal contextual information in a sequence of temporal images was added in the classification process in the paper by Melgani and Serpico (2003). They did use multitemporal datasets in a MRF based approach extending the temporal information from one image to the other. This improved classification, for they used two multisensors and reduced the uncertainty, however the classification process takes long to execute. Arachchige (2011) integrated information from VHR multispectral and panchromatic imagery, to be able to determine the building footprints. His research study had a number of drawbacks: Firstly, there is spectral confusion between some roof colours with shades. Secondly, individual buildings which are closer to each other are identified as one building whereas smaller buildings were not detected. Lastly, the accuracy of the method developed is directly related to class definition, that is, the classes should be defined with better spectral separation to achieve good classification accuracy. This is usually very difficult.

Probabilistic models with statistical parameters to be estimated are studied by Hassner and Sklansky (1980). The statistical parameters control the size and direction of adjacent similar pixels hence used in texture classification with MRF algorithm.

The related work analysed show that the classification methods employed in the different studies, prior information has not, or has not been efficiently integrated at the object level of classification. Since prior information can improve coherent interpretation of a remote sensing scene, we therefore strive to model the same in contextual classification with MRF in order to enhance classification accuracy.

2.2. The MRF theory

MRF theory is defined by (Li, 2009)as a branch of probability theory for analysing the spatial or contextual dependencies of physical phenomena. It is used in visual labelling to establish probabilistic distributions of interacting labels as shown in figure 2.1. Class label pixels do interact with neighbours to determine the probability that a given class label belong to a certain pixel. They interact with their four and/or eight neighbouring pixels. In the first-order neighborhood system (four neighborhood system), every site has four neighbours whereas in the second order neighborhood system (eight neighborhood system), there are eight neighbours for every (interior) site.

F is said to be a Markov random field on S, the set difference, with respect to a neighborhood system N if and only if the following three conditions are satisfied:

- Positivity: $P(f) > 0, \forall f \in F$
- Markovianity: $P(f_i \mid f_{S-\{i\}}) = P(f_i \mid f_{N_i})$
- Homogeneity: $p\ (w_r / w_{N_r})$ is the same for all sites.

  where $S - \{i\}$ is the set difference, $f_{S-\{i\}}$ denotes the set of labels at the sites in $S - \{i\}$.

A set of random variables F is said to be a Gibbs random field (GRF) on S with respect to N if and only if its configurations obey a Gibbs distribution. A Gibbs distribution takes the form

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \tag{2.1}$$

Where,

$$Z = \sum_{f \in F} e^{-\frac{1}{T}U(f)} \tag{2.2}$$

Z is a normalizing constant called the partition function, T is a constant called the temperature, which shall be assumed to be 1 unless otherwise stated, and $U(f)$ is the energy function. The energy

$$U(f) = \sum_{c \in C} V_c|(f) \tag{2.3}$$

$U(f)$ *is* a sum of clique potentials Vc(f) over all possible cliques C. The value of $Vc(f)$ depends on the local configuration on the clique c. The Gaussian distribution is a special member of this Gibbs distribution family. The posterior, prior and likelihood energies are expressed as: Posterior = Prior + Conditional / Likelihood.

$$U\ (w|d) = U\ (w) + U\ (d|w) \tag{2.4}$$

Where,

- w= class value
- d= pixel value (DN)

The probability density function in equation 2.6 and 2.7 define:

P = Probability, *Z*= Partition function, *T*=Temperature, *U (w)* = Energy function.

$$P\ (w|\ d) = p\ (w) \times p\ (d\ |\ w) \tag{2.5}$$

Thus,

$$P = \frac{1}{Z}\ exp\left(-\ \frac{U(w)}{T}\right) \tag{2.6}$$

$$P \sim exp\left(-\ \frac{U(w)}{T}\right) \tag{2.7}$$

	0	
0	**x**	**0**
	0	

0	**0**	**0**
0	**X**	**0**
0	**0**	**0**

5	**4**	**3**	**4**	**5**
4	**2**	**1**	**2**	**4**
3	**1**	**X**	**1**	**3**
4	**2**	**1**	**2**	**4**
5	**4**	**3**	**4**	**5**

(a) (b) (c)

Figure 2.1: Neighborhood and cliques on a lattice of regular sites

2.2.1. The Smoothness prior

Smoothness is a generic contextual constraint as defined by (Li, 2009). The smoothness assumes that physical characteristics in a neighbourhood of space or in an interval of time present some coherence and generally do not change abruptly. For instance, the surface of a table is flat, a meadow presents a texture of grass, and a temporal event does not change abruptly over a short period of time. We can indeed find regularities of a physical phenomenon with respect to certain properties. The smoothness is expressed as the prior probability or equivalently an energy term $U(w)$, measuring the extent to which the smoothness assumption is violated by w.

2.2.2. Maximum Likelihood

Given a realization f of a single MRF, the maximum likelihood (ML) estimate maximizes the conditional probability $P(w|d)$ (the likelihood of d) or its log-likelihood ln $P(w|d)$. Note that in this case f *is* the data used for the estimation. When the prior probability density function of the parameters, $p(d)$, is known, the MAP estimation that maximizes the posterior density can be sought. The ignorance of the prior probability density function leads the user to a rather diffuse choice. The prior probability density function is assumed to be flat when the prior information is totally unavailable. In this case, the MAP estimation reduces to the ML estimation.

2.2.3. Supervised classification

Classes are the ground features, for example, grass, bare soil, trees, buildings. Thus, supervised classification as defined by (ITC Core book, 2010) is the 'partitioning' of the feature space by an operator who defines the spectral characteristics of the classes by identifying sample areas (training areas). The operator should be familiar with the area of interest or information derived from dedicated field observations.

A sample of a specific class, comprising a number of training cells, forms a cluster in the feature space. The clusters, as selected by the operator:

- Should form a representative data set for a given class. This means that the variability of a class within the image should be taken into account. Also, in an absolute sense, a minimum number of observations per cluster is required. Although it depends on the classifier algorithm to be used, a useful rule of thumb is 30 × n (n-number of bands) observations.
- Should not or only partially overlap with the other clusters, otherwise a reliable separation is not possible. For a specific data set, some classes may have significant spectral overlap, which, in principle, means that these classes cannot be discriminated by image classification. Solutions are to add other spectral bands, and/or add images acquired at other moments.

2.2.4. Maximum Likelihood Classifier

The Maximum Likelihood (ML) classifier considers not only the cluster centres but also the shape, size and orientation of the clusters. This is achieved by calculating a statistical distance based on the mean values and covariance matrix of the clusters. The statistical distance is a probability value: the probability that observation d belongs to specific cluster. A cell is assigned to the class (cluster) to which it has the highest probability.

$$P(d,w) = P(w|d)\,P(d) = P(d|w)P(w) \qquad (2.8)$$

P(d,w) is the probability of coexistence (or intersection) of events *d* and *w*, *P(d)* and P(w) are the probabilities of events *d* and w, and $P(w/x)$ is the conditional probability of the event *d* given event w. If event d_i is the i^{th} pattern vector and w_j is the information class j then, according to Equation 2.8, the probability that d_i belongs to class w_j is thus:

$$P(w_j|d_i) \propto P(d_j|w_i)\, P(w_j) \,|\, P(d_i) \qquad (2.9)$$

Since *P(x)* is set to be uniformly distributed (i.e., the probability of occurrence is the same for all pixel features), then:

$$P(w_j|d_i) \propto P(d_j|w_i)\, P(w_j) \qquad (2.10)$$

One can thus allocate pixel i to the class k, which has the highest value of the term $P(w_j/x_i)$ in the equation above. The classification criterion can be expressed as:

$$W_k = \underset{\forall w_j}{\arg\max} \left\{ P(d_i|w_j) P(w_j) \right\} \qquad (2.11)$$

where arg denotes argument. The criterion shown in the equation above is the Maximum A Posteriori (MAP) solution, which maximizes the product of conditional probability and prior probability.

2.2.5. Class Separability

Class separability is a quantitative measure of how well classes can be separated. There are four widely used quantitative measures for the class separability: divergence D, transformed divergence TD, Bhattacharyya distance *B*, and Jeffries–Matusita distance JM. For any pair of classes $\propto$ and β, these measures are defined as follows.

1) Divergence

$$D_{\propto\beta} = \frac{1}{2}(\mu_{\propto} - \mu_{\beta})'(\mathbf{C}_{\propto}^{-1} + \mathbf{C}_{\beta}^{-1})(\mu_{\propto} - \mu_{\beta}) + \frac{1}{2}\mathrm{Tr}\left[\frac{1}{2}(\mathbf{C}_{\propto} - \mathbf{C}_{\beta})(\mathbf{C}_{\beta}^{-1} - \mathbf{C}_{\propto}^{-1})\right] \qquad (2.12)$$

2) Transformed divergence

$$TD_{\propto\beta} = 2(1\text{-}e^{-D_{\propto\beta}/8}) \qquad (2.13)$$

3) Jeffries–Matusita distance

$$JM_{\propto\beta} = 2(1\text{-}e^{-B_{\propto\beta}}) \qquad (2.14)$$

Transformed divergence works well for both small and large $E_{\alpha\beta}$ Transformed divergence and Jeffries–Matusita distance take values between 0 and 2, whereas divergence vary from 0 to ∞. If the parameters of two classes are identical, i.e., $\mu_{\propto} = \mu_{\beta}$ and $C_{\propto} = C_{\beta}$, then $D_{\propto\beta} = TD_{\propto\beta} = B_{\propto\beta} = JM_{\propto\beta} = 0$, indicating that it is not possible to discriminate between the classes $\propto$ and β based on spectral information. Separability between the classes increases with increasing TD values. Transformed divergence and Jeffries–Matusita distance are preferred over the other two measures because of their saturated behaviour for the large values of $\mu_{\propto}$ - μ_{β}. Both TD and JM are widely used for both feature selection and refinement of spectral signatures. μ is the mean vector of the classes whereas $\mathbf{C}$ is the covariance matrix.

2.3. Software

In Imagine objective and Ecognition, tree crowns are measured by circularity and shape index respectively. These measurement approaches are described in url1 and url2 in the list of references. These two measurement approaches shall be integrated in the Hierarchical Markov Random Field (HMRF) classification technique in the methods chapter. They shall be incorporated in MRF object based level.

ECognition by Definiens Imaging interpret and analyse remote sensing imagery from an object view point instead of pixels alone. This remote sensing tool not only considers spectral value information but also spatial relationships between objects.

2.3.1. Imagine objective

2.3.1.1. Circularity

This metric is one method of measuring how close an object is to a circle. The result is computed as follows: Firstly, a centre point is computed by averaging the coordinates of all points in the raster object, secondly, the distances from each point on the raster object to the centre point is calculated, thirdly, the standard deviation of the distances is computed and lastly the standard deviation is subtracted from 1.0. If the result is less than zero, it is set to 0.0.

A perfect circle will have a circularity of 1.0. Some other shapes such a squares, rectangles and regular polygons will also have very high circularity. Also concave polygons shaped like a thin letter "C" will rank high in this metric. To distinguish between true circles and other polygons that would rank high using this metric, other metrics such as area or compactness are used. Even though the results of circularity range from 0 to 1, it cannot be considered a probabilistic metric perimeter point.

Tree crown circularity is shown in figure 2.2(a). The diameter axes are m and n, whereas b and a are the x and y pixels that define the entire tree crown.

2.3.2. Area and simple complexity descriptors

This metric computes the area in square metres of each polygon shape in the input shapefile. The area of islands within the other polygon is subtracted from the outer polygon's area. The output shapefile will have an area attribute containing the result of this metric.
Complexity is an important property of shapes. Geometric properties of shapes such as spatial coverage are of essence in classification. Circularity is one such shape descriptor that is defined thus:

Circularity = P^{2} / A, where P, denotes the shape perimeter and A, the shape area.

2.4. Ecognition

2.4.1. Shape Index: [for 2D Image Objects]

The Shape index (SI) describes the smoothness of an image object border. The smoother the border of an image object is, the lower its shape index. It is calculated from the Border Length feature of the image object divided by four times the square root of its area.

Parameters

- b_v is the image object border length
- $\sqrt[2]{P_v}$ is the border of square with area P_v

Expression

$$\mathrm{SI} = \frac{b_v}{\sqrt[2]{P_v}} \qquad 2.15$$

Feature Value Range

[1;∞] ; 1 = ideal.

(a) (b)

Figure 2.2: Shape index of an image object v as shown in (url 1)

2.5. Shape classification

In (Da Fontoura Costa and Cesar 2001) , it is illustrated that shape classification portray a recognition problem when given an input shape to decide whether or not it belongs to some specific predefined class. This shape recognition problem is known as supervised classification. When the shape classes are predefined, or examples are available for each class, it's often desirable to create algorithms that take a shape as input and assign it to one of the classes.

2.5.1. Shape in images

Shape in images is explored by (Dryden and Mardia 1998) where low and high level image analysis techniques are employed to be able to extract object features from remotely sensed images. Low level image analysis involves techniques at a pixel by pixel level, for instance, classifying each pixel into classes. On the other hand, high level image analysis entails direct modeling of objects in images, object recognition and object location. In high level image analysis, shape analysis has a prominent role in improving classification accuracy.

2.5.2. Prior models for objects

Dryden and Mardia (1998) demonstrate the key to the successful inclusion of prior knowledge on classification in high level Bayesian image analysis through specification of the prior distribution. The prior can be specified either through a model with known parameters or with parameters estimated from training data.

2.5.3. Simulated Annealing

Simulated annealing is defined in the book (Li, 2009) as a stochastic algorithm for combinatorial optimization. It simulates the physical annealing procedure in which a physical substance is melted and then slowly cooled down in search of a low energy configuration. The following formula shows the relation in a system for any f, the probability and temperature.

$$P_T(f) = P(f)^{1/T} \qquad 2.16$$

$$P_T(f) = e^{-E(f)/T} \mid \sum_f e^{-E(f)/T} \qquad 2.17$$

Where,

$T > 0$ is the temperature parameter

$T \to \infty$, $P_T(f)$ is a uniform distribution

$E(f)$ is the energy of a set to a random configuration

Initially, T is set very high and f is set to a random configuration. At a fixed T, the sampling is according to the Gibbs distribution. After the sampling converges to the equilibrium at current T, T is decreased according to a carefully chosen cooling schedule. This continues until T is close to 0, at which point the system is "frozen" near the minimum of $E(f)$. Simulated annealing states: If the decreasing sequence of temperatures satisfies the following equations, then the system converges to the global minimum regardless of the initial configuration $f^{(0)}$.

$$\lim_{t\to\infty} T^{(t)} = 0 \qquad 2.18$$

$$T^{(t)} \geq \frac{m \times \Delta}{\ln(1+t)} \qquad 2.19$$

Where,

$\Delta = max_f\, E(f) - min_f\, E(f)$

According to (Geman et al., 1984), temperature is defined as shown in the formula below. Where for every t, c is a constant independent of t and is set to c = 3.0 or c = 4.0, then with probability converging to one ($t \to \infty$), the configurations generated by the algorithm will be those of minimal energy. Put another way, the algorithm generates a Markov chain which converges in distribution to the uniform measure over the minimal energy configurations:

$$T^{(t)} = T \times T_{upd} \;, \qquad 2.20$$

T_{upd} is the temperature update which takes values between zero and close to one.

$$T^{(t)} = \frac{c}{ln(1+t)} \qquad 2.21$$

The logarithmic annealing schedule in equation 2.23 takes iterations in the tune of more than five thousand for convergence to be reached. Equation 2.22 is thus preferred, so long as Tupd is not set close to 1, as the cooling annealing schedule for the HMRF method.

3. Study area and data preparation

3.1. Study area

Wanjohi area in the Lake Naivasha basin is the application area of this research topic. It's located in the north western part of central Kenya on latitudes 0^0 16' 39" and 0^0 18' 53", and, longitudes 36^0 28' 28" and 36^0 30' 41".

The study area has individual tree crowns of specific species that have a round shape and a definite size.

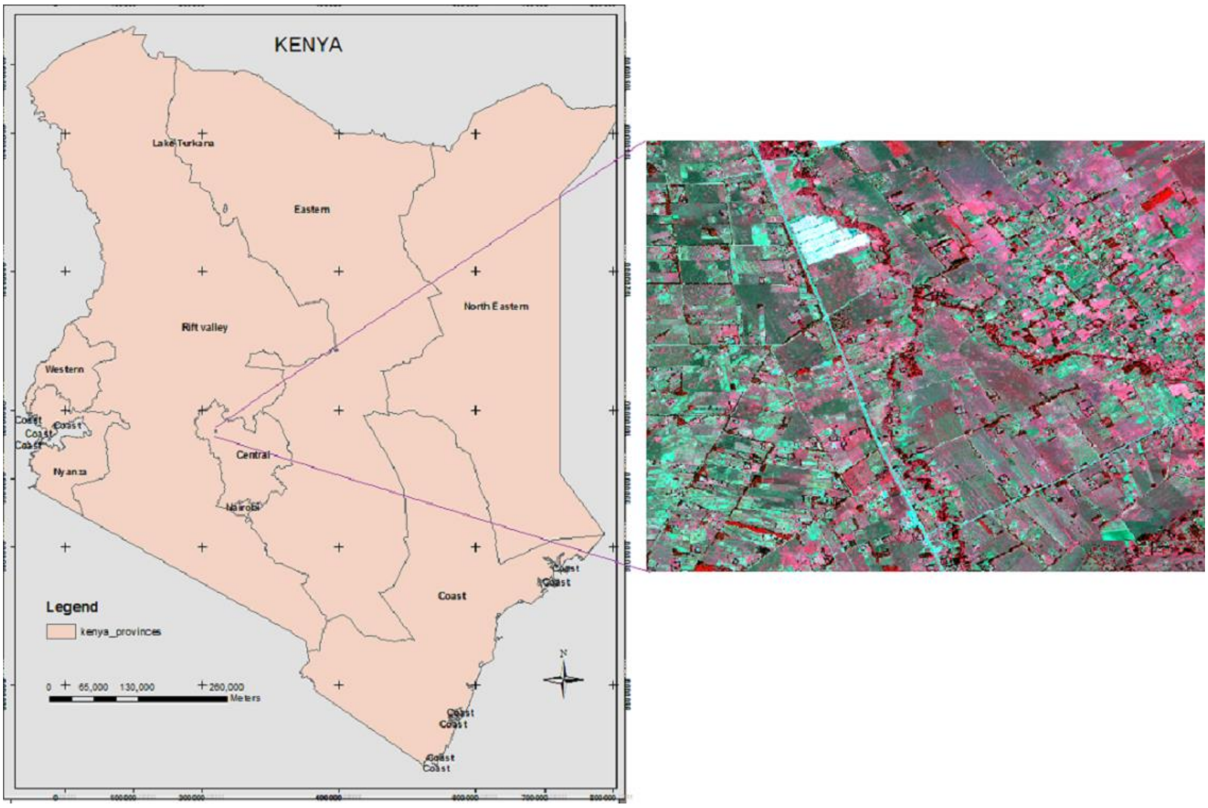
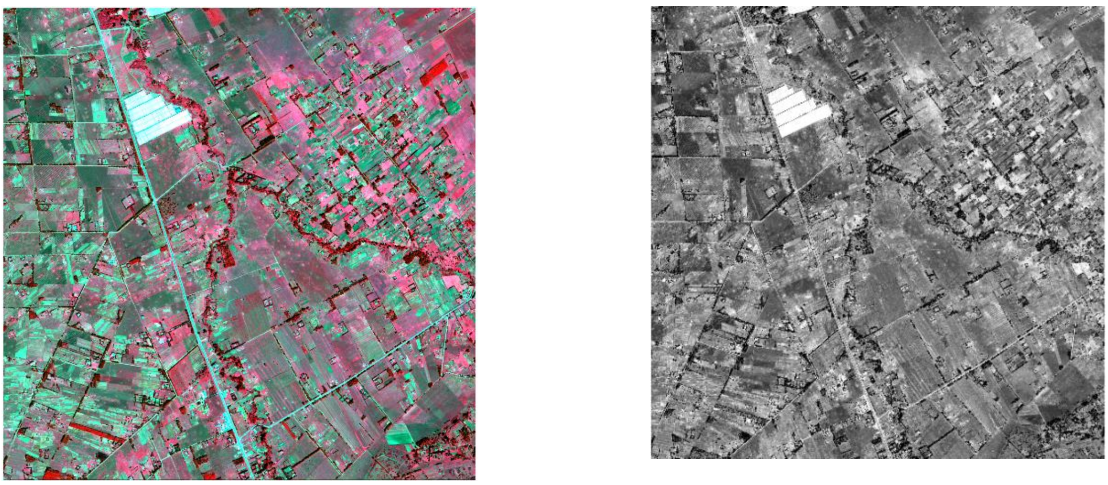


Figure 3.1: Map of Kenya showing study area on the left side and multispectral image on the right.



(a) (b)

Figure 3.2: (a) and (b) WorldView-2 multispectral and panchromatic (PAN) image

3.2. WorldView-2

In this research, a high-resolution 8-band multispectral imagery was used. This image is of 2.0 metre spectral resolution. The swath width is 16.4 km at nadir, collecting a capacity of 975,000 km² / day with average revisit of 1.1 days.

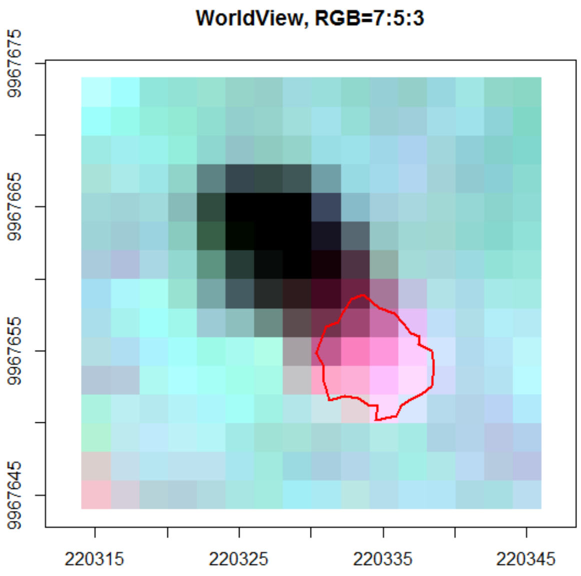
Bands NIR1, Red and Green (753) to RGB false color composite were used for visualisation. They clearly identify healthy vegetation. Bands 532 to RGB are the true color composite.

3.3. Subsets for this study

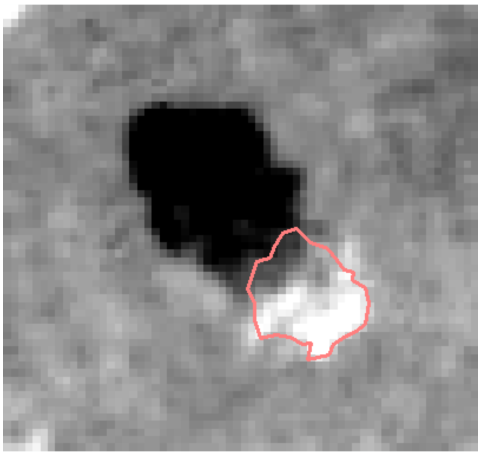
Three subset images were extracted from the high resolution multispectral image. Subset one is the tuning subset whereas subset two and three are the implementation areas. The band combination for the multispectral subsets is 753 (NIR1, Red, Green) for visualisation.

The panchromatic subset images in figure 3.3, 3.4 and 3.5 show the object tree crown, shadow and background classes. Tree crown is digitized and used as a reference vector layer. The panchromatic image of 0.5 meter is of higher accuracy than the multispectral image, referred in (url4), the reason why we do use it as reference.

3.3.1. Subset one



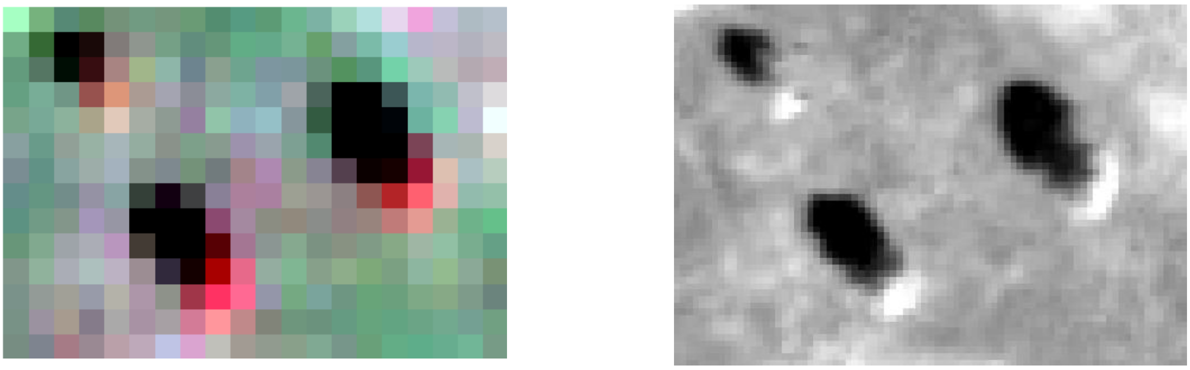
(a)



(b)

Figure 3.3: Tuning subset (a) of multispectral image in WGS84 system and reference subset (b) of PAN image.

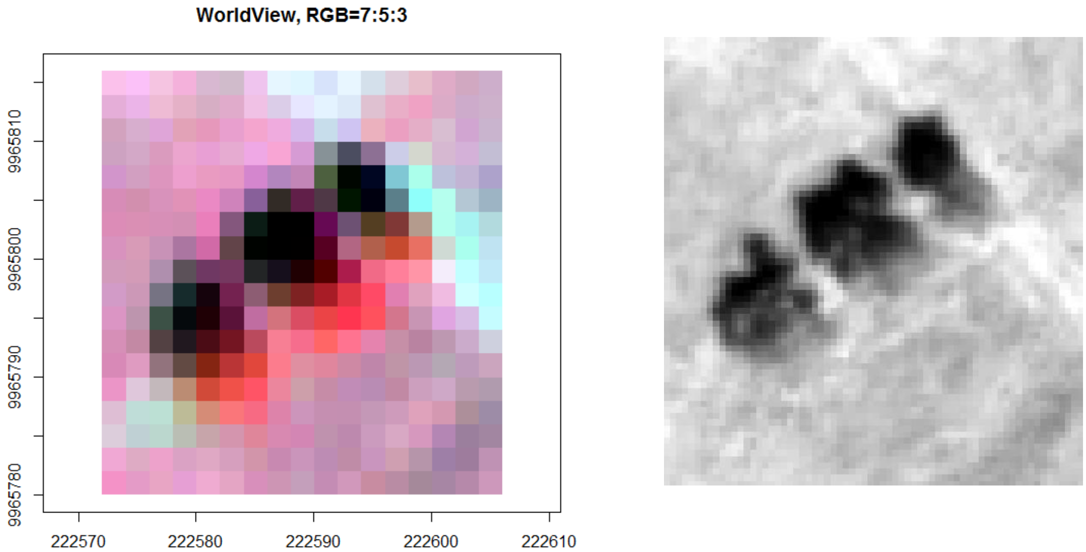
3.3.2. Subset two



(a) (b)

Figure 3.4: (a) and (b) Three tree crowns separated on a MS image (left) and PAN (right)

3.3.3. Subset three



(a) (b)

Figure 3.5: Three tree crowns partially interlocked on a MS image (left) and PAN (right)

3.3.4. Worldview-2 spectral bands and their characteristics (roles)

Band label	Band (nm)	Characteristic and role
Coastal Blue Band 1	(400-450 nm)	• Absorbed by chlorophyll in healthy plants and aids in conducting vegetative analysis
Blue Band 2	(450-510 nm)	• Identical to QuickBird • Readily absorbed by chlorophyll in plants
Green Band 3	(510-580 nm)	• Narrower than the green band on QuickBird • Able to focus more precisely on the peak reflectance of healthy vegetation • Ideal for calculating plant vigor • Very helpful in discriminating between types of plant material when used in conjunction with the Yellow band
Yellow Band 4	(585-625 nm)	• Very important for feature classification • Detects the “yellowness” of particular vegetation, both on land and in the water
Red Band 5	(630-690 nm)	• Narrower than the red band on QuickBird and shifted to longer wavelengths • Better focused on the absorption of red light by chlorophyll in healthy plant materials • One of the most important bands for vegetation discrimination • Very useful in classifying bare soils, roads, and geological features
Red-Edge Band 6	(705-745 nm)	• New band • Centered strategically at the onset of the high reflectivity portion of vegetation response • Very valuable in measuring plant health and aiding in the classification of vegetation
NIR1 Band 7	(770-895 nm)	• Narrower than the NIR1 band on QuickBird to provide more separation between it and the Red-Edge sensor • Effectively separates water bodies from vegetation, identifies types of vegetation and also discriminates between soil types
NIR2 Band 8	(860-1040 nm)	• New band • Overlaps the NIR1 band but is less affected by atmospheric influence • Enables broader vegetation analysis and biomass studies

Table 3-1: Worldview-2 spectral bands and their characteristics (roles)

Source: http://worldview2.digitalglobe.com/docs/WorldView-2_8-Band_Applications_Whitepaper.pdf

4. Methods

The book (Li, 2009) defines MRF theory as a branch of probability theory for analyzing the spatial or contextual dependencies of physical phenomena. It is used in visual labeling to establish probabilistic distributions of interacting labels. The MRF model exploits spatial class dependencies (spatial context) between neighboring pixels in an image, and temporal class dependencies between different images of the same scene as demonstrated in (Solberg et al., 1996).

Markov random field (MRF) models play an important role, due to their ability to integrate the use of contextual information associated with the image data in the analysis process, through the definition of suitable energy functions (Serpico and Moser 2006).

In this chapter, a Hierarchical MRF technique is described. This technique is two leveled, level one is MRF pixel-based whereas level two is MRF object-based.

4.1. Supervised maximum likelihood classification

Supervised classification was performed in ENVI software. Training data was taken for defining classes: Tree crowns=190pixels, bare soil bright=262pixels, grass dense=185pixels, earth road=111pixels, tarmac road=132pixels, buildings=481pixels, shadow=327pixels, grass sparse=187pixels, bare soil dark=262pixels. These pixels were picked the number of pixels being more than 10n and less than 100n (n=number of bands). Mean vectors and covariance matrices for the classes were derived in ENVI, and the class separability by transformed divergence demonstrated in the software and the feature space.

The subset image for this method has four classes: Tree crown, bare soil dark, shadow and grass dense. For each of the classes, mean vectors, covariance matrices for all the eight bands of the multispectral image were used. Tree crown spectral class is highly confused with grass and other vegetation hence in this chapter we develop a contextual HMRF method that incorporate shape and area to improve classification of tree crowns. The supervised maximum likelihood classification result is shown in figure 5.1 in the chapter 5.

4.2. The MRF Pixel based level

At this level, temperature is optimised by energy minimization technique simulated annealing.

4.2.1. Prior energy functions

The prior energy function models spatial context and penalizes the occurrence of pixels with different class labels in the neighbourhood system N. This means that spatial configurations of adjacent pixels labelled as tree crowns are more likely to occur than isolated ones. Weights $w(a_i)$ are chosen inversely proportional to the distance d (a_j) between the central pixel $a_{j/i}$ and the pixel a_i

$$U(c) = \sum_{i,j} U(C(a_{j/i})) = \sum_{i,j} \sum_{a_i \in N(a_{j/i})} w(a_i)\, I(C(a_{j/i}), C(a_{j/i})) \qquad 4.1$$

$$w(a_i)\ a\ \frac{1}{d(a_{j/i}, a_i)} \qquad 4.2$$

Where,

$N(a_{j/i})$ = The neighbourhood system

$U(C(a_{j/i})$ = The local contribution to the prior energy from pixel $C(a_{j/i})$

$w(a_i)$ = The weight of the contribution from neighbour pixel $a_i \in N(a_{j/i})$ and I (α,β) takes the value 0 if α=β and 1 otherwise

4.2.2. Prior probability

Probability *p(w)* is inversely proportional to energy *U(w)*. Energy need to be minimised in order to increase probability.

$$P\ (w) = \frac{1}{Z}\ exp\left(-\ \frac{U_{(w)}}{T}\right),\ Z= exp\ \left(-\ \frac{U_{(w)}}{T}\right) \qquad (4.3)$$

Where,

U (w) is the prior energy for a class label w
P (w) is the probability for class label w
T is a constant termed temperature
Z is the partition function

4.2.3. Conditional energy functions

In the paper (Ardila et al., 2011), the conditional energies consider the proximity of observed pixel values *y* to each land cover class. They model spectral values *x* of a class α with the Gaussian distribution. An assumption is made that values *x* are spatially uncorrelated given their class association. In this case the spectral values in *y* also follow the Gaussian distribution. The conditional term $U\ (y/c\ )$ for the multispectral image is given as:

$$U\ (d/w) = \sum_i \frac{1}{2}\left[M(d(b_i), \mu_i, \mathbf{C}_i) + \frac{1}{2}\ln|\det \mathbf{C}_i|\right] \qquad (4.4)$$

Mahalanobis distance (MD) is described by (De Maesschalck et al., 2000) in the equation below:

$$MD_i = \sqrt{(x_i - \bar{x})\mathbf{C}_x^{-1}(x_i - \bar{x})^T} \qquad (4.5)$$

Where,

$\mathbf{C}_x = 1|(n-1)(x_c)^T(x_c)$

$M(d(b_i), \mu_i, C_i)$ = Mahalanobis distance between $d(b_i)$ and μ_i with C_i.

μ_i = Mean vector

$\mathbf{C}_i$ = Covariance matrix

The values of μ_i and C_i are modelled as linear mixtures of mean vectors and covariance matrices based on area proportions of respective land cover classes $C(a_{j/i})$ inside the pixel b_i.

4.3. The MRF object based level

4.3.1. Shape and area contribution

Prior information in this method is the shape and size of tree crown. Shape of tree crown is circular and the area of the crown is computed as in equation 4.6. Lambda shape parameter that controls the balance between the shape and the area models in the segment energy functions gives more weight to shape of tree crown than area.

$$A=\pi \times (D/2)^2 \qquad 4.6$$

Where,

A = Area

D = Diameter of tree crown

Modeling shape and area is done at the prior term of the HMRF model. Lambda, lambda segments, lambda shape and area are the smoothness assumption based on the concept of context (Smoothness prior).

Tree crown diameters were measured on google earth by measurement tools. A sample size of hundred tree crowns was taken. The tree crown statistics were then computed in Microsoft access: Standard deviation, area, mean,median. The probability density functions for tree crown area and diameter are illustrated in figure 4.1

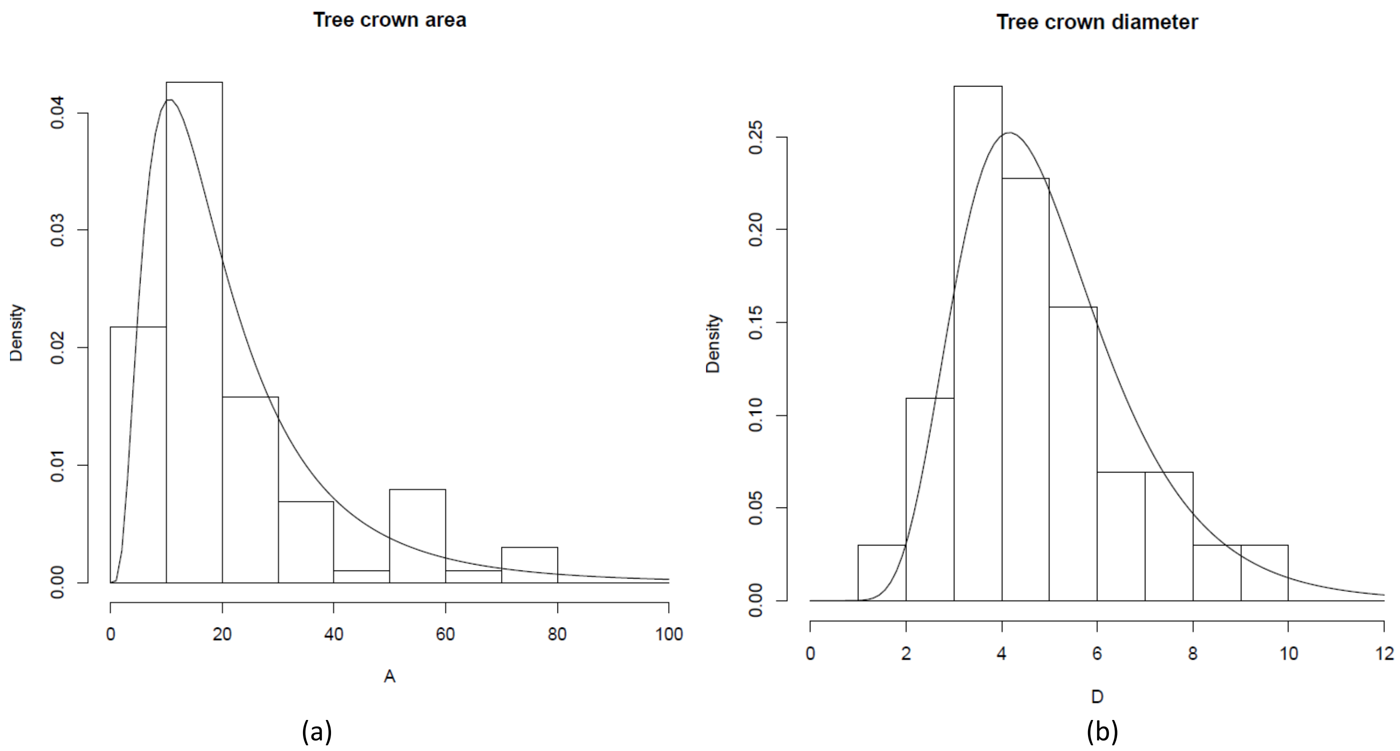


Figure 4.1: (a) and (b) Probability density function for tree crown area and diameter

The MRF theory formulae,

$$U(w|d) = \lambda U(w) + (1-\lambda)U(d|w) \qquad 4.7$$

Where,

$U(w/d)$ = Posterior energy
$U(w)$ = Prior energy
$U(d/w)$ = Likelihood energy
λ = Lambda parameter

From the MRF and upon introduction of lambda segments, lambda shape parameters, it can be deduced thus:

$$U(w/d) = (1-\lambda_{seg})(\lambda U_{prior}(i,j) + (1-\lambda)U_{pix}(i,j)) + \lambda_{seg}(U_{seg}(i,j,Sc)) \qquad 4.8$$

Where,

λ_{seg}= Lambda segments
U_{prior} (i, j) = Prior energy
U_{pix} (i, j) = Likelihood energy
U_{seg} (i, j, Sc) = Energy for segments in shape analysis.

4.4. Accuracy assessment

Accuracy assessment as described by (ITC Core book, 2010), is performed by an error/confusion matrix which compares samples taken and evaluates them with the reference data. This matrix allows calculation of quality parameters: Overall accuracy, error of omission and error of commission.

Producer accuracy also termed the error of omission or type II error is the probability that a sampled point on the map is indeed that particular class. Producer accuracy= A|B; A= number of correctly classified tree crown pixels, B=number of all tree crown pixels.

User accuracy or the error of commission (type I error), is the probability that a certain reference class has indeed actually been labelled as that class. The error of commission refers to incorrectly classified samples.

Overall accuracy is the number of correctly classified pixels, (that is, sum of the diagonal cells in the error matrix) divided by the total number of pixels checked. The overall accuracy yields one value for the result as a whole. Overall accuracy= A|C; A=number of correctly classified tree crowns, C= number of total pixels for all classes.

Kappa is a measure of accuracy which takes into account the fact that even assigning labels at random will result in a certain degree of accuracy.

4.5. Workflow diagrams for HMRF method

The workflow diagram below is a sequence of steps in the methodology that describes the execution of Hierarchical Markov Random Field (HMRF) contextual classification technique. MLC is performed on one tile of the multispectral image; the resultant classified image is used as a control in reference to the HMRF classification result to compute accuracy and analysis. Extracted subset from the multispectral image is used in this method for tuning MRF parameters to optimization.

The HMRF is a two tier contextual classification technique that does classify tree crowns in enhanced accuracy. Stage one is the MRF pixel based classification approach followed by stage two which is MRF object based approach. MRF pixel based classification is the first level in the development of the method. Lambda segments parameter is set to zero resulting a maximum likelihood classification. Simulated annealing algorithm estimates the optimal values for temperature and temperature update. MRF object based level involves shape and area contribution in the classification process where lambda and lambda segments are estimated to output the most accurate and reproducible MRF classification.

On the other hand, lambda shape parameter controls the balance of contribution in shape and area in the MRF objective energy functions. Shape analysis being of critical importance in the classification of tree crown, it is awarded more weight than the area contribution. Reference data stage entails the vectorisation of a tree crown vector that best discerns the crown boundary. This reference is derived from a panchromatic image, moreover, used in the computation of the confusion matrix that determines the classification accuracies. Accuracy assessment is done to define whether or not the method does improve classification.

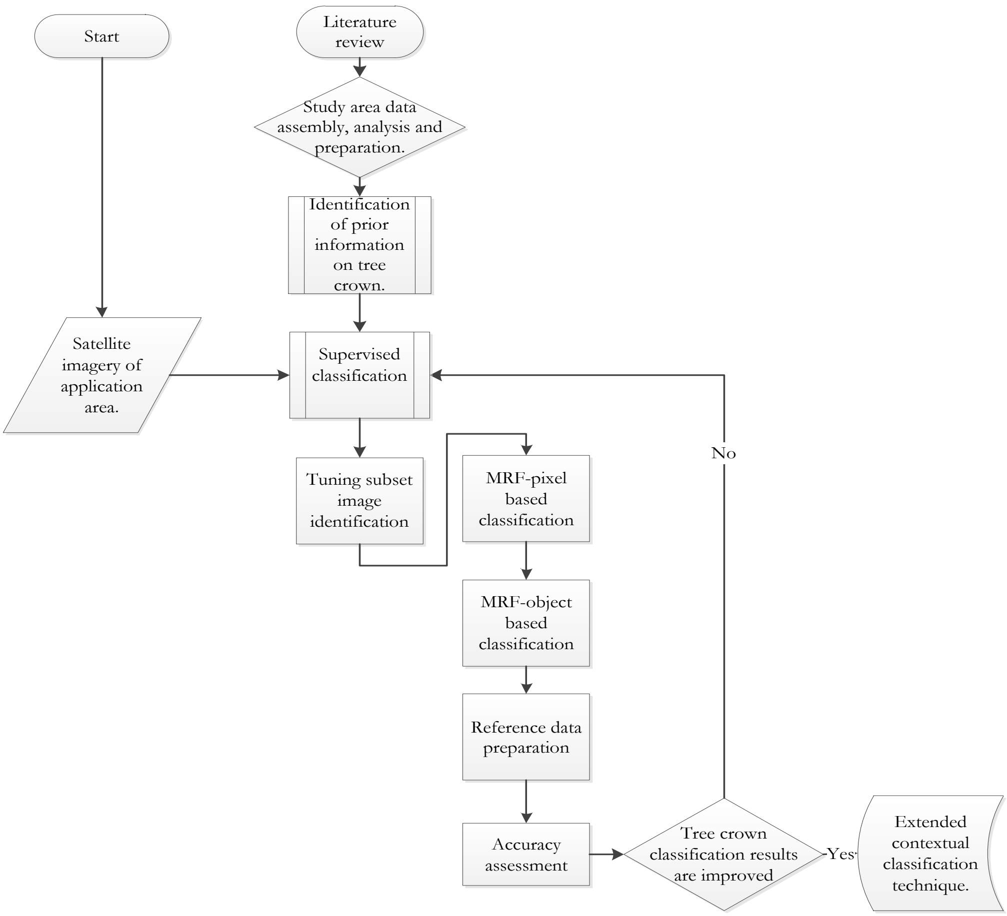


Figure 4.2: Sequence of steps in HMRF approach

The workflow diagram describes the supervised image classification process illustrated in section 2.2.3 in the Literature review chapter. Training set data is picked from the multispectral image. Nine classes are defined out of which mean vectors and covariance matrices of four classes of the subset are taken. Class separability is analysed and this data is hence used in our HMRF method.

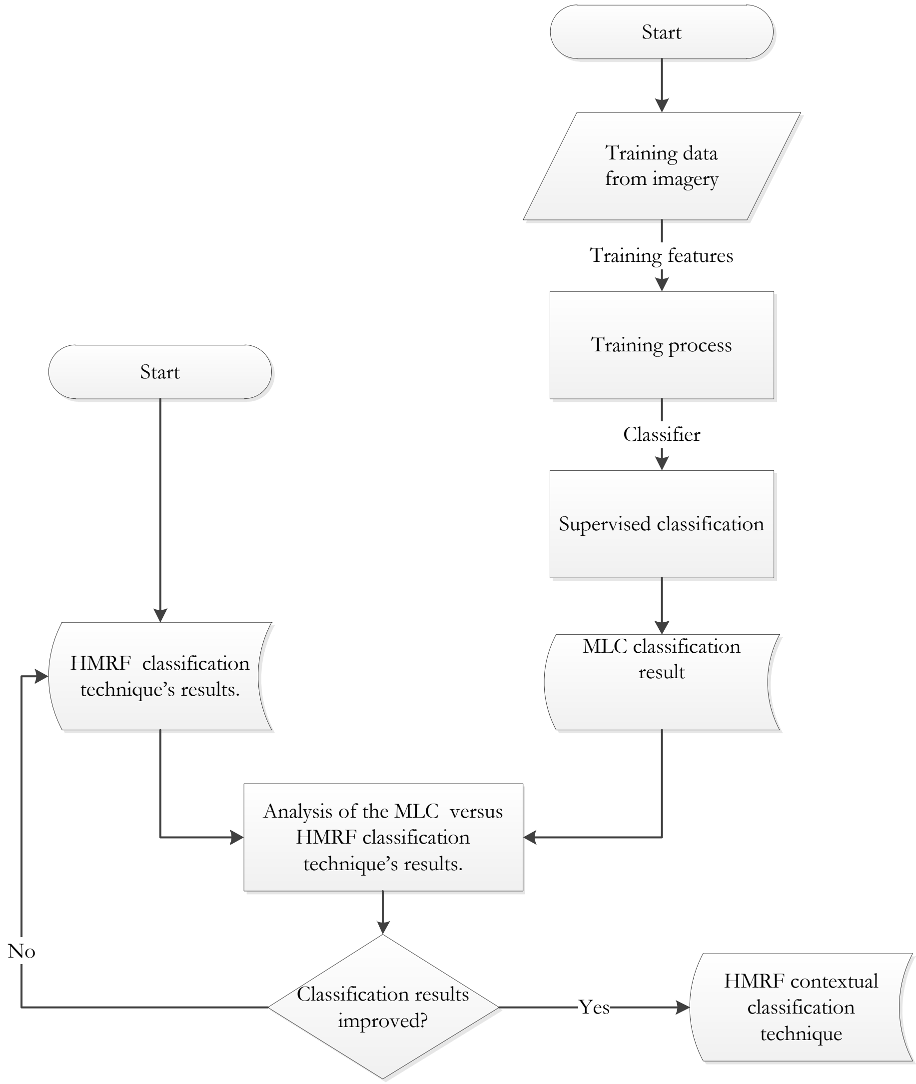


Figure 4.3: Supervised maximum likelihood image classification procedure

5. Results

The output of maximum likelihood classification is shown in figure 5.1 and expounded in section 4.1.

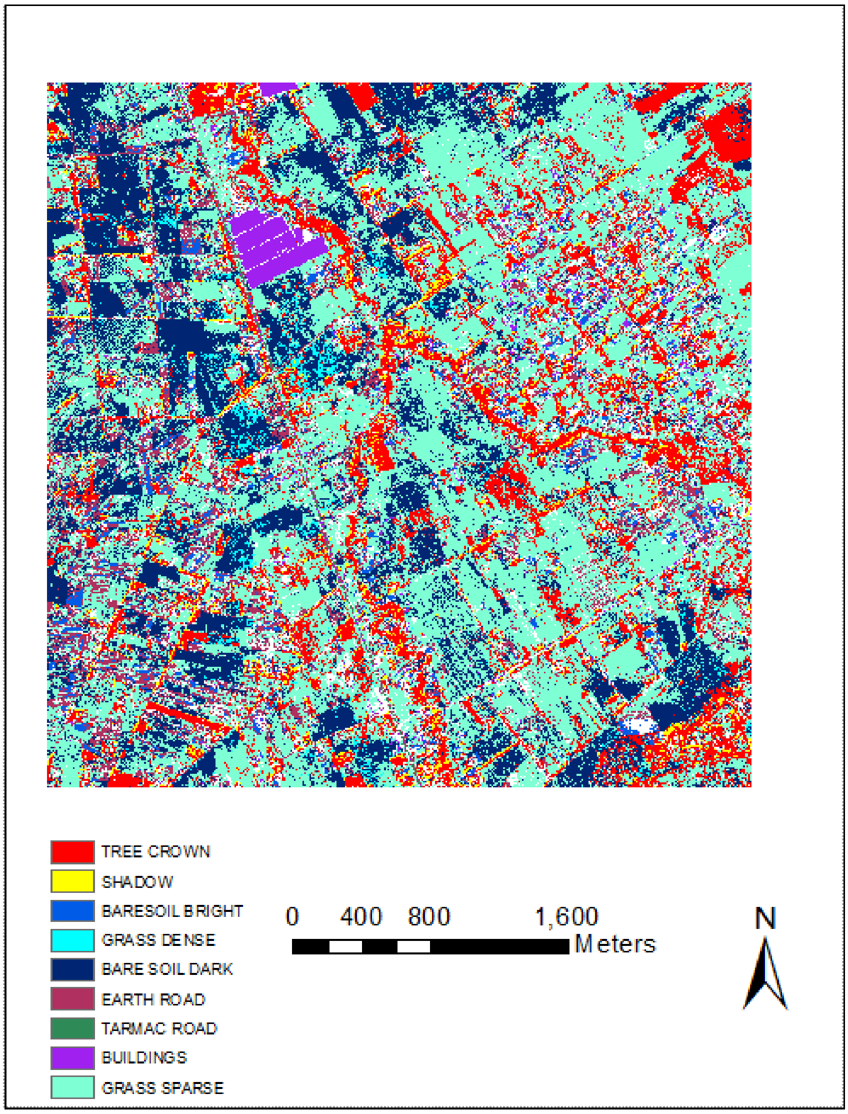


Figure 5.1: Supervised maximum likelihood classification of the multispectral image of the study area.

5.1. Temperature and temperature update optimization

Energy minimization for varying temperature update (Tupd) and temperature (T0) by simulated annealing. Fourty runs were used for different values of T0 and Tupd for convergence. Error bars were generated to portray the standard deviation. The standard deviation for Tupd 0.8, 0.9 are large as described in figure 5.2 compared to 0.95, 0.97 and 0.99.

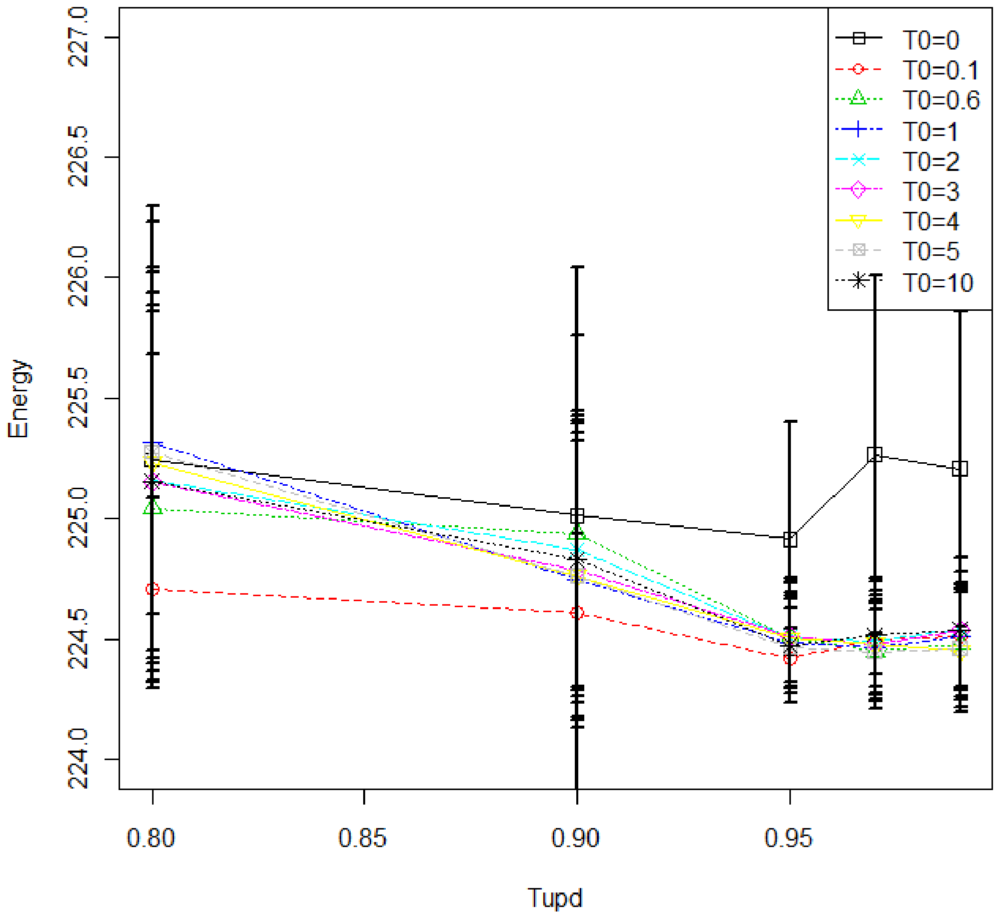


Figure 5.2: Energy minimization for varying temperature update (Tupd) and temperature (T0) by simulated annealing.

The table below shows the energy summary for T0_list (0.0,0.1,0.6,1.0,2.0,3.0,4.0,5.0,10.0) and Tupd_list (0.8,0.9,0.95,0.97,0.99).

Table 5-1: The table above shows the energy summary for T0 and Tupd values

T0 / Tupd	0.8	0.9	0.95	0.97	0.99	Min. energy T0
0.0	225.2452	225.0133	224.9156	225.2665	225.2066	224.9156
0.1	224.7076	224.6120	224.4209	224.5018	224.5126	224.4209
0.6	225.0411	224.9399	224.5046	224.4506	224.4809	224.4506
1.0	225.3151	224.7481	224.4843	224.4639	224.5115	224.4639
2.0	225.1578	224.8722	224.5065	224.4893	224.5444	224.4893
3.0	225.1542	224.7868	224.5140	224.4812	224.5371	224.4812
4.0	225.2339	224.7632	224.5023	224.4802	224.4534	224.4534
5.0	225.2781	224.7632	224.4658	224.4483	224.4613	224.4483
10.0	225.1545	224.8310	224.4738	224.5186	224.5411	224.4738
Min. energy Tupd	224.7076	224.6120	224.4209	224.4483	224.4534	Least Min.energy 224.42.9

The figure 5.2 and table 5-1, demonstrate the simulated annealing optimisation of temperature and temperature update in the cooling schedule of the tuning subset. The set optimal parameters after energy minimisation are T0=3.0 and Tupd=0.99. These values were arrived after a careful examination of the energy values with respect to computational time. The optimal values taken for simulated annealing are T0=3.0 and Tupd=0.99 for they have a small mean, low standard deviation and high reproducibility of MRF results.

5.2. MRF Lambda estimation

Lambda and lambda segment are estimated by use of the kappa coefficient, User accuracy, producer accuracy and reproducibility of several repetitions of MRF results with varying lambda and lambda segment to estimate the most optimal values. Fourty runs for each λ were used to illustrate error bars.

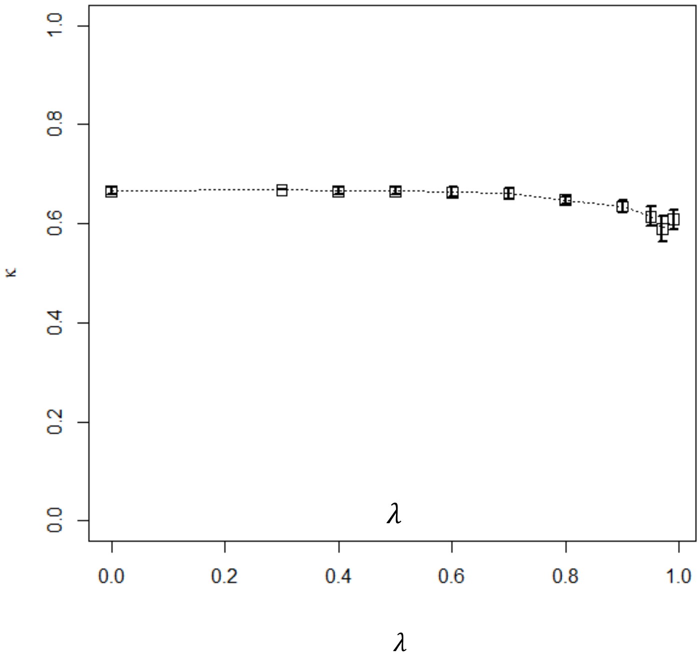
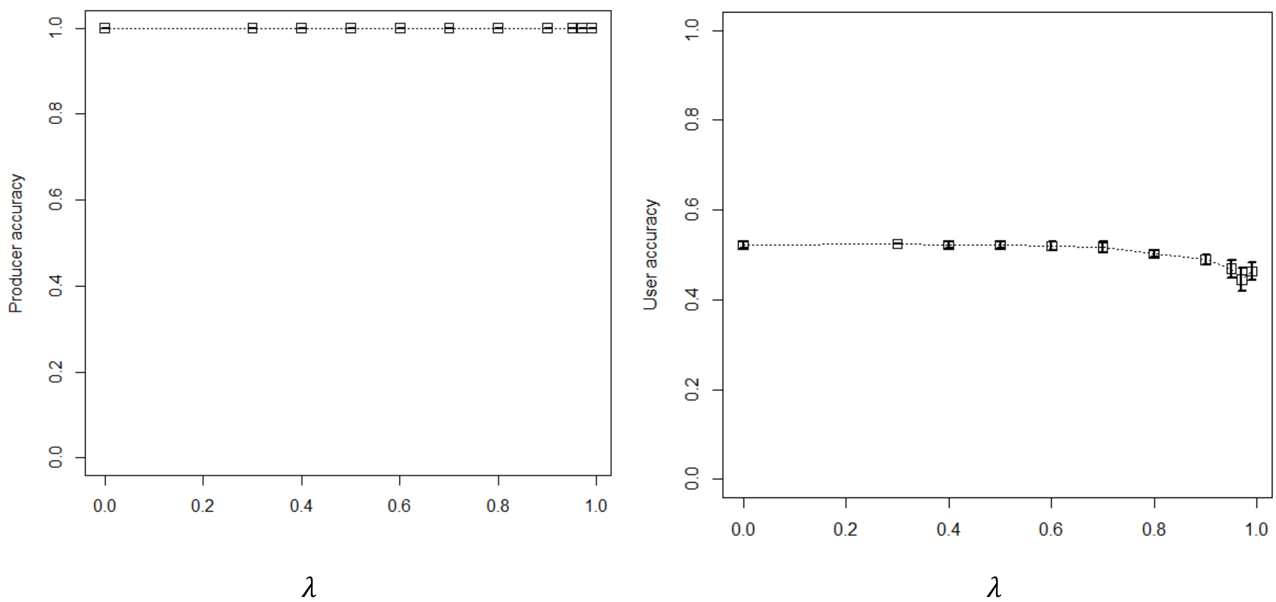


Figure 5.3: Kappa (κ) coefficients versus λ values.



(a) (b)

Figure 5.4: (a) and (b) Producer and user accuracies versus λ

Table 5-2: λ values with respect to kappa, kappa standard deviation, User and producer accuracy

λ	Kappa	Kappa_sd	User accuracy	Producer accuracy
0	0.665	0.0071	0.521	1.00
0.3	0.667	0.0000	0.524	1.00
0.4	0.665	0.0071	0.521	1.00
0.5	0.665	0.0071	0.521	1.00
0.6	0.663	0.0095	0.519	1.00
0.7	0.661	0.0109	0.517	1.00
0.8	0.647	0.0071	0.502	1.00
0.9	0.634	0.0111	0.489	1.00
0.95	0.614	0.0197	0.469	1.00
0.97	0.589	0.0266	0.445	1.00
0.99	0.608	0.0198	0.463	1.00
Max	0.667	0.0266	0.524	1.00
Min	0.589	0.0000	0.445	1.00

The figure 5.3, figure 5.4(a-b) and table 5-2 illustrate the estimation of lambda by analysing the kappa values, kappa standard deviation, user accuracy, producer accuracy and reproducibility of MRF results. The lambda value with the highest kappa, user accuracy, least standard deviation with high reproducibility was taken as optimal, λ =0.3. Lambda value of 0.3 is the most optimal since it has the least standard deviation and high user accuracy. The error in λ=0.3 is null hence the preference of the value as optimal.

5.3. HMRF Lambda segments optimization

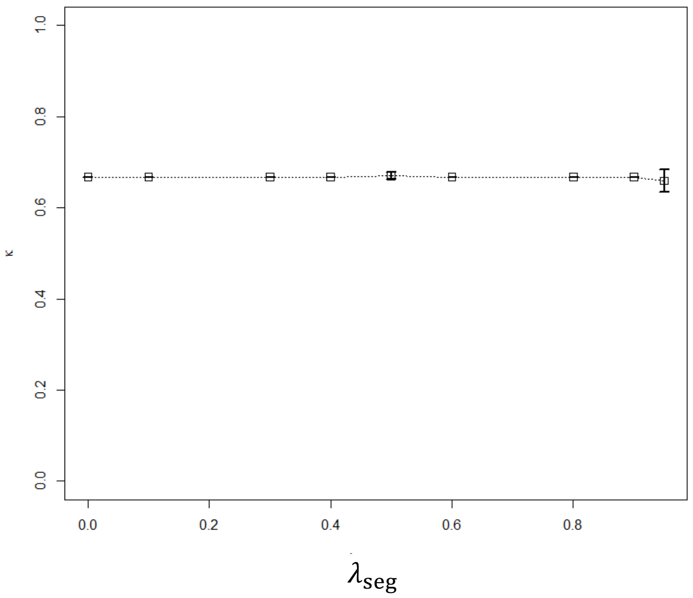
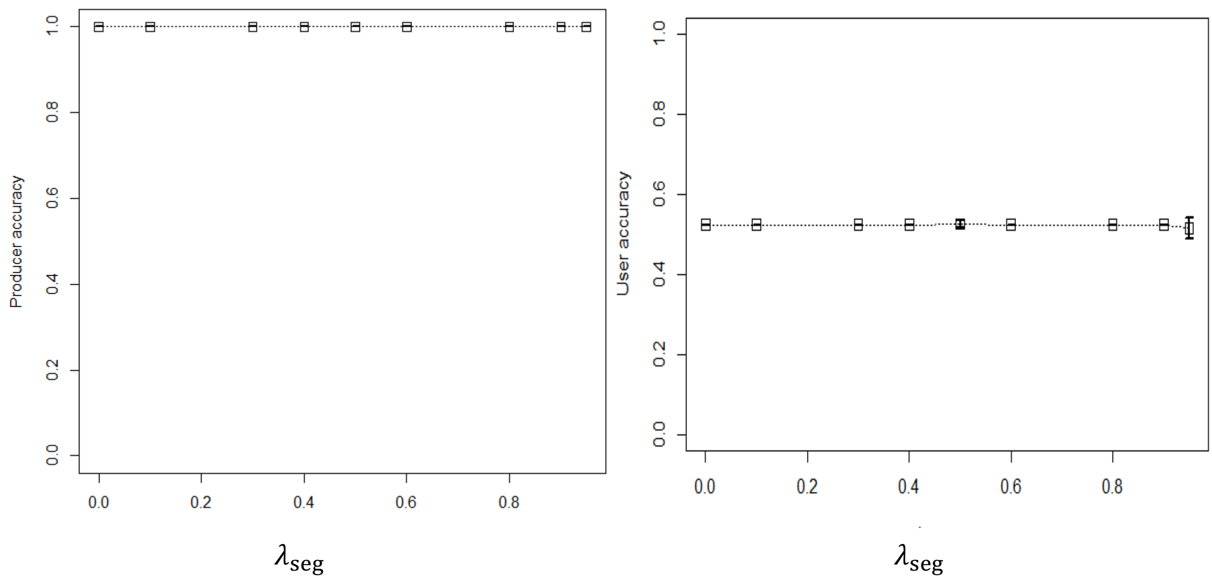


Figure 5.5: Kappa versus λ_{seg}values.



(a) (b)

Figure 5.6: (a) and (b) Producer and user accuracies versus λ_{seg}

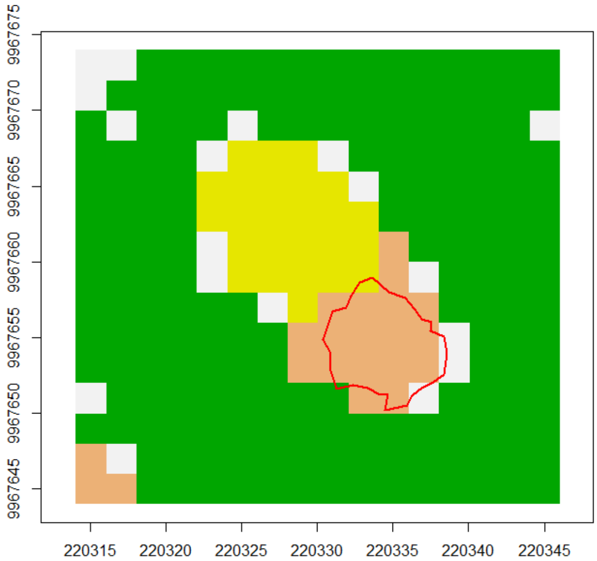
Table 5-3: Lambda segments values with respect to kappa, kappa standard deviation, User and producer accuracy

λ_{seg}	Kappa	kappa_sd	User accuracy	Producer accuracy
0.0	0.667	0.000	0.524	1.000
0.1	0.667	0.000	0.524	1.000
0.3	0.667	0.000	0.524	1.000
0.4	0.667	0.000	0.524	1.000
0.5	0.670	0.0076	0.526	1.000
0.6	0.667	0.000	0.524	1.000
0.8	0.667	0.000	0.524	1.000
0.9	0.667	0.000	0.524	1.000
0.95	0.659	0.0255	0.515	1.000
Max	0.670	0.0255	0.526	1.000
Min	0.659	0.000	0.515	1.000

The λ_{seg} parameter was optimised as shown in figure 5.5, 5.6(a-b) and table 5-3. The same criterion earlier applied in the estimation of lambda was applied to arrive at λ_{seg} =0.5 as the most optimal MRF parameter. Lambda value of 0.5 is the most optimal since it has the high user accuracy and minimal standard deviation.

5.4. Hierarchical MRF output

Hierarchical MRF output for optimal parameter values:
T0 =3.0, Tupd=0.99, Lambda = 0.95 and Lambda segments = 0.9. The figure in 5.7 shows an MRF result whose colors do represent the following: Brown=tree crown, yellow=shadow, green=grass and white=bare soil.



Legend

Class	color
Tree crown	Brown
Shadow	
Grass	Green
Bare soil	White

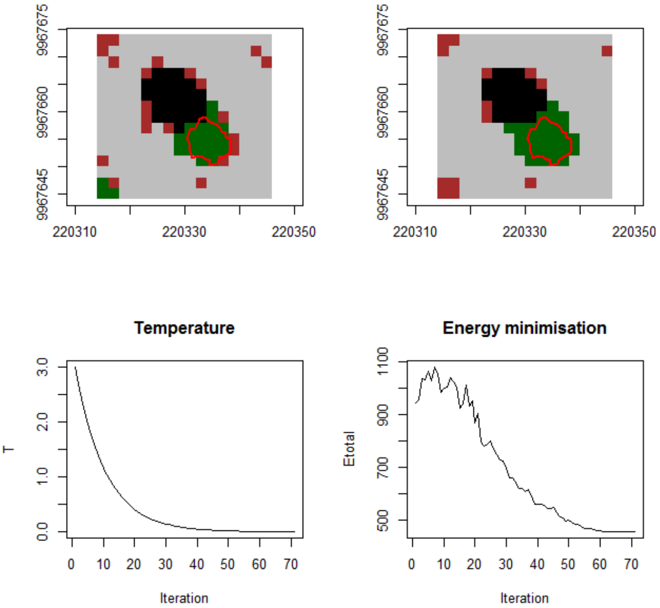
Figure 5.7: MRF result for λ, λ_{seg} set at 0.95 and 0.9 respectively, at T=3.0 and Tupd=0.99

Table 5-4: Confusion matrix for MRF output

	Tree crown	Background	Total	Error of commission	User accuracy
Tree crown	11	10	21	47.619	52.381
Background	0	219	219	0.000	100.00
Total	11	229	240		
Error of commission	0	4.367			
Producer accuracy	100	95.633			

T0=3.0, Tupd=0.9, Lambda=0.9, Lambda segments=0.95 and Lambda shape=0.0

The figures in 5.8, shows an MLC and MRF result whose color choice does represent the following: Dark brown=bare soil, black=shadow, green=tree crown and gray=grass. In the first two outputs in a row, the first figure illustrates MLC whereas the second shows MRF.



Legend

Class	Color
Tree crown	Green
Shadow	
Grass	Grey
Bare soil	Brown

Figure 5.8: Comparison of MLC versus MRF in classification of subset image

The MRF result is less noisy in comparison to the MLC; the confusion matrix for the MRF output is shown in the table 5-5.

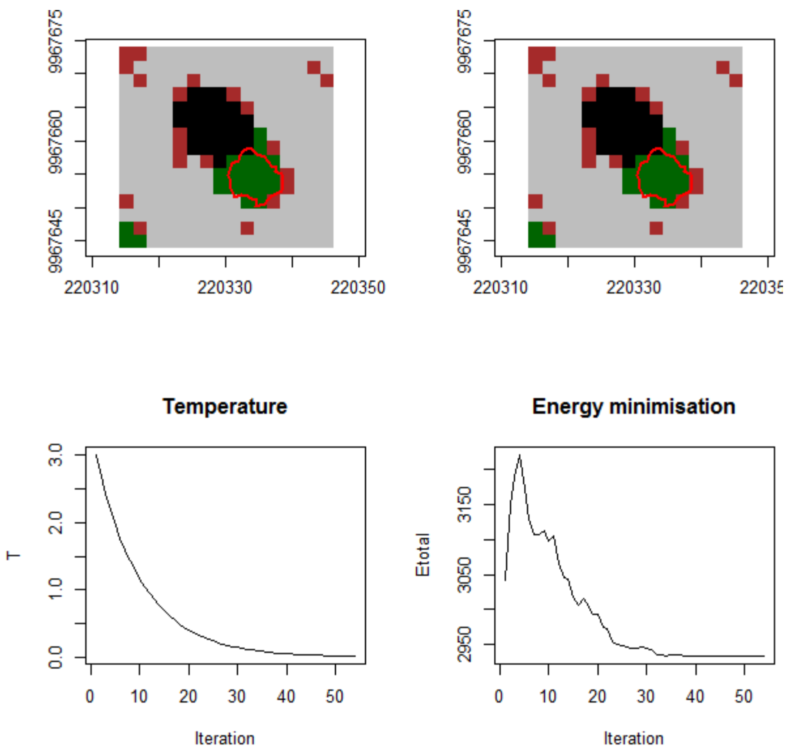
Table 5-5: Error of commission and omission in the MRF result

	Tree crown	Background	Total	Error of commission	User accuracy
Tree crown	11	16	27	59.259	40.741
Background	0	213	213	0.000	100.00
Total	11	229	240		
Error of omission	0.000	6.987			
Producer accuracy	100.00	93.013			

The figure 5-8, show experimental results of the comparison of MLC versus HMRF. The HMRF result is smooth than the MLC result but the error of commission is 59.26%. Sixteen pixels of tree crown are misclassified in HMRF output whereas ten pixels are misclassified in MLC. This is so because lambda and lambda segment parameters are assigned high values. The MRF result indicate clearly that the shadow is detected well than the shadow in MLC. This is attributed to the reason that most pure pixels were picked for shadow than tree crown.

T0=3.0, Tupd=0.9, Lambda=0.3, Lambda segments=0.5 and Lambda shape=0.1

The parameters optimal for energy minimisation (T0 and Tupd) and lambda, lambda segments optimal parameters were used in the contextual classification technique to output the following result.



Legend

Class	Color
Tree crown	Green
Shadow	
Grass	Grey
Bare soil	Brown

Kappa is 0.6674979

Figure 5.9: MLC versus MRF result for lambda, lambda segments set 0.3 and 0.5 respectively at T=3.0 and Tupd=0.9

Table 5-6: Confusion matrix for one tree crown

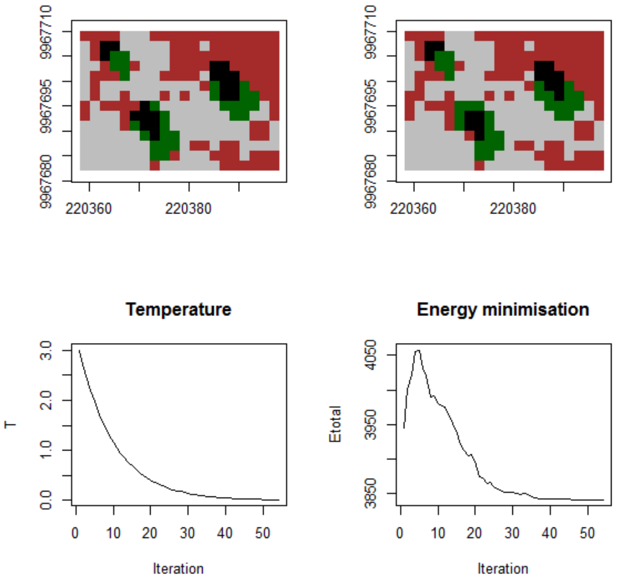
	Tree crown	Background	Total	Error of commission	User accuracy
Tree crown	11	10	21	47.619	52.381
Background	0	219	219	0.000	100.00
Total	11	229	240		
Error of omission	0.000	4.367			
Producer accuracy	100.0	95.633			

Optimal HMRF parameters on the tuning subset were applied in figure 5-9 producing a user accuracy of 52.38% as tabulated in table 5-6. Shadows in this case are identical, and tree crown was detected in both MLC and MRF.

5.5. Application areas

5.5.1. Three separated tree crowns

T=3.0,Tupd=0.9,lambda=0.3,lambda segments=0.5,lamshape=0.1



Legend

Class	Color
Tree crown	Green
Shadow	
Grass	Grey
Bare soil	Brown

Figure 5.10: HMRF method implemented on three individual tree crowns

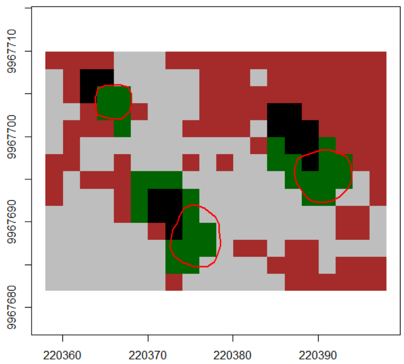


Figure 5.11: Reference individual tree crown vectors from a panchromatic image

Table 5-7: Accuracy measures in implementation of HMRF on separates tree crowns.

Kappa	User accuracy	Producer accuracy	Overall accuracy
0.6467795	0.581	0.818	93.929

In figure 5.10, it is noted that shadows are estimated well in MRF than in MLC. The user accuracy for detecting tree crown and shadows is 58.1% as tabulated in 5-7.

5.5.2. Tree crowns almost interlocked

The figure 5.12, 5.13 and table 5-8, 5-9 demonstrate that shadows are estimated acceptably unlike the tree crowns. Indeed, the tree crowns are over detected in relation to the reference data. The user accuracy is below 50%. However, the overall and producer accuracies are 88.562% and 96.80% respectively. Therefore, we draw a general observation that this method does not perform in tree crowns almost interlocked.

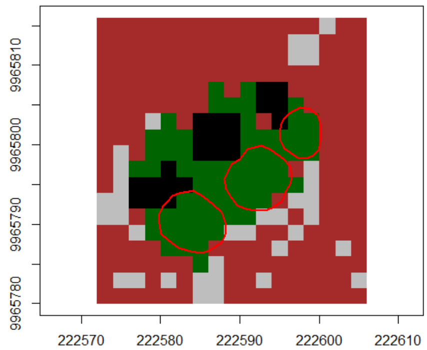
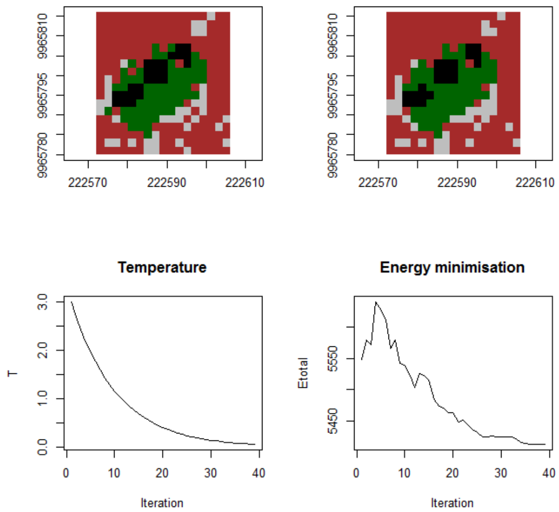


Figure 5.12: Tree crowns with the reference vector

Table 5-8: HMRF accuracies obtained at optimal parameters and lambda shape 0.5

Producer accuracy	User accuracy	Kappa	Overall accuracy
0.968	0.469	0.573	88.562

T0=3.0,Tupd=0.9,lambda=0.3,lambda segments=0.5,lamshape=0.1



Legend

Class	Color
Tree crown	Green
Shadow	
Grass	Grey
Bare soil	Brown

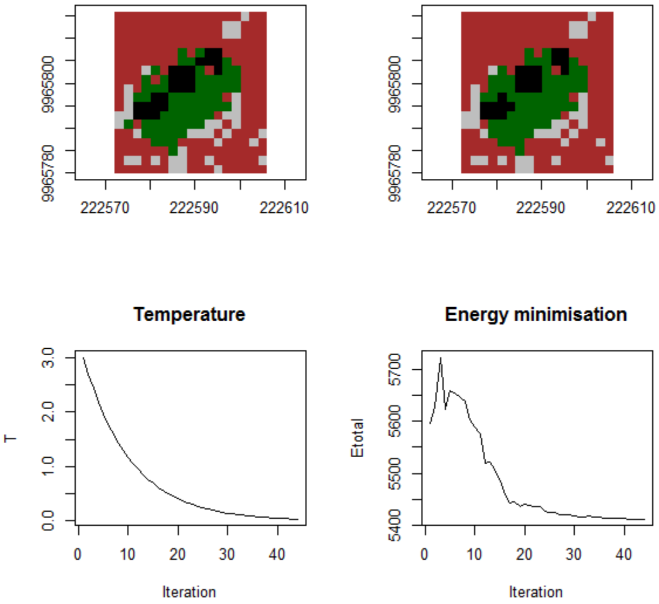
Figure 5.13: MLC versus MRF results in application of HMRF method on almost interlocked tree crowns.

Table 5-9: HMRF accuracies obtained at optimal parameters and lambda shape 0.1

Producer accuracy	User accuracy	Kappa	Overall accuracy
0.968	0.441	0.542	87.255

T0=3.0,Tupd=0.9,lambda=0.3,lambda segments=0.5,lamshape=0.8

It is observed in figure 5.14 that shadows are near perfect in the MRF output result in that they are well detected and separated in relation to their shape. The tree crown is over estimated in MLC and MRF. The overall and producer accuracy are 88.562% and 96.80% respectively. The user accuracy is 46.90%.



Class	Color
Tree crown	Green
Shadow	
Grass	Grey
Bare soil	Brown

Figure 5.14: HMRF method implemented on almost interlocked tree crowns

Table 5-10: (a) and (b) HMRF result accuracy measures and confusion matrix obtained at optimal parameters, lambda shape 0.8

Producer accuracy	User accuracy	Kappa	Overall accuracy
0.968	0.469	0.573	88.562

(a)

	Tree crown	Background	Total	Error of commission	User accuracy
Tree crown	30	34	64	53.125	46.875
Background	1	241	242	0.415	99.587
Total	31	275	306		
Error of omission	3.226	12.364			
Producer accuracy	96.774	87.636			

(b)

6. Discussion

In this research topic, the methodology employed was a two tier approach:
1) Level one is MRF pixel based classification
2) Level two is MRF object based classification

This two level contextual classification approach is the HMRF technique that is developed to try and improve classification of individual tree crowns in a high resolution multispectral image. The HMRF results obtained show that the developed method can improve classification on separated tree crowns. The user accuracy obtained for three separated tree crowns is 58.10% as shown in table 5-7 and a user accuracy of 52.38% for single tree crown as indicated in table 5-6. These results are well over the average, so acceptable, but they are not better than the MLC in the tree crown class.

In reference to figure 5.10, 5.13 and 5.14, the method estimates pixels of the shadow class outperforming the MLC result. This is attributed to the ease with which pure shadow pixels could be identified and picked during sampling. Therefore, it can be deduced that if pure tree crown pixels are sampled, the method can improve in the classification accuracy of trees.

On the other hand, the method does not perform well in interlocked tree crowns because: a) the shape of interlocked tree crowns is different. b) The sampling strategy did not base on field measurements hence mixed pixels could have been picked as training set data.

6.1. Strengths of the HMRF method

The method introduces the inclusion of object characteristics in classification. Cognitive characteristics of objects are modeled into the classification process with MRF energy functions. The method developed in this study, for instance, integrates shape and size in contextual classification. This does have classification accuracies above 50.00% in tree crown estimation as described in figure 5-11 and table 5-7. Besides, the method opens up the opportunity of integrating spatial object characteristics with MRF in classification to boost classification accuracies of objects.

Shape of tree crown is a geometric property that can aid in the detection of trees. However, this can improve classification results only if an authentic sampling strategy that makes a direct field contact is applied. Hence, shape and size as object traits are of paramount addition to the contextual classification of objects from images.

6.2. Opportunities of the HMRF method

This method can be improved further by including spatial information of panchromatic image in order to improve the tree crown detection. Ground data and information need to be acquired in order to verify pixels when performing sampling on a multispectral image.

6.3. Weaknesses of the HMRF method

The method has a weakness in the sense that only spectral information of classes has been considered and class pixels for training set data were picked from the multispectral image without ground verification to determine whether or not they are pure. The overestimation of tree crown pixels in figure 5-12, 5.13, and 5.14 is indicative of this weakness.

7. Conclusions and recommendations

7.1. Conclusions

1) Is shape and size of tree crown an important spatial characteristic in classification?

Shape and area: Tree crowns are round in shape with a definite area; this geometry makes it easier to discriminate the extents of a tree in a multispectral image when modelled with MRF. The results of MRF based contextual classification method, which integrates shape and size in the MRF energy functions, outperforms classification results obtained with MLC. Prior information therefore, does lead to an improvement of tree crown detection.

2) What quantifiable accuracy measures are established by the improved algorithm in comparison with maximum likelihood classification (MLC)?

   The HMRF based contextual classification method yielded a user accuracy of 58.06%, producer accuracy of 81.82%, kappa of 0.65 and 93.93% overall accuracy. These classification results are acceptable given that the sampling strategy of the training set can be improved by performing field observations.

3) What is the efficiency of this classification algorithm in relation to various determining factors: computational time and contextual classification accuracy?

   The HMRF method detects tree crowns that are separated in improved accuracy. The trees that form almost interlocked tree crowns obtain low accuracy less than 50% because the shape of these crowns is not round.
   The objective of this research, extending a classification technique by including prior knowledge on objects, has been achieved. The contextual classification technique so developed improves classification, however, issues raised in the recommendation section can make the technique improve classification results more.

7.2. Recommendations

Data aspect

1) Reference vector data can be obtained from high resolution aerial photographs of the study area, these are more accurate than the panchromatic image used in this study. Acquisition date of the aerial photographs should be when we have the good weather conditions for photography.
2) Panchromatic image with high spatial resolution can be integrated to improve classification accuracy. The mean and standard deviation of the classes from the PAN image can be incorporated in methods.

Method aspect

3) Sampling strategy of the training set data need improvement by performing a fieldwork (ground truth). Ground control points showing relevant class pixel labels can be used as training data to aid in the training process.
4) This study is focused on individual tree crown; a further method can be developed for interlocked tree crowns that exhibit other geometric shapes.

Application aspect

Post classification analysis is required to discriminate shrubs, large flowers and other vegetation types. This should then be classified in a different class.

LIST OF REFERENCES

Arachchige, W. (2011). *Accuracy assessment of fuzzy classification.* University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.

Ardila, J. P., Tolpekin, V.A., Bijker, W., Stein, A., (2011). Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(6), 762-775.

Da Fontoura Costa, L., & Cesar, R. M. (2001). *Shape analysis and classification: theory and practice*: CRC.

Derin, H., & Elliott, H. (1987). Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-9*(1), 39-55.

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems, 50*(1), 1-18.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis* (Vol. 4): John Wiley & Sons New York.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-6*(6), 721-741.

Hassner, M., & Sklansky, J. (1980). The use of Markov Random Fields as models of texture. *Computer Graphics and Image Processing, 12*(4), 357-370.

ITC Core book. (2010). GI Science and Earth Observation, a process - based approach.ITC,Enschede, the Netherlands.

Kamal, M., & Phinn, S. (2011). Hyperspectral Data for Mangrove Species Mapping: A Comparison of Pixel-Based and Object-Based Approach. *Remote Sensing, 3*(10), 2222-2242.

Khedama A., B.-A. (2004). *Contextual classification of remotely sensed data using map approach and MRF.*

Li, S. Z. (2009). *Markov Random Field Modeling in Image Analysis* (Third Edition ed.): springer.com.

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

Moskal, L. M., Styers, D. M., & Halabisky, M. (2011). Monitoring Urban Tree Cover Using Object-Based Image Analysis and Public Domain Remotely Sensed Data. *Remote Sensing, 3*(10), 2243-2262.

Phillips, D. B., & Smith, A. F. M. (1994). bayesian faces via hierarchical template modeling.*Journal of the American Statistical Association, 89*(428), 1151-1163.

Serpico, S. B., & Moser, G. (2006). Weight Parameter Optimization by the Hondash;Kashyap Algorithm in MRF Models for Supervised Image Classification. *Geoscience and Remote Sensing, IEEE Transactions on, 44*(12), 3695-3705.

Serpico, S. B., & Moser, G. (2006). Weight Parameter Optimization by the Ho–Kashyap Algorithm in MRF Models for Supervised Image Classification. *Geoscience and Remote Sensing, IEEE Transactions on, 44*(12), 3695-3705.

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

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

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

URL1: E-cognition
http://www.pcigeomatics.com/products/pdfs/definiens/ReferenceBook.pdf

URL2: Imagine Objective
http://www.uwf.edu/gis/manuals/IMAGINE_Objective.pdf

URL3: ITC geodata warehouse
http://intranet.itc.nl/support/it/pop_links/gts/earth%20observation/geodatawarehouse.aspx
(25th october, 2011)

URL4: WorldView2 satellite information
http://worldview2.digitalglobe.com/docs/WorldView2_8Band_Applications_Whitepaper.pdf
(26th october, 2011)

APPENDIX A

```
Ulsegshape <- function(nr,Seg)
{
	# Shape contribution

	val <- 0.5*((seg_shape(nr,Seg)-1)^2)

	# Area contribution
	Area <- unlist(Seg[2])

	lA <- log(Area[nr]/pixArea)

	# compare area to PDF
	val <- lamshape*val + (1-lamshape)*(((lA - meanArea)^2)/(2*stdArea^2) + lA)

	return(val)
}


Ush_seg<-function(i,j,Sc)
{
	Seg <- Segment(F,Sc)	# only class Sc is considered in shape analysis

	Nobj <- unlist(Seg[1])

	temp <- 0

	if(Nobj>0)
	{
		Nxext <- c(i,Nx[i,j,])
		Nyext <- c(j,Ny[i,j,])

		if(sum(Nxext==0)>0)
		{
			l1 <- which(Nxext==0)
			Nxext <- Nxext[-l1]
			Nyext <- Nyext[-l1]
		}

		coords <- cbind(Nxext,Nyext)
		Fvec <- F[coords]

		N1 <- sum(Fvec==Sc)
		N2 <- sum(Fvec!=Sc)

		if(N1>0)
		{
			if(N2>0)
			{
				l1 <- which(Fvec!=Sc)
				coords<-coords[-l1,]
			}

			Snrs <- array(0,N1)
```

```
                    if(N1==1)
                    {
                            Snrs <- segm_number(coords,Seg)
                            temp <- temp + Ulsegshape(Snrs,Seg)

                    }else
                    {
                            for(k1 in 1:N1) Snrs[k1] <- segm_number(coords[k1,],Seg)

                            # delete repeating numbers in Snrs
                            Snrs <- compress_arr(Snrs)
                            N1 <- length(Snrs)
                            if(N1==1)
                            {
                                    temp <- temp + Ulsegshape(Snrs,Seg)
                            } else for(k in 1:N1)temp <- temp + Ulsegshape(Snrs[k],Seg)
                    }
            }
        }

        return(temp)
}

U <- function(i,j)
{
   val <- (1-lamseg)*(lambda * Uprior(i,j)+ (1-lambda)*Ulpix(i,j)) + lamseg * Ush_seg(i,j,Sc)
        return(val)
}

TotalEnergy<-function()
{
        val <- 0

        for(i in 1:M)
        for(j in 1:N)
        {
          val <- val + lambda * Uprior(i,j) + (1-lambda) * Ulpix(i,j)
        }

        Seg <- Segment(F,Sc)     # only class Sc is considered in shape analysis

        Nobj <- unlist(Seg[1])
        Area <- unlist(Seg[2])

        temp <- 0
        for(k in 1:Nobj)
        {
                    temp <- temp + Ulsegshape(k,Seg)
        }

        val <- val + lamseg * temp


        return(val)
}
```