

**School of Environment, Faculty of Science and Engineering
Flinders University, South Australia**

**Pixel- and Feature-level Fusions of Aerial Imagery with LiDAR
Data for Landscape Object Extraction**

Syed Sohel Ali

**This thesis is presented as part of the requirements for the Degree of
Doctor of Philosophy
at the Flinders University**

May 2011

Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university.

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

Signature:..... Syed Sohul Aji

Date:..... 9 March 2012

ABSTRACT

The objective of this research is to investigate different fusion models that integrate aerial imagery with LiDAR data for landscape object extraction. Pixel- and feature-level fusions are particularly investigated in data- and user-driven scenarios to delineate a range of landscape objects on forest and semi-urban study areas. Thematic accuracy is evaluated against field-surveyed data and optimum fusion models for each study area is identified.

The complementary nature of aerial imagery and LiDAR data is the main reason for their selection in this research. LiDAR data provides an accurate measurement of landscape structure in the vertical plane; however, LiDAR sensors have limited coverage in the electromagnetic spectrum. By contrast, aerial imagery provides extensive coverage of landscape classes in the electromagnetic spectrum but is relatively insensitive to variations in height of objects. As a result, the fusion of aerial imagery with LiDAR data has the potential to significantly improve mapping of the landscape. Since small footprint LiDAR and aerial imagery systems can achieve very high spatial, spectral and textual resolutions and suitable for site-specific landscape mapping. However, the direct relationship between spatial resolution and landscape classification does need to be considered before apply any fusion model.

The forest study area is in the Moira State Forest, which form a part of Barmah and Millewa forests near Mathoura, New South Wales on a site 1.25km x 1km consisting of native Eucalyptus forest. The semi-urban study area 1.25km x 1km is also in Mathoura township itself and contains typical semi-urban landscape objects such as residential and commercial buildings, open spaces, roads and gardens.

Geometric corrections of multi-source data are a prerequisite for any data fusion study. Variations in sensor altitude, attitude, and calibration affect the quality of the fusion results. A parametric rectification model was implemented to correct these influences in high spatial resolution aerial imagery. The aerial triangulation-derived RMSE values for colour and multispectral imagery are between 0.70 and 1.17 microns. These results meet the established standard 1-micron, or one-quarter of a

pixel, benchmark and provide a sound geometric basis for multi-source data fusion. The original 0.88m spatial-resolution aerial multispectral imagery is resampled into 0.5m resolution for uniformity with LiDAR data. The processed LiDAR data do not require orthorectification, however an nDSM needs to be generated from the LiDAR source for use in the fusion process. The nDSM represents mean height of landscape objects and is computed as the difference between first and last LiDAR returns. LiDAR-derived object heights are compared with the field-surveyed data to check height accuracy. The estimates of heights from LiDAR were generally within one metre of field measurements, although discrepancies as high as 3m were observed.

An extensive field survey was conducted to collect training data and gather reference data for thematic accuracy assessment. A complex sampling strategy was developed combining random and systematic sampling techniques that provide a good balance between statistical validity and practical application. For thematic accuracy assessment, error matrices were generated using reference pixel data derived from field-survey and aerial photo interpretation with corresponding pixels of the fusion results. Overall thematic accuracy as well as User's and Producer's accuracies were computed to measure the success of fusion models. Kappa analysis tested the significance of each matrix and determined whether the results presented in the error matrix were significantly better than a random result.

For the forest study area, data-driven pixel-level fusion was implemented using an unsupervised model. In feature-level fusion, a watershed transformation algorithm was utilised for delineating tree crowns; a masking techniques was used for collecting tree feature attributes; and finally an unsupervised model was applied for delineating tree species. Thematic accuracies of pixel- and feature-level fusion results are assessed through error matrices derived from fusion results and field-surveyed reference data. In pixel-level fusion results, the overall thematic accuracy was 64.67 percent and the Kappa Coefficient value was 0.52. The Kappa Coefficient of the pixel-level fusion results indicates *moderate* agreement between the fusion results and the reference data. In feature-level fusion results, the overall thematic accuracy was 86.33 percent and the Kappa Coefficient was 0.82, which being close to 1, indicated *substantial* agreement between fusion results and reference measurement. Statistical comparison of the pixel- and feature-level fusion results

indicated that, at the 95 percent confidence level, the standard normal deviation of the Kappa Coefficient was 6.68, the results were significantly better than a random result and feature-level fusion achieves better results than the pixel-level fusion. Segmentation and subsequent feature classification using a data-driven model provided superior results to pixel-level fusion.

In a user-driven scenario, a hierarchical landscape classification scheme was developed for the delineation of semi-urban landscape objects using pixel- and feature-level fusions. 4-band multispectral imagery and LiDAR-derived nDSM have incompatible statistics and unable to represent into a normal class model as a result statistical methods of supervised fusion is not considered. Pixel-level fusion utilises the supervised parallelepiped technique to fuse these datasets. In the feature-level fusion, the feature delineation is achieved through multi-resolution segmentation and subsequently classifies features using knowledge-driven rules. The spectral, spatial and contextual properties of the features are utilised to develop these knowledge rules.

For the semi-urban study area, the thematic accuracy of pixel-level fusion results was 73.25 percent with Kappa Coefficient 0.67. There is thus *substantial* agreement between pixel-level fusion results and reference data. By contrast, feature-level fusion of the same datasets gave 88.38 percent overall accuracies and Kappa Coefficient 0.86 indicating *excellent* agreement between fusion results and reference data. Comparing fusion results indicated a significant difference between the pixel- and feature-level fusions results. At the 95 percent confidence level, the standard normal deviation of the Kappa Coefficient for pixel- and feature-level fusions using multispectral imagery with LiDAR data was 3.70, well above the standard 1.96 threshold. This indicates a significantly better than random result and shows that feature-level fusion performs better than the pixel-level fusion. Particularly in the delineation of *Shadow* classes, the User's and Producer's Accuracies for feature-level fusion results were substantially better than pixel-level fusion results.

With high spatial resolution data the interclass variability within the class was also high meaning that the 'pepper and salt' effect was widespread in pixel-level fusion results for both study areas. In particular, the pixel-level fusion in the forest study

area did not clearly delineate individual trees when they were clumped. It also suffered from mixed-pixel effect for individual trees as it showed multiple tree species within a single canopy. Feature-level fusion overcame this problem by defining the tree canopy areas first, then extracting feature attributes from the segments, and finally identifying the tree species using unsupervised feature classification.

Replacement of multispectral imagery with colour imagery revealed results were not much different in feature-level fusion but for pixel-level fusion of multispectral imagery lead to less misclassification for the higher radiometric depth (16-bit) than did colour imagery (8-bit). The exclusion of the LiDAR data greatly reduced the quality of the fusion results. The results of the fusion accuracy for inclusion and exclusion of LiDAR are small in pixel-level fusion but are large in feature-level fusion. The same class of objects has conflicting spectral properties but inclusion of LiDAR-derived height resolves that conflicts and plays a vital role in landscape mapping. The research has lead to the conclusion that feature-level fusions perform better in classifying landscape objects than pixel-level fusions in both data- and user-driven scenarios.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude and appreciation to all the people who have facilitated the completion of this dissertation. In particular, I sincerely thank Dr. Paul Dare for his continued support and valuable guidance throughout this research. This thesis is an outcome of his constructive comments, unsparing criticism and painstaking editing. Without his help, this work would not be possible. I would also like to thank Professor Simon Jones for his constructive comments and for giving advice and guidance throughout this research.

I would like to express my gratitude to Flinders University and Australian Research Council (ARC) for academic and financial support. I also thank Australian Department of Defence for continuous support. In particular, I thank Gary Palmer for his encouragement to complete this research.

I would like to record my appreciation to a number of organisations that provided most valuable support through sharing data. Without them I could not have completed the thesis. Especially, I am thankful to IFMS Germany for the provision of Ultracam-D airborne data. I would also like to extend my thanks to AEROMETREX for the provision of colour aerial photography, Spatial Scientific Technologies for the small format aerial photography and Airborne Research Australia at Flinders University for administrative support. I also thank Kim Wells for his contribution to finalising this manuscript.

I want to thank my parents for their wholehearted support. I also thank Abu and Rawshon Abdullah for their good wishes and endless support to my study.

Finally, to my dear wife Jesmin, son Shayaan and daughter Abia, I express my deepest gratitude for your understanding during the preparation of this thesis. Without your constant support it would not have been possible.

TABLE OF CONTENTS

ABSTRACT.....	III
ACKNOWLEDGEMENTS	VII
TABLE OF CONTENTS	VIII
LIST OF ACRONYMS	XIII
LIST OF FIGURES	XIV
LIST OF TABLES	XVII
CHAPTER 1.....	1
1 INTRODUCTION.....	1
1.1 DATA FUSION IN REMOTE SENSING	1
1.1.1 Pixel-level fusion.....	3
1.1.2 Feature-level fusion.....	3
1.1.3 Decision-level fusion.....	4
1.2 FUSION OF AERIAL IMAGERY WITH LiDAR DATA	4
1.3 PROBLEM STATEMENT	5
1.4 RESEARCH OBJECTIVES	7
1.5 SIGNIFICANCE AND BENEFITS OF THE RESEARCH.....	8
1.6 RESEARCH METHODOLOGY	9
1.7 THESIS STRUCTURE.....	10
CHAPTER 2.....	13
2 A REVIEW OF REMOTELY SENSED DATA FUSION PROCESSES ...	13
2.1 INTRODUCTION	13
2.2 REVIEW OF SCALE EFFECT IN IMAGE FUSION.....	13
2.3 REVIEW OF HIERARCHICAL LANDSCAPE CLASSIFICATION SCHEMES	16
2.3.1 Urban landscape classification scheme	19
2.3.2 Forest landscape classification scheme	21
2.4 REVIEW OF AERIAL IMAGERY AND LiDAR DATA FUSION.....	22
2.5 REVIEW OF DATA FUSION TECHNIQUES	25
2.6 PIXEL-LEVEL FUSION	25
2.6.1 Colour related fusion models	25
RGB coding based fusion.....	25
RGB to IHS based fusion	26
2.6.2 Numerical fusion models.....	27
Arithmetic combinations	27
Adding and multiplication.....	27
Difference and ratio	28
2.6.3 Statistical fusion models.....	29
Unmixing based fusion.....	29
Principal Component Analysis (PCA)	29
Markov Random Field (MRF)	31
Linear regression	31
2.6.4 Signal-processing-based Fusion Models	32
High-pass filter	32
Wavelets	32
Pyramid algorithms	34
2.6.5 Classification-based fusion models	34
Unsupervised ISODATA approach.....	36

Supervised approach	38
2.7 FEATURE-LEVEL FUSION	43
2.7.1 Segmentation.....	45
Data-driven segmentation	45
User-driven segmentation	55
2.8 CLASSIFY THE SEGMENTED FEATURES	59
2.8.1 Fuzzy classification of segmented features.....	60
2.9 EVALUATION OF FUSED RESULTS	63
2.9.1 Accuracy assessment of fusion results.....	64
Descriptive techniques	65
Analytical techniques.....	67
2.9.2 Sampling design for accuracy assessment	70
Sampling scheme	70
Sample size	72
2.10 COMPARISONS OF PIXEL- AND FEATURE-LEVEL FUSION	73
2.10.1 Qualitative comparisons	73
2.10.2 Quantitative comparisons	73
2.11 CONCERNS ABOUT PER-PIXEL BASED THEMATIC ACCURACY ASSESSMENT	78
2.12 SUMMARY OF REVIEWS	81
2.13 CHAPTER CONCLUSIONS.....	86
CHAPTER 3	88
3 STUDY AREA AND DATASETS.....	88
3.1 INTRODUCTION	88
3.2 SELECTION OF THE STUDY AREA AND FUSION MODEL.....	88
3.3 GEOGRAPHY OF THE STUDY REGION	91
3.4 MAJOR LANDFORMS OF THE STUDY REGION	91
3.5 CLIMATE OF THE STUDY REGION	92
3.6 CHARACTERISATION OF THE LANDSCAPE TYPES	93
3.7 THE FOREST STUDY AREA.....	94
3.7.1 River red gum (RRG), <i>Eucalyptus camaldulensis</i>	95
3.7.2 Black box (BB), <i>Eucalyptus largiflorens</i>	97
3.7.3 Grey box (GB), <i>Eucalyptus microcarpa</i>	97
3.8 THE SEMI-URBAN STUDY AREA	98
3.8.1 Open space/road.....	98
3.8.2 Roofs	98
3.8.3 Tree/grass	99
3.8.4 Shadows	100
3.9 DATA SELECTION FOR THE FUSION MODELS	100
3.10 ACQUISITION AND PROPERTIES OF LiDAR DATA.....	101
3.10.1 Principles of LiDAR remote sensing.....	101
3.10.2 LiDAR range measurement.....	103
3.10.3 The properties of a LiDAR system.....	104
3.11 ACQUISITION AND PROPERTIES OF COLOUR IMAGERY	105
3.12 ACQUISITION AND PROPERTIES OF MULTISPECTRAL IMAGERY	105
3.13 OTHER DATA SOURCES	106
3.13.1 Supporting digital aerial photography collections.....	106
3.13.2 Field data collection	107
3.14 SUMMARY	109

CHAPTER 4	110
4 GEOMETRIC CORRECTION OF THE MULTI-SOURCE DATA	110
4.1 INTRODUCTION	110
4.2 REVIEW OF THE IMAGE RECTIFICATION METHODS	111
4.2.1 Non-parametric rectification methods.....	112
4.2.2 Parametric rectification	112
4.3 REVIEW OF IMAGE RESAMPLING METHODS	117
4.3.1 Resampling methods	117
4.3.2 Resampling interval.....	118
4.4 SELECTION OF THE IMAGE RECTIFICATION METHOD	119
4.5 SELECTION OF THE IMAGE RESAMPLING METHOD.....	120
4.6 ORTHORECTIFICATION OF THE AERIAL IMAGERY	121
4.6.1 Extraction of interior and exterior orientation parameters	122
4.6.2 Selection of GCPs and checkpoints.....	123
GCP requirements	123
Collection of GCPs for forest area	124
Collection of GCPs for semi-urban area	125
4.6.3 Evaluating GCPs location precision.....	126
4.6.4 Performing aerial triangulation	128
4.6.5 Performance of the resampling.....	130
4.7 LiDAR DATA PROCESSING	131
4.7.1 Normalised DSM Generation from LiDAR Data.....	131
4.7.2 Normalised DSM accuracy analysis.....	133
4.8 DISCUSSION	135
4.9 SUMMARY.....	137
 CHAPTER 5	 139
5 IMPLEMENTATION OF DATA-DRIVEN FUSION MODELS FOR TREE ATTRIBUTE DELINEATION*	139
5.1 INTRODUCTION	139
5.2 TREE PROPERTY ANALYSIS.....	140
5.3 METHODOLOGY FOR TREE DELINEATION	141
5.4 DATA PROCESSING	142
5.5 PIXEL-LEVEL FUSION MODEL FOR TREE SPECIES DELINEATION	143
5.6 FEATURE-LEVEL FUSION MODEL FOR TREE SPECIES DELINEATION.....	146
5.6.1 Watershed segmentation for tree feature delineation	147
Definition of internal markers	149
Definition of external markers	149
Application of watershed transformation	150
5.6.2 Attributes delineation for tree features	151
5.6.3 Filling tree features with attributes.....	151
5.6.4 Unsupervised classification of tree features	151
5.7 SUMMARY.....	152
 CHAPTER 6	 154
6 IMPLEMENTATION OF USER-DRIVEN FUSION MODELS FOR SEMI-URBAN LANDSCAPE MAPPING*	154
6.1 INTRODUCTION	154
6.2 PROPERTIES OF THE SEMI-URBAN LANDSCAPE OBJECTS.....	155

6.3	METHODOLOGY FOR SEMI-URBAN MAPPING.....	158
6.4	PIXEL-LEVEL FUSION FOR THE SEMI-URBAN STUDY AREA.....	158
	6.4.1 Selection of the training landscape objects.....	159
	6.4.2 Signature extraction of semi-urban landscape objects.....	161
	6.4.3 Fusion using parallelepiped classifier.....	162
6.5	FEATURE-LEVEL FUSION FOR THE SEMI-URBAN STUDY AREA.....	163
	6.5.1 Multi-resolution segmentation.....	164
	6.5.2 Class hierarchy.....	166
	Fusion based on spectral properties.....	168
	Fusion based on LiDAR-derived nDSM properties.....	169
	Fusion using contextual information.....	170
6.6	SUMMARY.....	171
CHAPTER 7		172
7	RESULTS AND DISCUSSION.....	172
7.1	DATA-DRIVEN FUSIONS FOR THE FOREST STUDY AREA.....	172
	7.1.1 Pixel-level fusion results.....	173
	7.1.2 Thematic accuracy of pixel-level fusion results.....	174
	7.1.3 Feature-level fusion results.....	175
	7.1.4 Thematic accuracy of feature-level fusion results.....	177
	7.1.5 Comparative accuracy of data-driven fusion results.....	178
7.2	USER-DRIVEN FUSIONS FOR THE SEMI-URBAN STUDY AREA.....	180
	7.2.1 Pixel-level fusion results.....	181
	7.2.2 Thematic accuracy of pixel-level fusion results.....	183
	7.2.3 Feature-level fusion results.....	184
	7.2.4 Thematic accuracy of feature-level fusion results.....	185
	7.2.5 Comparative accuracy of results for user-driven fusions.....	187
7.3	DISCUSSION.....	189
	7.3.1 Methodologies for the fusion of aerial imagery with LiDAR data.....	189
	7.3.2 The use of colour imagery as an alternative multispectral imagery in data fusions.....	192
	7.3.3 The effect of height data in fusion models.....	194
	7.3.4 Landscape object height derived from different DSM sources.....	196
	DSM for forest study areas.....	197
	DSM for urban study areas.....	199
	7.3.5 Issues with comparative accuracy assessment using error matrices 200	
7.4	SUMMARY.....	202
CHAPTER 8		204
8	CONCLUSIONS AND RECOMMENDATIONS.....	204
8.1	CONCLUSIONS.....	204
	8.1.1 Tree species identification in a forest area.....	205
	8.1.2 Identifying objects in a semi-urban area.....	206
8.2	RECOMMENDATIONS FOR FUTURE RESEARCH.....	207
	8.2.1 New data sources.....	207
	8.2.2 Automatic registration of multi-source data.....	208
	8.2.3 Virtual field data collection.....	208
	8.2.4 Automatic fusion models.....	209

8.2.5 Machine-generated accuracy assessment	209
REFERENCES.....	211
APPENDIX.....	235
APPENDIX 1. THE FOREST STUDY AREA FIELD DATA COLLECTION FORM.....	235
APPENDIX 2. THE SEMI-URBAN STUDY AREA FIELD DATA COLLECTION FORM.....	236
APPENDIX 3. SUMMARY STATISTICS OF THE INITIAL 10 CLUSTERS OF DATA-DRIVEN PIXEL-LEVEL FUSION.....	237
APPENDIX 4. 3-BAND COLOUR IMAGERY AND LIDAR DATA FUSION 238	
APPENDIX 5. ACCURACY ASSESSMENT OF PIXEL- AND FEATURE- LEVEL CLASSIFICATION USING COLOUR IMAGERY ONLY	242

LIST OF ACRONYMS

3D	Three Dimension
AHD	Australia Height Datum
BD	Bhattacharya Distance
CW	Continuous Wave
DSE	Department of Sustainability and Environment (Victoria)
DGPS	Differential Global Positioning System
DN	Digital Number
DSM	Digital Surface Model
ERS	European Remote sensing Satellite
ETM+	Enhanced Thematic Mapper plus
FNEA	Fractal Net Evolution Approach
GCP	Ground Control Point
GDA94	Geocentric Datum of Australia 1994
GIS	Geographic Information System
GPS	Global Positioning System
GEOBIA	Geospatial Object Based Image Analysis
GSD	Ground Sampling Distance
IFOV	Instantaneous Field Of View
IMU	Inertial Measurement Unit
INS	Internal Navigation System
ISI	Institute of Scientific Information
ISODATA	Iterative Self-Organizing Data Analysis Technique
LRF	Laser Range Finder
LiDAR	Light Detection And Ranging
MLC	Maximum Likelihood Classification
nDSM	normalised Digital Surface Model)
NDVI	Normalised Difference Vegetation Index
OBIA	Object Based Image Analysis
PDoP	Position Dilution of Precision
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SD	Standard Deviation
UTM	Universal Transverse Mercator
USGS	United State Geological Survey

LIST OF FIGURES

	Page
Figure 1.1 Processing levels of data fusion, after Pohl and van-Genderen (1998).....	3
Figure 2.1 Influence of spatial resolution on measurement of selected landscape parameters (<i>after</i> Benson and Mackenzie, 1995)	15
Figure 2.2 Spatial resolutions as a function of the land cover classification level, after Welch (1982)	19
Figure 2.3 Illustration of the gradient operators and their components	46
Figure 2.4 Flooding of the relief and dam building (a), catchment basins and divide lines (b), after Beucher (1991).	54
Figure 2.5 Example for three fuzzy sets on feature x . The membership functions on feature x define the fuzzy set <i>low</i> , <i>medium</i> and <i>high</i> for this feature, after Benz <i>et al.</i> , (2004).....	61
Figure 2.6 A conceptual model of error matrix.	65
Figure 3.1 Location of the study areas	89
Figure 3.2 Rainfall statistics in the study region, Source: BOM (2006).....	93
Figure 3.3 Characterisation of the dominant trees in forest study area	96
Figure 3.4 Dominant landscapes in the semi-urban area	99
Figure 3.5 A typical laser scanner system producing three dimensional data (XYZ coordinates), based on Wehr and Lohr (1999).....	102
Figure 3.6 Conceptual basis of LiDAR range measurement.	103
Figure 3.7 Sample locations for the forest site	108
Figure 4.1 Elements of exterior orientation, after Mikhail <i>et al.</i> , (2001)	114
Figure 4.2 Flowchart showing the orthorectification of the aerial imagery	122
Figure 4.3 GCPs location for the forest area orthorectified images	125
Figure 4.4 Distribution of GCPs for orthorectification of the semi-urban area images 125	125
Figure 4.5 Positional precisions compared with the PDoP for the GCPs in the forest study area	126
Figure 4.6 Positional precisions compared with the PDoP for the GCPs in the semi-urban study area	127

Figure 4.7 Mean tree height derived by subtracting the last return from first return of the LiDAR data	132
Figure 4.8 Mean semi-urban objects height derived by subtracting the last return from first return of the LiDAR data	133
Figure 4.9 Relationship between individual tree mean heights as estimated in the field and from LiDAR data.....	134
Figure 4.10 Regression analyses of the field and mean LiDAR height data for the semi-urban study area.....	135
Figure 5.1 Tree species spectral profile of the forest site.....	141
Figure 5.2 A conceptual model for comparing different data-driven fusion levels for tree species delineation.....	142
Figure 5.3 (a) An RGB composite image of the forest study area (R: IR, G: nDSM and B: Texture); (b) Red band versus IR band scatter diagram with the tree species cluster centre	144
Figure 5.4 Feature-level fusion of multispectral imagery with LiDAR data for tree species delineation.....	146
Figure 5.5 An illustration of watershed segmentation. (a) A canopy model derived from nDSM, (b) 3D view of the canopies, and (c) Segmentation results with dams built at the divide line	147
Figure 5.6 Flow chart of marker-controlled watershed transformation for tree feature delineation	148
Figure 5.7 Tree crown map in the forest study area.....	150
Figure 6.1 Spectral profiles of semi-urban landscape objects.....	156
Figure 6.2 A conceptual model for comparing different user-driven fusion levels for semi-urban landscape mapping	158
Figure 6.3 Flowchart for pixel-level fusion of aerial imagery with LiDAR-derived nDSM.	159
Figure 6.4 The training objects distribution on the subset of multispectral imagery of the semi-urban study area.....	160
Figure 6.5 Flowchart for the user-driven feature-level fusion of multispectral imagery with LiDAR-derived nDSM for semi-urban landscape mapping.....	164
Figure 6.6 Scale-parameter for the multi-resolution segmentation: (a) coarse size features with large scale-parameter of 40, (b) medium size features with moderate scale-parameter of 25 and (c) small size features with small scale-parameter of 10	165

Figure 6.7 The class hierarchy for the feature-level fusion of semi-urban study area	167
Figure 6.8 Fuzzy membership functions for the discrimination of <i>Colour</i> and <i>Grey Roofs</i>	168
Figure 6.9 Fuzzy membership functions for the discriminating <i>Grass</i> and <i>Tree</i> classes in multispectral imagery	169
Figure 6.10 Fuzzy membership function for the discrimination of <i>House</i> and <i>Infrastructure</i> using nDSM values.....	170
Figure 7.1 Tree species identification using pixel-level fusion of multispectral imagery with LiDAR-derived height and texture data	173
Figure 7.2 Tree species identification using feature-level fusion of multispectral imagery with LiDAR-derived height and texture data	176
Figure 7.3 Visual comparisons of the pixel- and feature-level fusion results for tree species delineation	179
Figure 7.4 The results of the pixel-level fusion of multispectral imagery with LiDAR derived nDSM data	182
Figure 7.5 The results of the feature-level fusion of multispectral imagery with LiDAR-derived nDSM data.....	185
Figure 7.6 Final tree species map of the forest study area derived from feature-level fusion of multi spectral imagery and LiDAR data.....	190
Figure 7.7 the 3-band colour imagery with nDSM data fusion results (a) pixel-level fusion and (b) feature-level fusions	192
Figure 7.8 (a) Pixel-level classification of colour image, (b) Pixel-level fusion of colour image with LiDAR-derived nDSM data.....	194
Figure 7.9 (a) Feature-level classification of colour image, (b) Feature-level fusion of colour image with LiDAR-derived nDSM data.....	195
Figure 7.10 Summary of the overall object delineation accuracy of pixel- and feature-level fusions for the semi-urban study area	196
Figure 7.11 3D models of the same semi-urban area created from (a) LiDAR-derived DSM, and (b) stereo photogrammetry	199

LIST OF TABLES

	Page
Table 2.1 Data characteristics equivalent to levels of interpretation for remotely sensed data (after Anderson <i>et al.</i> , 1976)	17
Table 2.2 Residential land cover classification subsystem	18
Table 2.3 Relationship between the level of detail required and the spatial resolution of various remote sensing systems for vegetation inventories, after Botkin <i>et al.</i> , (1984).	21
Table 2.4 Qualitative descriptions for the strength of agreement for the Kappa statistics, after Landis and Koch, (1977).....	69
Table 2.5 A summary of the literature review presenting major component of the current research.....	81
Table 3.1 Major landscape types in the study region	94
Table 3.2 Summary descriptions of the datasets	106
Table 4.1 Summary of the camera models	122
Table 4.2 Summary of the orthorectification results with quality evaluation (RMSE in microns).....	128
Table 4.3 Summary of residuals of exterior orientation parameters for the multispectral images of forest study area	129
Table 4.4 Summary of residuals of exterior orientation parameters for the colour image of the semi-urban study area.....	129
Table 5.1 Original and derived layers used in the proposed fusion procedures.....	143
Table 5.2 Mean values and standard deviation for each of the tree species generated from pixel-level fusion	145
Table 5.3 Mean value and standard deviation for each of the tree species generated from feature-level fusion	152
Table 6.1 Landscape classification system developed for the semi-urban study area	156
Table 6.2 Training statistics derived from 4-band multispectral image with nDSM data	162
Table 6.3 Segmentation parameters for the aerial imagery with LiDAR-derived nDSM data fusion for level III landscape classification scheme	166
Table 7.1 Error matrix for pixel-level fusion using multispectral imagery with LiDAR-derived nDSM data	175

Table 7.2 Error matrix for feature-level fusion using multispectral imagery with LiDAR-derived nDSM data.....	177
Table 7.3 Comparison of pixel- and feature-level fusion accuracy at the 95 percent confidence level	180
Table 7.4 Error matrix for the pixel-level fusion using multispectral imagery with LiDAR data.....	183
Table 7.6 Error matrix for the feature-level fusion using multispectral imagery with LiDAR data.....	186
Table 7.8 Comparison of pixel- and feature-level fusion results at the 95 percent confidence level	188
Table 7.9 Comparison of pixel- and feature-level fusion for colour image and LiDAR data fusion (summary of Table A4 a & c)	193
Table A4a Error matrix for pixel-level fusion using colour imagery with LiDAR data	238
Table A4b Segmentation parameters for the colour imagery with LiDAR-derived nDSM data fusion for level III landscape classification scheme	240
Table A4c Error matrix for the feature-level fusion using using colour imagery with LiDAR data.....	240
Table A5a Error matrix for pixel-level classification of colour imagery	242
Table A5b Error matrix for feature-level classification of colour imagery	243

CHAPTER 1

INTRODUCTION

This chapter discusses the concept of remote sensing data fusion as a framework to extract different landscape objects. Different data fusion-levels are outlined and their comparative advantages and disadvantages are discussed. The objectives of the research are stated, as well as the benefit and significance of the research. Lastly, the research methods are formulated and the thesis structure is outlined.

1.1 Data Fusion in Remote Sensing

Remotely sensed data have been used widely for monitoring and mapping landscapes at different spatial and temporal scales. Today, the number of Earth observation satellites and airborne sensors has increased substantially, leading to the increased availability of data at many spatial, temporal and spectral resolutions. It is now possible to have several concurrent images of the same scene providing different types of information. Therefore, remotely sensed data fusion has emerged as a new and important research area. Data fusion is a process that deals not only with combining multi-source images, but also provides improved information for decision-making (Hall and McMullen, 1997). The most common case is the fusion of two images acquired respectively by a multispectral sensor and by a panchromatic sensor having higher spatial resolution. The improvement in spatial resolution of the resulting fused image reveals more detail at a larger scale than the original multispectral image. This improvement is noteworthy also in respect of the panchromatic image given the enhanced spectral content of the new image.

In the literature a large number of definitions can be found for the term data fusion. A comprehensive list of definitions and their sources are found in Wald (1999) and Pohl and van-Genderen (1998). Keys *et al.* (1990) and Franklin and Blodgett (1993) described the term *fusion* in a pixel context. However, according to Wald (1999, p. 1191) “data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of ‘greater quality’ will depend

upon the application”. *Image merging* (Carper *et al.*, 1990) and *image integration* (Wehch and Ehlers, 1988) are equivalently used in the fusion context. *Data integration* (Nandhakumar, 1990) represents a broader view and comprises image fusion algorithms as well. In this case, not only are remote sensing images fused, but also further ancillary data (such as topographic maps, GPS coordinates and geophysical information) contribute to the resulting image (Harris and Murray, 1989). According to Wald (1998) the term *integration* implicitly refers more to concatenation (*i.e.* increasing the stack vector) than to the extraction of relevant information. According to Richards and Jia (2005), labelling pixels by drawing inferences from several available sources of data is referred to as *data fusion* or *multi-source classification*. As with the treatment of single image datasets, analysis of mixed data types can be carried out photo-interpretatively or by using matching analyses.

In this study, the term *data fusion* is defined to cover the whole aspect of the *multi-source classification* (Richards and Jia, 2005) and is not restricted only to pixel domain, but is also extended into feature domain. This term also includes the geometric fusion of aerial imagery with LiDAR data as well as the extraction of landscape information from their complementary properties.

In general, data fusion techniques can be divided into three categories according to the stage at which the fusion is performed: pixel-, feature- and decision-level fusion (Schistad-Solberg *et al.*, 1994; Pohl and van-Genderen, 1998). The concept of data fusion is visualised in Figure 1.1.

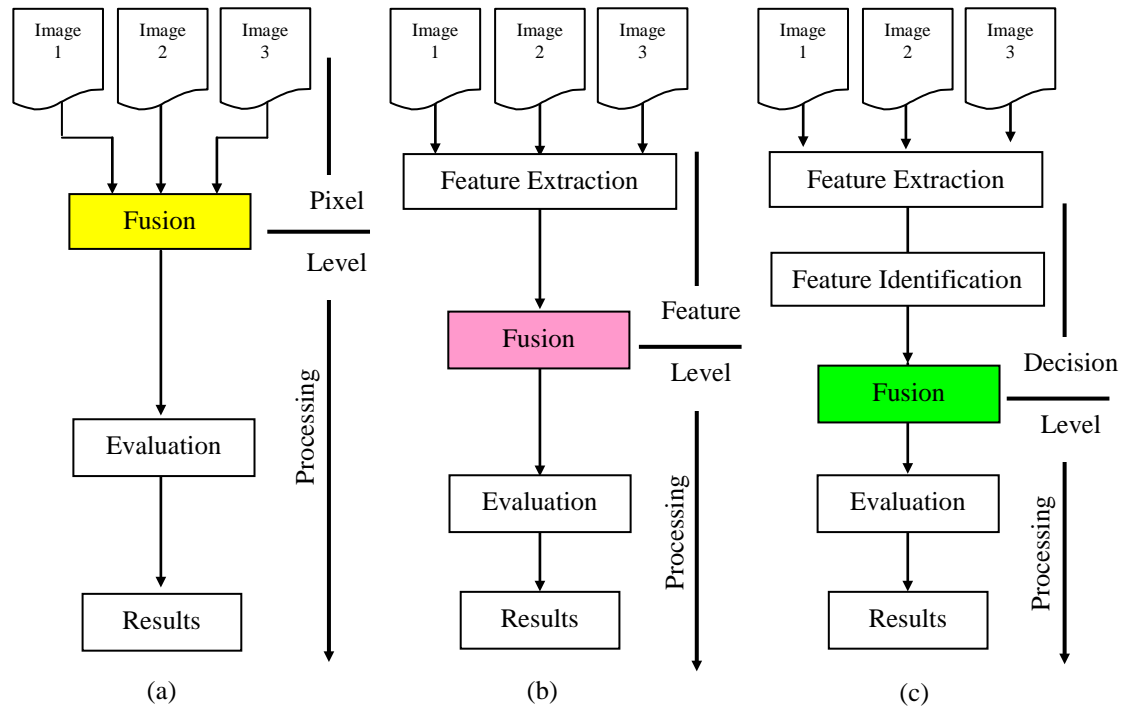


Figure 1.1 Processing levels of data fusion, after Pohl and van-Genderen (1998)

1.1.1 Pixel-level fusion

At the lowest level, pixel-level fusion uses the registered pixel data from all image sets to perform detection and classification functions. This level has the potential to achieve the greatest fusion performance only at the highest computational expense (Waltz, 2001). Figure 1.1(a), illustrates the concept of pixel-level fusion. In this process, registration and subsequent resampling play a vital role in aligning all image data into a common pixel spacing and map projection. Misregistration errors cause artificial colour of objects in multi-sensor data, which falsifies the interpretation later on (Pohl and van-Genderen, 1998).

1.1.2 Feature-level fusion

At the intermediate level, feature-level fusion combines features that are detected and segmented in the various data sources. Figure 1.1(b), shows a conceptual model of a feature-level fusion. Features that correspond to the characteristics of landscape objects are extracted from initial data sources dependent on their characteristics such as extent, shape and neighbourhood. Features identified from multiple sources create a common feature space for further classification.

1.1.3 Decision-level fusion

Fusion at the decision level combines decisions of independent sensor detection/classification paths by applying decision rules. The main drawback of this process is that decision uncertainty in each sensor chain is maintained and combined with a composite measure of uncertainty (Waltz, 2001). Figure 1.1(c) illustrates a conceptual framework of the decision-level fusion.

The above mentioned three-levels of processing are the basic building blocks of multi-source data fusion. During a complex process, these levels might be combined. In all cases, the aim is the extraction of useful information included in the source data while avoiding the introduction of artefacts harmful to human observations or matching analyses (Laporterie and Flouzat, 2003). Fusion can also be divided into two broad categories: data- and user-driven models. The data-driven model is dependent on the data itself and user-driven model is closely controlled by the users. Remotely sensed data fusion requires well defined techniques as well as a good understanding of the input data (Pohl and van-Genderen, 1998). An introduction to fusion of aerial imagery and LiDAR data is presented in the following section.

1.2 Fusion of Aerial Imagery with LiDAR Data

LiDAR is a relatively recent development in remote sensing with great potential for creating high resolution DSMs. LiDAR is an active sensor that uses its own laser beam for acquiring 3D point clouds of the Earth's surface. In contrast, multispectral aerial imagery systems use passive sensors and utilise the sun's illumination as a source for capturing spectral information about the Earth's surface.

LiDAR data allow an accurate assessment of landscape objects in the vertical plane (height). However, current LiDAR sensors have limited coverage in the horizontal plane or electro-magnetic spectrum. Conversely, multispectral imagery provides extensive coverage of landscape classes in the colour spectrum but is relatively insensitive to height variation. The complementary nature of these two datasets has the potential to significantly improve the extraction and measurement of landscape objects.

The strong argument in favour of the fusion of aerial imagery with LiDAR data is that LiDAR points are not distributed evenly and usually have gaps. As a result, the 3D structure of objects might not be well defined (Baltasvias, 1999). It thus becomes fairly complex to obtain a good 3D model of the Earth's objects with a low density of LiDAR returns. The idea of exploiting the complementary properties of aerial imagery and LiDAR data is to extract additional information from the aggregated data for a more complete surface description.

1.3 Problem Statement

Pixel-level fusion is simply the merging of pixels from images that correspond to the same feature space. Although this concept is easy to understand, several issues require consideration.

Firstly, consider the geographic registration and measurement scale of each data source. Since different data sources are likely to have different registration and different measurement scales, it is required to bring all datasets into the same registration and same spatial scale. In passive aerial imagery, data is collected in the electro-magnetic spectrum as a ground reflectance unit. In active LiDAR sensors, object height is recorded as a phase difference between the transmitted and received signal backscattered from the objects. When these two data sources are fused then the choice of a data normalisation scheme is conceptually complex.

A second issue is the determination of the contribution of various data sources in a fusion model. Data sources that display a large-scale variation dominate the fusion process. For instance, if the first data source in a two-source case has a dynamic range of 8 bits, and the second has a dynamic range of 16 bits, then the second data source is more likely to dominate the fusion process due to the effects of the measurement scale.

A third issue concerns the uncertainties inherent in different data sources. In pixel-level fusion, each data source contributes equally to the fusion process in determining the location of the decision boundaries in feature space. In practice, it is not possible to get equally reliable data, hence reducing the possibility of achieving highly accurate fusion results.

Finally, fusion applications can be developed through data- and user-driven approaches. The data-driven technique is dependant on the data itself and used when little is known about the data before fusion. In contrast, user or user-derived rules closely control user-driven techniques. Results can vary awing to these different approaches so we need to address this for different landscape types.

Traditionally, data fusion is tackled at the *pixel-level*, however *feature-level* fusion has gained attention recently due to the availability of robust segmentation algorithms and the reduction of computational expense. Feature-level fusion starts with the crucial initial step of grouping neighbouring multi-source pixels so that they represent meaningful features. In the segmentation and topology generation process, the resolution and the scale of the expected features play a vital role (deKok *et al.*, 1999). Once the overall image has been segmented, measurements are performed on each region and adjacency relations between regions can be investigated (Soille, 2003).

The major drawback of pixel-level fusions is that they often lack spatial consistency. A pixel is fused depending on its spectral values, regardless of its neighbour values. Fusions are therefore very sensitive to noise and do not take spatial information into account (Soille, 2003), as a result, the spectral information is not sufficient to recognize objects in high spatial resolution images, particularly when the GSD is less than 5m (Rego and Koch, 2003; Blaschke *et al.*, 2001). In these situations, sun illumination angle and shadow can play a vital role in recognizing objects in the landscape. In pixel-level spectral analysis, the same objects might be recognized as different objects. Although the improvement in spatial resolution can reduce the problem of mixed-pixels, the internal variability and the noise within object are increased. As a consequence, traditional pixel-level fusions produce too many, or poorly defined, classes because their clusters are built upon spectral homogeneities only (Schiewe *et al.*, 2001). As an alternative to pixel-level fusion, feature-level fusions have an ability to improve accuracy and interpretability in high spatial resolution images (Aplin *et al.*, 1999).

Although feature-level fusions have some promising attributes, there are several critical points that need to be considered. Currently, feature-level fusion algorithms

are heavily reliant on users' input for the knowledge rules. A common criticism of this user-derived approach is that the user needs to have a significant knowledge of the features of interest in order to choose the best parameters for identifying and classifying the features from multi-source data (deKok *et al.*, 1999). The user needs to be aware of the spatial and spectral behaviour of the features in different data sources, understand the underlying processes, and have good ground information to define the decision key. Sometimes the 'ideal user' does not exist. Another problem area is the transferability of an existing protocol of a feature-level fusion to a new area. The user must use his or her judgment (Flanders *et al.*, 2003). With respect to the degree of automation, the feature-level fusion is time-consuming and requires manual interactions (Schiewe *et al.*, 2001). Firstly, during the segmentation process the choice of scale parameters, as well as the setting of weights for the input data sources, has to be chosen by the user. Secondly, in contrast to human analysis, the segmentation and the fusion steps are strictly separated. Lastly, the segmentation and subsequent processes that will produce a fusion result need to be evaluated in terms of their accuracy before they can be used for their designated purpose.

Decision-level fusion is not included in this research. It is more complex and has more uncertainty than pixel- and feature-level fusions.

1.4 Research Objectives

The aim of this research is to investigate data- and user-driven fusion models for extracting forest and semi-urban landscape objects by the synergy of active and passive sensor-derived data. Pixel- and feature-level approaches are particularly investigated in order to evaluate the potential of aerial imagery and LiDAR data fusion and also to ascertain the best fusion model. Thematic accuracies of the fusion results are evaluated against known landscape objects derived from field-surveyed data. The best performing fusion model is selected for landscape object extraction for each of the study areas. The fulfilment of these research objectives should enable answers to be found to the following research problems identified for the particular study area:

- (a) can a robust methodology be developed for fusing aerial imagery with LiDAR data?

- (b) what are the relative merits of pixel- and feature-level fusion methodologies in data- and user-driven domains?
- (c) is thematic mapping accuracy improved by fusing aerial imagery with LiDAR data?

1.5 Significance and Benefits of the Research

The number of active and passive sensing systems has increased substantially in recent years and data are available in many spatial, spectral and temporal resolutions. Therefore, remotely sensed data fusion has emerged as an important research area. The fusion of disparate and complementary data not only enhances the image quality, but also increases the reliability of the interpretation and classification. The fused data provide robust operational performance, such as increased confidence, reduced ambiguity, improved reliability and improved classification (Schistad-Solberg *et al.*, 1994; Rogers and Wood, 1990).

Data fusion techniques take advantage of the different physical natures of aerial imagery and LiDAR systems and enhance the mapping of various landscape objects. The object enhancement capability of fusion is visually apparent in pan-sharpened images (fusion of panchromatic and multispectral images) that are superior to both original images. The fusion of aerial imagery with LiDAR sensor-derived data enhances semantic capability of the images and yields information that is otherwise unavailable or hard to obtain from single-source data. Fused images maximise the amount of information that can be extracted from remotely sensed data.

Thematic mapping accuracy is improved when multiple remotely sensed data are introduced in the fusion process. The classification of multispectral data relies solely on the spectral signature of landscape objects. Some vegetation species cannot be separated due to their similar spectral response. LiDAR data can contribute in a different way, through difference in height, shape and roughness of the observed landscape objects. The use of multi-source data fusion becomes more popular with increased availability of sophisticated software and hardware facilities to handle the increased volume of data (Pohl and van-Genderen, 1998). A new trend in this respect is feature-level fusion or Object Based Image Analysis (OBIA), because the conventional pixel-level fusion is less capable of processing multi-source data. This

research is significant for the fusion of aerial imagery with LiDAR data because it provides:

- (a) a comprehensive approach to the analysis of landscape structure using multi-scale and multi-source remote sensing data. Different pixel- and feature-level fusions are compared, and the potential to improve thematic mapping accuracy is demonstrated
- (b) a better understanding of the performance of data- and user-derived fusion models for landscape mapping
- (c) a robust methodology to interpret complex landscape classes, and
- (d) results can be adopted in other environments with similar conditions

1.6 Research Methodology

The sources of multi-scale remotely sensed data used for this research are aerial imagery and LiDAR data. Broadly the approaches used in this research are to:

- (a) review the use of remotely sensed data fusion techniques for environmental information extraction, focusing on different pixel- and feature-level fusion techniques. The effect of data scale in relation with the landscape scale is reviewed to find an optimum scale for site-specific fusion analysis. Thematic accuracy as well as comparative accuracy are assessed
- (b) define a landscape classification scheme suitable for mapping a range of landscape objects using the spectral and spatial properties of the multi-source remotely sensed data. Select the study areas for implementing fusion models and formulate a sample data collection strategy for evaluating the thematic accuracy of the fused maps
- (c) compile relevant aerial imagery and LiDAR data for the study areas in a format suitable for analysis and fusion. Also collected ancillary data such as obtained from topographic maps and satellite images for better understanding of the study areas

- (d) formulate a comprehensive field-surveying strategy for collecting GCP positions, validating LiDAR-derived height data and collecting samples for training and accuracy assessment. A complex sampling strategy is developed to balance statistical validity and practical application.
- (e) orthorectify aerial imagery with collected GCPs using digital planimetric data derived from onboard navigation systems. A rigorous geo-referencing system is employed to fine-tune the geometric registration of the optical imagery and combine all the multi-source data in the same pixel spacing. A normalised DSM is prepared from the LiDAR-derived first and last return data for representing a landscape object's height. The LiDAR-derived height data is validated using field data
- (f) develop a comprehensive fusion analysis framework to fuse aerial imagery with LiDAR data at pixel- and feature-level. Different data- and user-driven fusion processes are developed to better delineate landscape-specific objects in different study areas
- (g) assess the thematic accuracy of the fusion results using field-derived sample data. Evaluate the recognition of different landscape objects, considering variations in spectral and spatial properties of the fused data
- (h) summarise landscape properties through analysis of fused results. Compare thematic accuracies of the fusion results and evaluate the suitability of the fusion techniques for each of the study areas

1.7 Thesis Structure

The thesis comprises of eight chapters. Chapter 1 outlines the research objectives and the background of data fusion as a tool to improve landscape mapping.

Chapter 2 reviews the principles of different data fusion techniques and their comparative advantages and disadvantages. Previous research regarding remotely sensed data fusion and its application for landscape mapping is discussed. The techniques that are available for the fusion of aerial imagery with LiDAR data are

reviewed. Accuracy assessments of pixel- and feature-level fusions are discussed and compared using statistical, visual and graphical analysis techniques.

The location and the characteristics of the major landscape classes of the study areas are given in Chapter 3. The aerial imagery and the LiDAR data are introduced and fusions for extracting different landscape objects are justified. In addition to aerial imagery, the topographic maps and extensive field-surveyed data that are used in this analysis are presented.

Chapter 4 presents the geometric correction procedures for the datasets. Geometric correction consists of two steps: orthorectification of aerial imagery and nDSM derived from LiDAR data. The geometric corrections are an essential precursor to the fusion of aerial imagery with LiDAR data. Orthorectification provides a common planimetric base for the fusion of multi-source data. LiDAR-generated nDSM provides an object's height, which is later used in the fusion process.

Chapter 5 focuses on the implementation of data-driven pixel- and feature-level fusion models for delineating individual tree attributes in the forest study area. The data-driven fusion models are applied to the forest study area for the presence of three Eucalyptus species (*camaldulensis*, *largiflorens* and *microcarpa*). Data-driven fusion model interpretations solely on the data are used when little is known about data before fusion. In pixel-level fusion, an unsupervised algorithm is implemented for the fusion of multispectral imagery with LiDAR data. In feature-level fusion, a watershed segmentation technique is used for delineating individual tree features from the forest study area then masking technique is used for collecting tree feature attributes from multi-source data. Finally an unsupervised algorithm is applied to the tree features to discriminate different tree species.

Chapter 6 focuses on the implementation of different user-driven fusion models for differentiating between six classes of landscape structures in a semi-urban study area. Users closely control user-driven fusion models and knowledge rules are applied in the fusion process. In pixel-level fusion, supervised algorithms are implemented for the fusion of aerial imagery with LiDAR-derived nDSM data. In feature-level fusion, multi-resolution segmentation helps to define landscape features, and subsequently a rule-based fuzzy algorithm is applied.

Chapter 7 presents the overall findings of this research. The accuracies of pixel- and feature-level fusion results are discussed for both the forest and semi-urban study areas. Fusion results are compared and performances of the fusion models are discussed. In the end, research objectives are re-examined in light of the findings of this research.

Chapter 8 presents the conclusions based on the findings of this research. Experimental results are reviewed in the context of overall research and recommendations are made for future work.

CHAPTER 2

A REVIEW OF REMOTELY SENSED DATA FUSION PROCESSES

This chapter reviews the principles of remote sensing data fusion for distinguishing a range of landscape objects. The effect of scale in a hierarchical landscape classification scheme is reviewed in the framework of extracting different landscape objects from multi-source data. Different fusion techniques are presented and the accuracies of results are evaluated to ensure the quality of fusion products. Research and development regarding the fusion of aerial imagery with LiDAR data are also presented.

2.1 Introduction

Fusion techniques for remotely sensed data are used to enhance image interpretation in an ever-growing number of cases. Objectives of data fusion, as well as the methods used, are various; exhaustive overviews of the arguments can be found in Wald (1999), and Pohl and van-Genderen (1998). Among the most important aims of data fusion is the production of spatially improved images suitable for classification (Teggi *et al.*, 2003). For a classification task, the goal of fusing data from different sensors is to reduce the classification error compared to single source classification (Schistad-Solberg *et al.*, 1994). A decision on which fusion technique is the most suitable is very much driven by its application. Therefore, it is very difficult to make a general statement on the comparative efficiency of different fusion techniques (Pohl and van-Genderen, 1998).

Before applying any fusion technique two most fundamental points need to be considered: the effect of scale on multi-source remotely sensed data and its relation with the objects to be extracted (Quattrochi and Goodchild, 1997). The following sections review these issues in the context of fusion applications.

2.2 Review of Scale Effect in Image Fusion

Scale is an important factor in the application of remote sensing that directly influences analysis. The scale effect was considered as the first challenge of using of

remotely sensed data for earth observation (Raffy, 1994). The fundamental reason for the continuing interest in scale in remote sensing is that spatial resolution is the primary scale of measurement (Atkinson and Aplin, 2004). As the number of sources of multispectral remotely sensed data increases and the variety of objects to be differentiated becomes greater the selection of appropriate spatial resolution is a critical factor identifying these objects.

In remote sensing, pixel size is considered as a measurement scale (spatial resolution), but it takes a number of pixels (operational scale) for a feature (e.g., buildings, roads, water bodies) to be recognized, and a much larger number of pixels (geographical scale) to understand spatial patterns in landscapes (Cao and Lam, 1996). To select imagery with appropriate spatial resolution for a study, one should examine the characteristics of the scene and how scale and resolution affects the determination of objects in that scene (Han *et al.*, 2008). Several methods of study the effects of scale and resolution have been suggested. Cao and Lam (1997) confirmed Benson and Mackenzie's (1995) experiment of scale effects on landscape. Landsat TM (band 4) data for an area in northeastern Wisconsin was acquired at a resolution of 30m per pixel. This area contained numerous lakes of various sizes. Application of a pixel aggregation algorithm successively generated 60, 120, 240, 480 and 960m per pixel resolution images. After six images were produced, three basic landscape parameters were extracted from each: (a) the percentage of the scene covered by water, (b) the number of lakes included, and (c) the mean surface area of the lakes. Figure 2.1 graphs the results of this data extraction process.

As the ground resolution cell size increases from 30 to 960m, the percentage of water defined as lake at first increased slightly and then continually decreased (Figure 2.1a). At the same time, the number of lakes decreased in a non-linear fashion and the mean lake surface area increased nearly linearly (Figure 2.1b,c). This study examines how landscape parameters change with scale (e.g., the effect of spatial resolution).

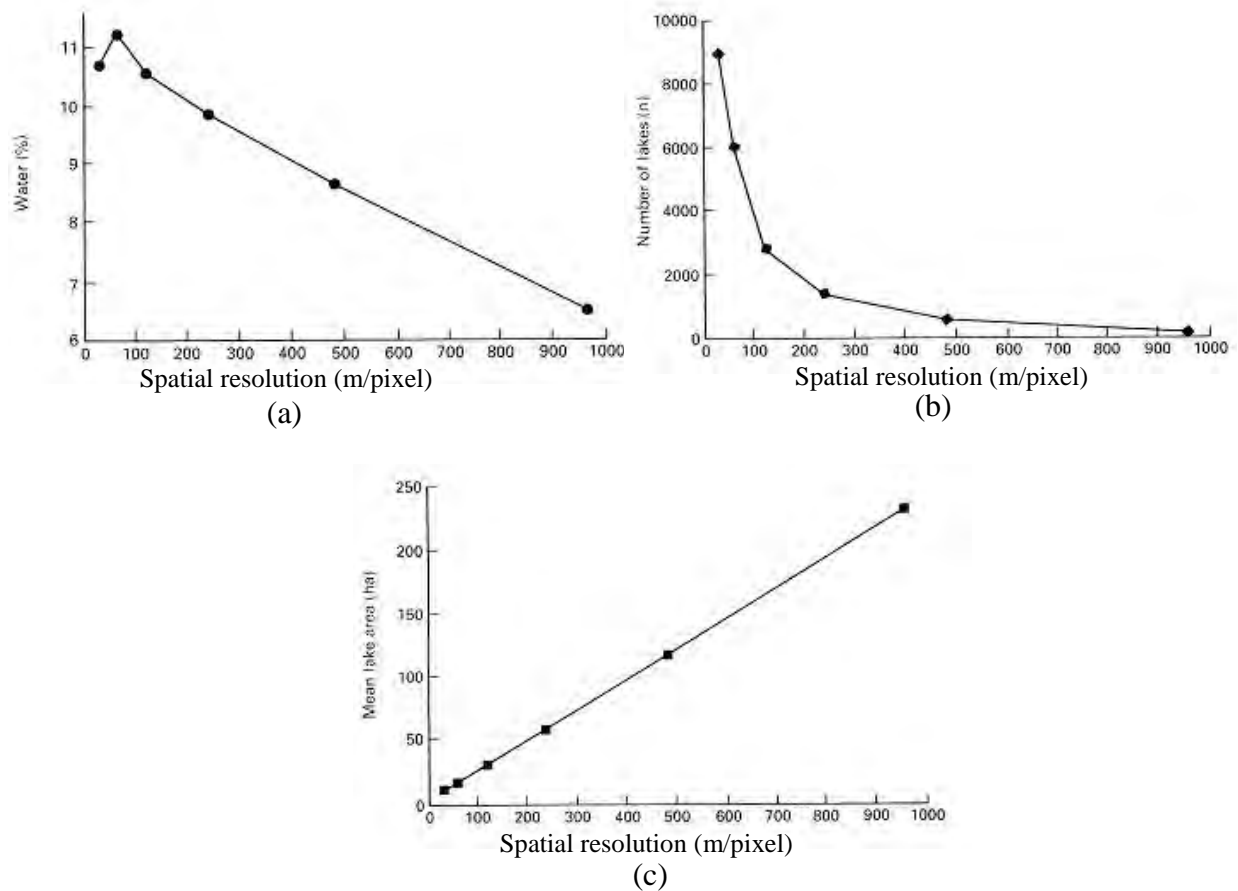


Figure 2.1 Influence of spatial resolution on measurement of selected landscape parameters (*after* Benson and Mackenzie, 1995)

The spatial resolution of an image has intricate effects on image classification. Markham and Townshend (1981) found that image classification accuracy is affected by two factors. The first factor is the influence of boundary pixels on classification results. As spatial resolution becomes finer, the proportion of pixels falling on the boundary of objects in the scene decreases. Boundary pixels have a mix of elements so reducing the number of boundary pixels reduces confusion in the classification process, resulting in higher classification accuracy. The second factor which influences classification accuracy is that the spectral variation of land cover types also increases in finer spatial resolution images. Variation within a land cover class decreases the spectral separability of classes and results in lower classification accuracy. The net effect of finer spatial resolution is a combination of these two opposing factors, which may vary in importance as a function of the relative size of the object under observation. The number of mixed pixels containing more than one

land cover type is a function of both the complexity of the scene and the spatial resolution of the sensor (Simonett and Coiner, 1971)

Woodcock and Strahler (1987) investigated the effect of spatial resolution on the local variance observed in remotely sensed imagery for object classes. Woodcock *et al.* (1988) and Atkinson and Aplin (2004) used a variogram to choose an appropriate spatial resolution for remote sensing investigations. A method based on statistical separability was developed to explore the scale effect on remotely sensed classification and to thereby determine optimal resolution (Bo *et al.*, 2005). An approach based on entropy was proposed by Han *et al.* (2008) to select an optimal scale in image classification with desirable overall classification accuracy. Entropy was used to describe uncertainty in image classification.

The choice of an appropriate scale, or spatial resolution, for a particular application depends on several factors. These include the information desired about the ground scene, the analysis methods to be used to extract the information, and the spatial structure of the scene itself (Woodcock and Strahler, 1987). As a result, the objects on the ground that are to be mapped play a vital role in the selection of the scale. Airborne LiDAR and optical imaging systems can achieve very high spatial, spectral and textual resolutions. As a result fusion of these sensor-derived data is very effective for site-specific landscape mapping; however, the direct relationship between spatial resolution and the landscape classification scheme need to be considered before implementing any fusion model.

2.3 Review of Hierarchical Landscape Classification Schemes

Standardised landscape classification schemes are required to encourage efficient use of the multi-source fusion methods and effective management of resources. Consistent and robust landscape data from multiple sources are essential for planning for management of the environment. Application of standardised land cover classification schemes provides a fundamental framework for the establishment of data fusion systems at local, regional and national scales.

There is no ideal land cover classification scheme and it is unlikely that one could ever be developed. There are different perspectives in the classification process and

the process itself tends to be subjective, even when an objective numerical approach is used (Anderson *et al.*, 1976). Geographic data are imprecise in nature due to multidimensionality and inherent fuzziness of many features and their inter-relationships. Current land cover classification schemes are based on hard boundaries between the classes, which are derived from *a priori* knowledge and are difficult to use. Nevertheless, land cover classification schemes are widely used because they are scientifically based and individuals using the same classification system can compare their results (Jensen, 1996).

The most widely utilised multi-level land cover classification scheme was developed by the USGS (Anderson *et al.*, 1976) to accommodate different sensor-derived data at a range of spatial resolutions. The USGS classification scheme comprises four levels of land use and land cover classes arranged in a hierarchical manner. Relationship among the classification levels, scale and pixel size are presented in Table 2.1.

Table 2.1 Data characteristics equivalent to levels of interpretation for remotely sensed data (after Anderson *et al.*, 1976)

Classification level	Scale	Pixel size (m)	Data source
I	Less than 1:100 000	80-30	Medium resolution satellite data
II	Less than 1:80 000	3-2	Moderate resolution satellite data and high altitude aerial photography
III	1:80 000 to 1:20 000	2-1	High resolution satellite data and medium altitude aerial photography
IV	Larger than 1:20 000	1-0.5	Very high resolution satellite data and low altitude aerial photography

Levels I and II classifications are suitable for employing data on a nationwide, interstate, or statewide basis. More detailed land cover data such as those obtained at Levels III and IV are used for regional, county or municipal level mapping. Only Levels I and II have been defined by Anderson *et al* (1976), and these are suitable for the classification of medium resolution satellite imagery. The USGS classification

system was designed primarily for interpretation of analog imagery and does not consider the availability of higher spatial resolution data nor the implications of multi-source data fusion techniques.

The USGS classification system can be extended to more detailed levels (Anderson *et al.*, 1976). For example, residential sub classes are given in Table 2.2. The breakdown of the residential class employs criteria of capacity, type and permanency of the residence as the discriminating factors among the sub-classes. Level III classification scheme is suitable for classifying images at scales ranging from 1:20,000 to 1:80,000 and requires use of supplementary information. The categories under this level are designed to be adaptable to the local needs of public agencies. Level IV classification scheme is most useful for aerial photos at scales larger than 1:20,000 and is suitable for multi-source data fusion. This level of classification is good for site-specific landscape mapping and assessment. As a whole, spatial resolution of the data plays a vital role in the selection of categories within classification scheme and allows aggregation and transfer between categories.

Table 2.2 Residential land cover classification subsystem

Level I	Level II	Level III
1 Urban or Built-up area	11 Residential	111 Single-family Units 112 Multi-family Units 113 Group Quarters 114 Residential Hotels 115 Mobile Home Parks 116 Transient Lodging 117 Others

There is a direct relationship between the level of detail in a classification scheme and the spatial resolution of remotely sensed data used (Jensen, 1996). Welch (1982) summarised the relationship between spatial resolution and the mapping requirement for urban land cover at levels I to IV in the United States systems as in Figure 2.2. The main finding of his study was that for mapping level II classes the spatial resolution is required to increase dramatically over level I.

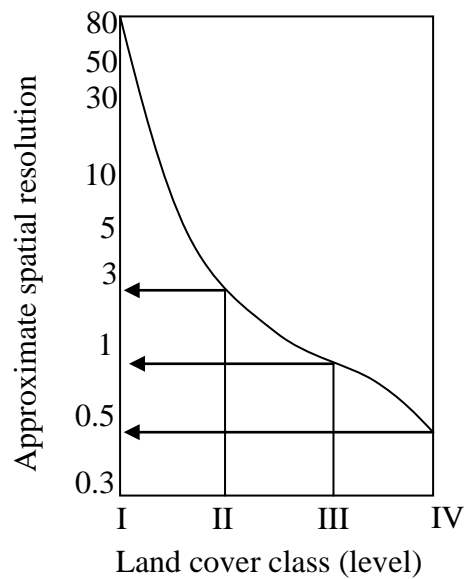


Figure 2.2 Spatial resolutions as a function of the land cover classification level, after Welch (1982)

2.3.1 Urban landscape classification scheme

Mapping land cover types requires a different perspective than for urban remote sensing (Herold *et al.*, 2006). Land cover considers characteristics in addition to those of different materials. The surface structure (roughness) affects the spectral signature as much as normal variations within the land cover type (e.g. buildings having different roof angle or cover area). Two different land cover types (e.g. asphalt roads and composite shingle or tar roofs) can be composed of very similar materials (hydrocarbons). From a material perspective, these surfaces would map as the same. Additional information (e.g. height data) can play a vital role in separating them. Thus, analysis of urban land cover would benefit by data fusion that contributes different sensor-derived unique information.

Another issue for urban land cover mapping is the problem of shadowing, where the object causing the shadowing shows a dramatically increased elevation across a short distance (Dare, 2005; Yuan, 2008; Zhou *et al.*, 2009). With the dominance of elevated objects such as buildings, bridges, towers and trees in the landscape, the proportion of the imagery that is affected by shadowing can be significant (Yuan and Bauer, 2006). Shadows in remotely sensed imagery occur when objects totally or partially obstruct the direct light from a source of illumination, casting shadows

(shadows cast on the ground by high-rise objects), and self shadows (the side of the object not illuminated) (Salvador *et al.*, 2001; Yao and Zhang, 2006). The influence of shadows in remote sensing has always been an important issue (Dare, 2005). Although shadows have been used to aid in reconstruction of three-dimensional geometry, such as measurement of the shape and height of buildings (e.g., Liow and Pavlidis, 1990; Shettigara and Sumerling, 1998), in most of the cases, shadows are considered a nuisance (Zhou *et al.*, 2009). Great difficulty arises in classification and interpretation of shaded objects in an image because of the reduction or total loss of spectral information from them (Dare, 2005; Yuan, 2008, Zhou *et al.*, 2009).

Multisource data fusion is one effective method for minimizing shadow problems in high resolution imagery (Zhou *et al.*, 2009; Yuan, 2008). A typical multi-source data fusion procedure for shadow restoration is to first identify the shaded pixels in the image of interest, and then replace the shaded pixels in the image with non-shaded pixels of the same region from another image acquired at a different time (Dare, 2005). Shackelford and Davis (2003) eliminated shaded areas in urban land cover classification by simply leaving them as unclassified areas, resulting in significant loss of land cover information (Zhou *et al.*, 2009). In addition to radiometric restoration and multi-source data fusion, spatial information such as adjacency relations can be used for classification of shaded areas in high spatial resolution imagery (Yuan and Bauer, 2006; Zhou and Troy, 2008). In this case, shadow classification refers to classification of the shaded areas into different land cover types, rather than classifying shaded areas into different types of shadows such as self shadow and cast shadow (Salvador *et al.*, 2001).

The design of an effective landscape classification scheme relies on the recognition of the target and sensor characteristics in conjunction with the interpretation approach to be applied. The urban landscape classification scheme is designed to utilise remotely sensed data as the primary information source, especially at the more detailed levels, and to incorporate multi-source data to assist in the understanding of landscape arrangements for more detailed interpretation. The range of data characteristics applicable to the derivation of each level of interpretation is given in Table 2.1. All low altitude airborne data are useful for interpretation of Level III to

IV classes. Higher spatial resolution generally enables more detailed categorisation of urban landscape, although in complex scenes the accuracy levels may be variable.

2.3.2 Forest landscape classification scheme

Botkin *et al.* (1984) developed a vegetation classification scheme, which is very similar to the USGS land cover classification scheme (Anderson *et al.*, 1976). The sensor systems and spatial resolution useful for discriminating between vegetation types at a global to *in situ* scale are summarised in Table 2.3.

Table 2.3 Relationship between the level of detail required and the spatial resolution of various remote sensing systems for vegetation inventories, after Botkin *et al.*, (1984).

Level	Geographic Scale	Spatial Resolution	Sensor
I	Global	1.1 km	AVHRR
II	Continental	1.1km-80m	AVHRR, Landsat Multi spectral Scanner
III	Biome	80m-30m	Landsat Multi spectral Scanner, Thematic Mapper, Synthetic Aperture Radar
IV	Region	30m-3m+	Landsat Thematic Mapper, High Altitude Aircraft, Large Format Camera, SPOT
V	Plot	3m-1m+	High and Low Altitude Aircraft, Small format Camera
VI	<i>In Situ</i> Sample Site	-	Surface Measurement and Observation

The Botkin *et al.*(1984) study showed that the level of detail required for the chosen classification system dictates what spatial resolution of the remotely sensed data is necessary. Spectral resolution is a very important consideration, however, it is not as critical a parameter as spatial resolution since most of the optical sensor systems record different light energy in approximately the same green, red, and near-infrared regions of the electromagnetic spectrum.

The classification scheme of Botkin *et al.*, (1984) was not directly relevant to multi-source data fusion in vegetation mapping. Table 2.3 shows the relationship between geographic scales, spatial resolutions and remote sensors. The main consideration in

determining tree species at plot level is to identify them in both spatial and spectral domains. The former is satisfied by the Level V classification, and the latter by maximizing the interclass variance and minimising the intraclass variance. Using single source remotely sensed data, these conditions may not be always achieved and other data sources are needed to compliment spectral and spatial information.

2.4 Review of Aerial Imagery and LiDAR Data Fusion

In landscape mapping, the fusion of aerial imagery with LiDAR data has a promising future. LiDAR provides very accurate position and height information, but less direct information of an object's geometrical shape, while aerial imagery offers very detailed information on the landscape objects, such as spectral signature, texture and shape. Fusing these two complementary datasets is quite promising for landscape object extraction and 3D modelling (Tao and Yasuoka, 2002).

The idea of exploiting the complementary properties of aerial imagery and LiDAR was first initiated by Schenk and Csatho (2002) to extract semantically meaningful information from the aggregated data for a more complete earth surface description. Haala and Brenner (1999) combined a LiDAR-derived DSM with three colour spectral bands of aerial imagery. In this context, the most problematic task is to separate trees from buildings with low-resolution LiDAR data and in the absence of the near-infrared band. Rottensteiner *et al.* (2004b) used a LiDAR-derived digital terrain model and NDVI from a multispectral image to separate buildings in densely built-up urban areas. Rottensteiner *et al.* (2007) showed that only LiDAR data can detect 95 percent of all buildings larger than 70 m²: buildings smaller than 30 m² could not be detected. Having additional multispectral imagery in the classification process improves the correctness of the results for small residential buildings by up to 20 percent.

Sohn and Dowman (2007) presented a building extraction method that automatically detects building objects and delineates their boundaries using IKONOS image and LiDAR data. The technique consists of a two-step procedure: building detection and building description. A building detection method was introduced to reduce a scene complexity in urban areas and simplify the building description process. The

technique subsequently detected dominant features comprising the urban scene, and finally isolated buildings from surrounding features. The results prove that terrain information extracted from LiDAR data together with chromatic cues provided by multispectral bands of IKONOS imagery can work together to detect buildings. Evaluation determined the completeness was 88.3 percent, the correctness was 90.1 percent and the quality was 80.5 percent. However, a qualitative error analysis showed that the extracted building polygons tend to overlook some buildings (false negative pixels) and misclassify some objects as buildings (false positive pixels) because of poor point density of the LiDAR data. These errors could be reduced if high-density LiDAR measurements are used.

Hofmann (2001) put forward the idea of object-based approach for detecting buildings and roads from high resolution satellite imagery using additional elevation information. Walter (2005) applied an object-based classification technique to multispectral and LiDAR data for detection of change in urban areas. Mutlu *et al.* (2008) fused LiDAR and multispectral data for predicting fire behavior using surface fuel models.

The fusion of high spatial resolution imagery with LiDAR data for determining individual tree attributes were reviewed by Baltsavias (1999) and Leckie *et al.*, (2003). The strong argument for fusion is that the LiDAR points are not distributed evenly and usually have gaps between them. As a result, the 3D structure of objects might not be very well defined (Baltsavias, 1999). It thus becomes fairly complex to obtain a good 3D model of tree canopy with low-density LiDAR returns. The idea of exploiting the complementary properties of aerial imagery and LiDAR data is to extract semantically meaningful information from the aggregated data for a more complete surface description (Tickle *et al.*, 2006). Sua´reza *et al.*, (2005) propose a data fusion technique using aerial photography with LiDAR data to estimate individual tree height in forest stands. The tree canopy model is derived from LiDAR layers as the difference between the first and last pulse returns. Information about individual trees is obtained by object-oriented image segmentation and classification. This analysis provides a good method of estimating tree canopies and heights. However, segmentation and classification are too user-dependent. The classification parameters are not defined automatically and exhibit no clear relationship to

allometry factors; instead, they are defined empirically following a trial-error process.

Bork and Su (2007) compared the suitability of LiDAR data, three-band multispectral data, and LiDAR data integrated with multispectral information, for classifying spatially complex vegetation in rangelands of the Aspen Parkland of western Canada. Classifications were performed for both three general vegetation classes (deciduous forest, shrub land and grassland) and eight detailed vegetation classes. A Digital Elevation Model (DEM) and Surface Elevation Model (SEM) developed from LiDAR data incorporated both topographic and biological biases in community positioning across the landscape. Using multispectral data, the original digital image mosaic, its hybrid colour composite, and an intensity–hue–saturation (IHS) image were each tested. Final vegetation classification was done through integration of information from both digital images and LiDAR data to evaluate the improvement in classification accuracy. Among the land cover schedules with three and eight classes of vegetation, classification from the multispectral imagery, specifically the hybrid colour composite image, had the highest accuracy, peaking at 74.6 percent and 59.4 percent, respectively. In contrast, the LiDAR classification schedules led to an average classification accuracy of 64.8 percent and 52.3 percent, respectively, for the general and detailed vegetation data. Subsequent integration of the LiDAR and digital image classification schedules resulted in accuracy improvements of 16 to 20 percent, with a superior final accuracy of 91 percent and 80.3 percent, respectively, for the three and eight classes of vegetation. A final land cover map including 8 classes of vegetation, fresh and saline water, as well as bare ground, was created for the study area with an overall accuracy of 83.9 percent, highlighting the benefit of integrating LiDAR and multispectral imagery for enhanced vegetation classification in heterogenous rangeland environments. This study comprehensively tested the suitability of LiDAR and multispectral imagery for rangeland application employing only a user-driven pixel-based classification technique.

2.5 Review of Data Fusion Techniques

In the previous chapter, it has already been mentioned that fusion can be achieved at three processing levels. Pixel- and feature-level fusions are the most prominent of them. Pixel-level fusion starts at the lowest processing (pixel) level whereas the feature-level fusion begins at the intermediate level after grouping the pixels into meaningful areas.

Pixel-level fusion has the potential to achieve the greatest performance with the highest computational cost. Feature-level fusion depends on the performance of the segmentation process and its computational cost is the lowest due to the number of features being much smaller than the number of pixels. The following sections review different pixel- and feature-level fusion techniques.

2.6 Pixel-level Fusion

Pixel-level fusion uses data from different sources at the pixel or resolution-cell level to perform detection and discrimination functions. One of the most important aspects of pixel-level fusion is that it relies heavily on the accuracy of the registration of the data sources being fused (Dai and Khorram, 1998). Misregistration errors cause artificial colours in pixel-level fusion, which falsify the interpretation later on. There are numerous different ways to achieve pixel-level fusion and highlight different aspects of the images in the final fused dataset. Some of the pixel-level fusions are reviewed in the following.

2.6.1 Colour related fusion models

There are a variety of techniques to display multi-source image data in colour. The following sections described most of those.

RGB coding based fusion

Input data are one high spatial resolution image and three coarse spatial resolution ones of higher spectral resolution. This method is mainly useful in the case of fusion between multispectral data and another image. After sampling all images at the same size, the high spatial resolution image replaces the coarse spatial resolution image having the closest spectral content. This is a very primitive way of fusing

multispectral image and is not very efficient with high spatial resolution imagery. This method is not used anymore due to the availability of more efficient methods (Laporterie and Flouzat, 2003).

RGB to IHS based fusion

The Red-Green-Blue (RGB) to Intensity-Hue-Saturation (IHS) colour transformation is common in colour image related fusion (i.e. conventional aerial photography). In an RGB image, three different images are assigned to red, green and blue channels of a colour image, giving a colour composite image. The RGB to IHS colour transformation effectively separates spatial (I) and spectral (H, S) information from the composite RGB image.

The IHS coding can be defined in several ways and so, the transform of RGB to IHS can be processed in several ways (Schetselaar, 1998). One definition can be as follows:

$$\begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.1)$$

$$H = \frac{v_1}{v_2} \quad (2.2)$$

$$S = \sqrt{v_1^2 + v_2^2} \quad (2.3)$$

The mathematical context is expressed by equation 2.1 to 2.3. I relates to the intensity, while v_1 and v_2 represent intermediate variables that are needed in the transformation. H and S stand for Hue and Saturation (Harrison and Jupp, 1990).

The RGB to IHS technique has become a standard procedure in image analysis (Pohl and van-Genderen, 1998). It has undergone several improvements enabling the fusion of more or less than three multispectral bands (Bethune *et al.*, 1998) or the

adjustment of the high resolution and intensity dynamics and or saturation dynamics (Carper *et al.*, 1990). Adjustment can also be done before the RGB to IHS transformation (Paradella *et al.*, 1997).

RGB to IHS transformation enables colour enhancement of highly correlated data (Gillespie *et al.*, 1986), the improvement of spatial resolution (Carper *et al.*, 1990; Welch and Ehlers, 1987) and the fusion of disparate data sets (Ehlers, 1991; Harris *et al.*, 1990).

The main drawback of the colour related methods is that they only allow for a limited number of input bands to be fused. Another issue is that these methods only enhance the colour composition therefore fused images are only suitable for visual interpretation but not for computer generated classification.

2.6.2 Numerical fusion models

Fusion by numerical methods is simply the combination of pixel values from multiple input images using some function or formula to give a new output pixel value. It comprises summation, subtractions, multiplication, and division (ratios) as well as combination of these operations.

Arithmetic combinations

The combinations of certain bands of image can lead to image sharpening, i.e., higher spatial resolution. Common combinations are the panchromatic and multispectral image. However, the fusion of SAR can also improve the spatial structure of an image because it introduces the surface roughness to the image (Pohl and van-Genderen, 1998).

Adding and multiplication

Adding and multiplication of images are useful to enhance the contrast of the fused image. An example is the multiplication process expressed by the following equations.

$$DN_f = A(w_1DN_a + w_2DN_b) + B \quad (2.4)$$

$$DN_f = A * DN_a * DN_b + B \quad (2.5)$$

A and B are scaling factors and w_1 and w_2 weighting parameters. DN_f , DN_a and DN_b refer to digital numbers of the final fused image and the input images a and b , respectively. A large number of publications contain suggestions on how to fuse high spatial resolution panchromatic images with low spatial resolution multispectral data to obtain high spatial resolution multispectral imagery. The choice of weighting and scaling factors may improve the resulting images. Details can be found in Welch and Ehlers (1987), Carper *et al.*, (1990) and Ehlers (1991).

Difference and ratio

Difference or ratio images are very suitable for change detection. The ratio method is even more useful because of its capability to emphasise the slight signature variations (Singh, 1989). In some cases the difference image contains negative values, therefore a constant has to be added to produce positive digital numbers.

The Brovey transform is an example of ratio fusion methods. It is named after its author and uses ratios to fuse the multispectral image (Pohl and van-Genderen, 1998). It is a formula that normalises multispectral bands used for a RGB display, then multiplies the results by the panchromatic image to add intensity or brightness to the image. The algorithm can be expressed as:

$$DN_{fusedMS_i} = \frac{DN_{bi}}{DN_{b1} + DN_{b2} + \dots + DN_{bn}} DN_{PAN'} \quad (2.6)$$

where $DN_{fusedMS_i}$ means the DN_b of the resulting fused image produced from the input data in n multispectral bands multiplied by the high-resolution panchromatic image $DN_{PAN'}$.

The numerical methods provide excellent contrast in the image domain but affect the spectral characteristics a great deal. The fused image derived from the numerical method is not suitable for further classification as the pixel values are changed drastically.

2.6.3 Statistical fusion models

In this section all operations that deal with statistical combinations of image bands are reviewed. They comprise unmixing, principal component analysis, regression and Markov Random Field based fusions.

Unmixing based fusion

This approach is widely used to improve the quality of information in wide field of view images for the joint use of wide field and ground truth data (Laporterie and Flouzat, 2003). The unmixing can be performed in two ways: spectral (Van-Der-Meer, 1999; Hu *et al.*, 2001) and spatial (Minghell-Roman *et al.*, 2001; Cherchali *et al.*, 2000).

Mezneda *et al.* (2010) developed methodology based on the linear spectral unmixing approach which was applied to ETM+ as well as ASTER VNIR and SWIR data for the mineral detection. The approach was organised in two steps; a coarse cartography based on the Landsat ETM+ spectral unmixing using image derived endmember and a detailed cartography based on a multispectral inter-images fusion using a simplified version of Multisensor Multiresolution Technique (MMT). The classification of hybrid multispectral data, which was based on the constrained linear spectral unmixing generated mineral detailed maps. The results showed that the fusion of Landsat ETM+ and ASTER SWIR multispectral image yielded the best mineral detection. The unmixing methods are commonly used to retrieve temporal reflectance profiles combining high spatial resolution images with high temporal frequency ones.

Principal Component Analysis (PCA)

The multi-source data can be accommodated by constructing a vector space with as many axes or dimensions as the components associated with each pixel. PCA is a statistical technique that transforms a multivariate dataset of correlated variables into a dataset of new uncorrelated linear combinations of the original variables. The approach for the computation of the principal components (PCs) comprises the calculation of covariance/correlation matrices and eigenvectors. Shettigara (1992) put forward the idea of Principal Component Substitution (PCS) to increase the

spatial resolution of a multi-channel image by introducing an image with higher spatial resolution. The channel that replaces the first principal component (PC1) is stretched to match the mean and variance of PC1. The higher resolution image replaces PC1 since this component contains the information common to all bands (Chavez *et al.*, 1991).

Richards and Jia (2005) used PCA for identifying bush fire damage areas from multi-temporal Landsat images. Initially coregistered each dated subscenes of 4-band data (total 8-band for two dates) were used for generating PCA. Automatic polarisation and scaling options were chosen in the transformation process as these gave component images with better visual dynamic range. The first four components were selected for further analysis as they contained most of the variances. The first component is tantamount to a total brightness image, whereas the later components highlight changes. It was the second, third and fourth components that were most striking in relation to the fire features of interest. An initial unsupervised classification of the first four principal components produced substantial confusion between water/land and fire burn/vegetation due to the nature of the first component which included all the brightness values. A second test using components 2, 3, and 4 was acceptable, although some of the richly revegetated regions were unclassified. Consequently, it was decided to use just components 3 and 4 in the classification since a visual inspection indicated that they contained most of the class/change information. The final classification map showed that there was no confusion between burn and revegetating pixels, and water edge regions. The reason for this is that the water edge pixels are approximately constant between dates and thus are correlated. Their map therefore to the midgrey constant background region of the higher order principal components. The fire burn pixels are vegetated in one data and burned in another and thus quite a different range of brightness can be found in transformed imagery.

The disadvantage of PCA approach is that it produces spatial distortion in the fused image. Spectral distortion becomes a negative factor for further processing, such as classification. Another difficulty of PCA is the interpretation of the colour composition from the three principal components, as a result interesting information may be lost in the other components (Chavez and Kwarteng, 1989).

Markov Random Field (MRF)

Markov random field theory is a branch of probability theory for analysing the spatial or contextual dependencies of physical phenomena. It is used in visual labelling to establish probabilistic distribution of interacting labels (Li, 2009). Traditionally, MRF models have been used in various fields ranging from statistical physics (Winkler, 1995; Bremaud, 1999) to remote sensing. The original work by Geman and Geman (1984) on MRF-based statistical methodology, has inspired a continuous stream of remote sensing researchers to employ the MRF model for a variety of image fusion analysis tasks (Nishii, 2003; Bruzzone and Prieto, 2000; Tso and Mather, 1999).

Solberg *et al.* (1996) developed MRF-based algorithms for image classification and change detection using multi-source data. A significant increase in classification accuracy was obtained using a MRF-based classification algorithm over other approaches. Kasetkasema *et al.* (2005) introduced a MRF model-based approach for super-resolution land cover mapping from multi-source remotely sensed imagery. This approach was based on an optimisation algorithm whereby raw coarse resolution images were first used to generate an initial sub-pixel classification that was then iteratively refined to accurately characterise the spatial dependence between the class proportions of the neighbouring pixels. Thus, spatial relations within and between pixels were considered throughout the generation process of the super-resolution map.

The drawback of employing MRF model is that it favours a more homogenous super-resolution map than the isolated pixels. Therefore, it may result in the loss of small targets of interest.

Linear regression

Multiple-regression derives a variable as a linear function of multi-variable data that has maximum correlation with univariate data. In image fusion the regression procedure is used to determine a linear combination (replacement vector) of an image band that can be replaced by another image band (Pohl and van-Genderen, 1998). Alternatively, an image is expressed according to another one by linear regression (Price, 1987). In this fusion process, the high spatial resolution band is

sampled from coarse spatial resolution and then local linear regression is used to extract high resolution data from coarse the resolution image (Laporterie and Flouzat, 2003). Wal and Herman (2007) applied a regression-based synergy of optical, short-wave infrared and microwave remote sensing for monitoring the grain-size of intertidal sediments.

A disadvantage of the regression method is that only reflectance from a limited number of wavelengths (bands) can be used to obtain significant regression models (as many bands are highly correlated with each other), and therefore, not all information in multi-source data is used (van-der-Wal and Herman, 2007).

2.6.4 Signal-processing-based Fusion Models

In this section, the signal processing related fusion techniques are reviewed. This type of fusion includes filtering, wavelet and pyramid algorithms.

High-pass filter

The basic principle consists in filtering the high spatial resolution band by a high-pass filter and then injecting the high frequencies in the coarse spatial resolution images. This method was first developed by Schowengert (1980). At first sight, it seems very natural. Indeed, the main objective of multi-resolution fusion is to inject high spatial frequencies in coarse resolution data. However, in remote sensing, the spectral mixing of data induces a decrease in spectral quality. That's why, Chavez (1988) suggests to decrease the content of the high frequencies extracted to limit the contribution of the high resolution band.

Wavelets

The wavelet transform is a multi-scale (multi-resolution) approach well suited to manage different image resolutions. The key step in image fusion based on wavelets is that of coefficient combination, namely, the processes of merging the coefficients in an appropriate way in order to obtain the best quality in the fused image. This can be achieved by a set of strategies. The most simple is to take the average of the coefficients to be merged, but there are other merging strategies with better performances. Pajares and Cruz (2004) reviewed three different wavelet

decomposition techniques: multi-scale decomposition (Zhang and Blum, 1999), the ARSIS (Amelioration de la Resolution Spatiale par Injection de Structures) concept (Ranchin and Wald, 2000) and a multi-sensor scheme (Li *et al.*, 1995). Zhang and Blum (1999) established a categorisation of multi-scale decomposition. The objective of this fusion is to achieve a high quality image from several degraded images. Ranchin and Wald (2000) introduced the ARSIS concept for making use of a multi-scale method for the description and modelling of the missing information between images to be fused. In ARSIS, the goal of the fusion is to achieve high spatial resolution together with a high-quality spectral content from two kinds of remote sensing images: images with high quality in the spectral content but low quality in the spatial resolution, and images with high spatial resolution but with a unique spectral band. Li *et al.*(1995) performed extensive experiments with several sets of images from multi-sensors including the fusion of multi-focus images.

In the Pajares and Cruz (2004) review, it was found that the wavelet-based methods achieve similar results compared to the classical methods such as principal component analysis (PCA), intensity-hue-saturation (IHS) or Brovey transformation. Nevertheless, their worst reported performance was due to the presence of a high number of edges in the source images. When the images were smooth, without abrupt intensity changes, the wavelets worked appropriately, improving the results over classical methods. The main drawback of the wavelet-based method is found in the decomposition level for multi-resolution approaches. Indeed, a decomposition of 6 requires that the source images have sizes greater than 2048×2048 pixels, so that the merging can be carried out to a resolution of 32×32 in wavelets with decomposition level of 6, otherwise it is ineffective. In classical approaches, with decomposition level of 4, the sizes can be of 512×512 pixels.

A well-known drawback of the wavelet transformation is the shift dependency, such as a simple shift of the input signal, may lead to completely different transform coefficients. When applying wavelet transformation for the decomposition of images, only a certain number of scales (levels) can be accommodated (Muhammad *et al.*, 2002). To overcome the shift dependency of the wavelet fusion scheme, the input images must be decomposed into a shift invariable representation. The main

drawback of the wavelet-based fusions is that they are more computationally complex and often require the user to determine appropriate values for certain parameters such as threshold (Amolins *et al.*, 2007). On another front, during signal processing the original data is not preserved; therefore it is hard to quantify the accuracy of the fused products (Phen-Lan and Po-Ying, 2008).

Pyramid algorithms

Pyramid algorithms utilise decomposition techniques to fuse images into coarse resolution and then extract data from recomposed images from the coarse resolution sensors. Pyramid algorithms can be divided into two distinct groups: linear decomposition (Aiazzi *et al.*, 1999) and non-linear decomposition (Pavel and Sharma, 1996). Morphological pyramid is one of the non-linear decomposition techniques (Laporterie and Flouzat, 2003). The morphological pyramid concept combines theories of multi-resolution analysis and mathematical morphology enabling decomposition of any image at several resolutions and recomposing it exactly from them (Flouzat *et al.*, 2001). The morphological pyramid is composed of two parts: the decomposition and the recomposition and can be considered as moving in the spatial domain.

The morphological pyramid has several advantages. Firstly, all the pyramid process can be parameterised. Enabling choice of well-fitted parameters if the user has *a priori* knowledge. Secondly, the decomposition is efficient with regards to the structure of natural landscapes, enabling easy extraction of characteristics objects. Within this process, multi-resolution fusion is feasible whatever the resolution ratio between images. Lastly, this method is particularly efficient for updating landscape information by fusing coarse spatial resolution images at high temporal frequency with high spatial resolution images at low temporal frequency. In this way, it is possible to build a new composite image of great interest when two spatial components have two different behaviours (Laporterie and Flouzat, 2003).

2.6.5 Classification-based fusion models

The above-described fusions can only give half the answer for aerial imagery with LiDAR data fusion for landscape mapping. These algorithms can produce fused

images rich in spectral and spatial resolution. However, the classification process (manual or automated) is quite separate. Generally, these algorithms are very good for fusing high spatial resolution panchromatic imagery with low spatial resolution multispectral imagery to produce high spatial and high spectral resolution fused images. However, the fusion of multi-sensor data from active and passive sensors does not enhance images much through these processes, therefore classification after the fusions is difficult. Consequently, an approach considering classification is needed to fully (or semi-) automate the mapping process. The classification-based fusion techniques fall into this category: they not only enhance the spectral and spatial content of the image but also delineate the object of interest from multi-source data.

A simple approach to deal with a pixel-level fusion problem is to extend the dimension of the data vectors to include various spectral and non-spectral data in the classification process. This approach is known as the stacked vector or augmented vector method (Tso and Mather, 2001; Vijayaraj, 2004; Richards and Jia, 2005). The stacked vector can be described by the following equation.

$$X = [x_1^t, x_2^t, \dots, x_s^t]^t \quad (2.7)$$

Where s is the total number of individual data sources with corresponding data vectors x_1, \dots, x_s , and the superscript t denote a vector transpose operation. The stacked vector X can, in principle, be used in standard classification techniques. However, some prerequisites need to be fulfilled before applying any classification techniques on the stacked vector (Schistad-Solberg *et al.*, 1994). They are: (1) various sources must be described by a common spectral model, (2) the sources ideally be acquired at the same time, and (3) have no changes in the pattern classes between the acquisition dates. This approach is certainly not valid for fusing imagery with other types of spatial data, e.g., topographic maps.

If the prerequisites are fulfilled, the data vector can be used as an input to any standard pixel-based classification techniques. Usually, classification is performed with a set of target classes in mind. A review of classification schemes has been presented in Section 2.2. The best classification scheme includes classes that are both

important to this research and discernible from the data sources. Traditionally, the classification techniques are divided into two broad categories: data-driven and user-driven techniques. The data-driven techniques are dependent on the data itself and used when less is known about data before classification (Jensen, 1996). *Unsupervised classification* technique is one example of this category. In contrast, the users closely control the user-driven techniques or user derived rules. *Supervised classification* falls into this category.

Unsupervised ISODATA approach

Unsupervised classification techniques detect clusters of pixels in feature space and categorise the pixels in the clusters based on the statistical patterns inherent in the data. Clusters are defined with a clustering algorithm that often uses all or many of the pixels in the input data file for its analysis. ISODATA presented by Ball and Hall (1965) is the most popular unsupervised classification algorithms.

An ISODATA algorithm is based on estimating some reasonable assignment of the pixel vectors into candidate clusters. They are then reclustered in such a way that the sum of squared errors (SSE) is reduced. The ISODATA algorithm is implemented by the following set of basic steps (Richards and Jia, 2005; 1993):

- (a) the procedure is initialised by selecting a number of clusters (C) in the multi-source feature space to serve as candidate cluster centres. Let these be called

$$m_i, i = \hat{m}_i \quad (2.8)$$

The selection of the \hat{m}_i at this stage is arbitrary with the exception that no two clusters may be the same. To avoid anomalous cluster generation with unusual data sets it is generally wise to spread the initial cluster means uniformly over the data. This can also serve to enhance convergence. Besides choosing the \hat{m}_i , the number of cluster C is specified beforehand by the user

- (b) the location x of each pixel in the cluster is examined and the pixel is assigned to the nearest candidate cluster. This assignment would be made on the basis of the Euclidean distance measure between clusters
- (c) The new set of means that result from the grouping in step (b) are computed. Let these be denoted

$$m_i, i = 1, \dots, C. \quad (2.9)$$

- (d) If $m_i, i = \hat{m}_i$ for all i , the procedure is terminated. Otherwise \hat{m}_i is redefined as the current value of m_i and the procedure return to step (b).
- (e) Once clustering is completed, or at any suitable intervening stage, the clusters can be examined to see whether any clusters contain so few points as to be meaningless (e.g. they would not allow acceptable statistics estimates), or some clusters are so close together that they represent an unnecessary or indeed an injudicious division of the data so should be merged.

Merging and splitting options are employed at the end of the iterations leading ultimately to the final clusters. The merging procedure commences by assuming that all initial clusters are individual classes, it then systematically merges neighbouring clusters by checking statistical attributes and visual interpretation. The clusters' mean separation are checked and merging starts with the shortest mean separations. This merging and splitting method starts with a large number of clusters and fuses them progressively into a small number of clusters. The agglomeration is continued until all clusters are accumulated in background classes as well as object classes.

Haala and Brenner (1999) applied an unsupervised classification algorithm to fuse multispectral imagery with LiDAR data for the extraction of buildings, trees and grass-covered areas. The basic idea of the proposed algorithm was to use geometric and radiometric information of the multispectral imagery simultaneously with nDSM derived from the LiDAR data. In this study, scanned images with near infrared, red and green spectral bands were used at a 30cm pixel footprint. For LiDAR data, terrain points were measured at approximately one point per square meter with an accuracy of 0.2m. The height data and the image were co-registered and brought into

the same spatial resolution of 0.5m. The applied unsupervised classification detects clusters of pixels in feature space and categorizes the pixels to the clusters based on the minimum distance criterion. For that purpose the ISODATA algorithm was utilized. With this approach, the optimal number of spectral clusters is automatically determined by iteratively applying split and merge operations.

The most problematic task of the unsupervised classification is to separate different classes through an automatic interpretation. Haala and Brenner (1999) stated that the separation of trees from buildings was a difficult task due to the relatively low spatial resolution of the LiDAR data. As an alternative of using all bands of multispectral images, the NDVI determined from the near infrared and the red portions of the spectrum can be used due to its potential in discriminating vegetation (Lu and Trinder, 2003).

An unsupervised approach is dependent upon the data itself for the definition of classes. This method is usually used when less is known about the data fusion. It is then the User's responsibility, after fusion, to attach meaning to the resulting classes (Jensen, 1996). The unsupervised approach is useful only if the classes can be appropriately interpreted.

Supervised approach

Supervised classification is the procedure most often used for quantitative analysis of multi-source remote sensing data. It is up to the user to define the training classes as *a priori* (before the fact) knowledge and using suitable algorithms to label the pixels from multi-source images as representing those classes. A variety of algorithms is available for this, ranging from those based on probability distribution models to those which partition class-specific regions using optimally located surfaces. According to Richards and Jia (2005), irrespective of the particular method chosen, the essential practical steps are as follows:

- (a) decide the set of landscape classes into which the multi-source images are to be classified. The landscape classification schemes, which were presented in Section 2.2, are generally used to derive these classes

- (b) choose representative or prototype pixels from each of the desired set of landscape classes. These pixels are said to form *training data*. Training sets for each class can be established using site visits, maps and aerial photographs. Often the training pixels for a given class will lie in a common region enclosed by a border. That region is often called a *training field* (Richards and Jia, 2005)
- (c) using the training data, the parameters of a particular classifier algorithm are estimated. These parameters are the properties of the probability model that are used to define the partitions in feature space. The set of parameters for a given class is sometimes called the *signature* of that class (Richards and Jia, 2005)
- (d) using the trained classifier, label or classify every pixel in the multi-source images into one of the desired landscape classes. Here the whole scene of the multi-source images is typically classified
- (e) produce tabular summaries or thematic maps, which summarise the results of the classification
- (f) assess the accuracy of the final product using a labelled testing data set

A range of algorithms could be used in steps (c) and (d). These algorithms can broadly be divided into parametric and nonparametric domains (ERDAS, 2002). A parametric decision rule is trained by the parametric signatures. These signatures are defined by the mean vector and covariance matrix for the data file values for the pixels in the signatures. When a parametric decision rule is used, every pixel is assigned to a class since the parametric decision space is continuous. Minimum distance and maximum likelihood are the common parametric classifiers. A nonparametric decision rule is not based on statistics; therefore, it is independent of the properties of the data. If a pixel is located within the boundary of a nonparametric signature, then this decision rule assigns the pixel to the signature's class. Basically, a nonparametric decision rule determines whether or not the pixel is located inside the nonparametric signature boundary. Parallelepiped classifier is one of the common nonparametric decision rules.

Parametric decision rule-Minimum distance classifier

The minimum distance decision rule (also called spectral distance rule) is based on spectral distance between the measurement vector for the candidate pixel and the mean vector for each signature. The equation for classifying by spectral distance is based on the equation for Euclidean distance (Eq. 2.2).

$$SD_{xyc} = \sqrt{\sum_{i=1}^n (\mu_{ci} - X_{xyi})^2} \quad (2.10)$$

where 'n' is the number of bands or dimensions, 'i' is a particular band, 'c' is a particular class, X_{xyi} is the data file value of pixel x, y, in band 'i', μ_{ci} is the mean of data file values in band 'i' for the sample for class 'c' and SD_{xyc} is the spectral distance from pixel x, y, to the mean of class 'c'. When spectral distance is computed for all possible values of c (all possible classes), the class of the candidate pixel is assigned to the class to which SD is the lowest. The main drawback of this classifier is that it does not take into account other statistics; as a result, every pixel is classified although some should be unclassified due to the large spectral distance to the mean of the sample.

Parametric decision rule-Maximum likelihood classifier

The maximum likelihood decision rule is based on the probability that a pixel belongs to a particular class. The basic equation assumes that these probabilities are equal for all classes, and that the input data have normal distributions. If the user has *a priori* knowledge that the probabilities are not equal for all classes, then weighting factors can be specified for particular classes. This variation of the maximum likelihood decision rule is known as the Bayesian rule (Hord, 1982). The equation for the maximum likelihood/Bayesian classifier is as follows:

$$D = \ln(a_c) - [0.5 \ln(|Cov_c|)] - [0.5(X - M_c)^T (Cov_c^{-1})(X - M_c)] \quad (2.11)$$

where D is the likelihood weighted distance, c is a particular class, X is the measurement vector of the candidate pixel, M_c is the mean vector of the sample of class c , a_c is the percent probability that any candidate pixel is a member of class c

(defaults to 1.0 or is entered from *a priori* knowledge). Cov_c is the covariance matrix of the pixels in the sample of class c , $|Cov_c|$ is the determinant of Cov_c , Cov_c^{-1} is inverse of Cov_c , \ln is the natural logarithm function and T is the transposition function (matrix algebra).

The pixel is assigned to the class for which the weighted distance, D , is the lowest. This classifier is very popular and more accurate than other classifiers if the training datasets have a normal distribution. This is often referred to as a ‘hard’ classification because a pixel is assigned to only one class, even though the sensor system records radiant flux from a mixture of biophysical materials within the IFOV (Foody *et al.*, 1992).

The maximum likelihood classifier has been widely used to fuse multi-source images. Mass (1999) applies a supervised maximum likelihood classifier to a LiDAR-derived DSM to extract different land cover features. Firstly, the original height data was used for discrimination between high objects such as buildings and trees on the one hand, and objects like streets and plain ground on the other. Secondly, a texture layer was prepared by applying a Laplace filter to the range (height) data. This emphasized edges or noise and thus delivered large values for vegetation, while flat roof faces had low or zero values. Finally, the slope around each pixel was determined from the local slopes in X and Y. The use of a slope image was valuable for distinguishing tilted roofs from flat roofs or streets as well as from trees. These three generated layers were used as inputs to a maximum likelihood classification, which was initialised by sparse training regions containing the object classes ‘flat roof’, ‘tilted roof’, ‘vegetation’, ‘flat terrain’, and ‘no data’. The author suggested that this process could be altered if *a priori* knowledge on building heights, roof slopes, tree height etc. is available.

Schistad-Solberg *et al.* (1994) applied a multi-source data fusion framework based on a Bayesian formula (maximum likelihood). They fused Landsat TM images and ERS-1 SAR images using a maximum likelihood logic. In this process, they pre-processed the SAR images in order to reduce noise, and included additional texture features that can help classification. The classification model included *a priori* information in various forms. This information was about sensor-specific noise

characteristics or information about the weather conditions at the time the images were captured. The proposed model gave significant improvement in the classification error rates compared to the conventional single-source classifiers. Overall classification accuracy of 95 percent for Landsat and 65-70 percent for SAR was achieved. The error rates using a non-contextual classifier were typically 3-5 percent higher. The fusion of Landsat image with the SAR image only reduced the error rate by 0.3 - 4.9 percent. The benefits of fusion were more clearly demonstrated when fusing the Landsat image with the three SAR images captured at different dates and taking into account the changes in the agriculture areas.

The main criticism of parametric classification, such as maximum likelihood, is that this may not be suitable when the various sources cannot be described by a common 'spectral' model especially when spectral and elevation data are fused (Schistad-Solberg *et al.*, 1994). For instance, neither the relative heights nor the spectral characteristics of landscape features can be assumed to be normally distributed. Features have different heights and spectral 'colour', so that they correspond to more than one cluster in feature space.

Nonparametric decision rule-Parallelepiped classifier

In parallelepiped decision rules, the data file values of the candidate pixel are compared to the upper and lower limits. These limits can be either: (1) the minimum and maximum data file values of each band in the signature, (2) the mean of each band, plus and minus of a number of standard deviations, or (3) any limits that a user specifies, based on knowledge of the data and signatures.

There are high and low limits for every signature in every band. When training data values are between the limits of a signature in every band, then the pixel is assigned to that signature's class. In cases where a pixel may fall into the overlap region, the user must define how the pixel should be classified. It may be classified by the order of the signatures or by the defined parametric decision rule, or it can be left unclassified. This classifier is often useful for a first-pass, broad classification when the parallelepiped decision rule quickly narrows down the number of possible classes to which each pixel can be assigned before the more time-consuming algorithms such as minimum distance and maximum likelihood classifiers are used.

Meyer *et al.* (1996) presented a comparison between parallelepiped and maximum likelihood classification techniques for identification of forest species from large-scale colour-infrared aerial photography. A semi-automated classification procedure was introduced and different combinations of the colour-infrared aerial photography were used. In all cases, the parallelepiped classification technique gave better results than the maximum likelihood classification technique. The accuracy was low for both of these classifications of a three-band data set. However, classification results were improved for the six new additional features generated from principal component transformation of original three-band data set. An average classification accuracy of 81 percent was achieved for the class set studied with the parallelepiped classification technique. The relative poor performance of the maximum likelihood classification technique is mainly due to the similar relationship of the class-related density function in the original and transformed data. In addition, the maximum likelihood classifier is only conditionally recommended for more than four to five band data (Lindenberger, 1973).

In pixel-level fusion, supervised classification is the procedure most often used for fusing multi-source data in stacked vectors. This presents a number of difficulties if parametric methods such as maximum likelihood classification are considered. These include incompatible statistics of the disparate data types, with some data unable to be represented by normal class models, and computational cost increasing with data dimensionality. A nonparametric classification such as parallelepiped classification technique could be an appropriate algorithm to adopt since it depends only on the application of threshold components of the multi-source data vector (Richards and Jia, 2005).

2.7 Feature-level Fusion

In the previous section, a detailed description of different pixel-level fusion procedures was given. Now, feature-level fusion is presented as part of the Object Based Image Analysis (OBIA) concept.

The idea of OBIA has been around since the early 1970s (de-Kok *et al.*, 1999), but implementation lagged due to lack of computing power. On a limited basis,

specialized object-oriented software packages were employed in the 1980s to extract roads and other linear features (McKeown, 1988; Quegan *et al.*, 1988). Since then, object-based classification techniques have developed considerably, with three main factors instrumental to their growing availability and use. Firstly, the sophistication of GIS technology, particularly raster/vector integration capabilities, has increased (Cowen *et al.*, 1995). Image pixels combined with vectors provides a relatively straightforward approach to object-based classification (Janssen and Molenaar, 1995). Secondly, a dedicated object-based image analysis system, known as eCognition was commercially released in the early 2000s (Flanders *et al.*, 2003; Platt and Rapoza, 2008). Finally, high spatial resolution imagery has become common, prompting a new emphasis on object-oriented techniques (Franklin *et al.*, 2003).

The term OBIA has recently been regarded as being too broad, as this term may include image analysis in disciplines such as computer vision and biomedical image (Blaschke, 2010). Hay and Castilla (2008) define Geospatial Object Based Image Analysis (GEOBIA) as a sub-discipline of Geographic Information Science, devoted to developing automated analysis of remotely sensed data and produced GIS-ready spatial information (Johansen *et al.*, 2010). The feature-level fusions in this study use the GEOBIA concepts specifically for the synergy of multispectral imagery and LiDAR data for landscape object extraction.

GEOBIA builds on the traditional segmentation, edge-detection, object extraction and classification concepts of remote sensing image analysis and the idea of incorporating contextual information in classification (Haralick and Shapiro, 1985; Kettig and Landgrebe, 1976). These approaches all recognise the hierarchical nature of objects in the environment and that context and pattern is critical for effective image analysis. Despite the initial foundation for GEOBIA being laid in the 1970s, it was not until the evolution of high spatial resolution digital imagery that a significant niche for GEOBIA was realised (Johansen *et al.*, 2010). A variety of software packages focusing on GEOBIA have been developed since the launch of eCognition in 2000 (Marpu *et al.*, 2010). Among them ENVI's Feature Extraction module (ITT, 2011) and Feature Analyst (VLS, 2011) are the most popular.

GEOBIA for feature-level fusion starts with the grouping of neighbouring pixels into meaningful regions, a process known as *segmentation* (Haralick, 1983; Haralick and Shapiro, 1985; Pal and Pal, 1993; Kartikeyan *et al.*, 1998). The elementary picture units are no longer the pixels, but connected sets of pixels. Once the image has been segmented, fusion of the features are performed using divergent spectral values as well as the additional spatial information (Blaschke and Strobl, 2001; Darwish *et al.*, 2003; Flanders *et al.*, 2003; Benz *et al.*, 2004; van-der-Werf and van-der-Meer, 2008; Hay and Castilla, 2008). Image segmentation is the essential pre-cursor to object-based image analysis for the fusion of multi-source data (Hay and Castilla, 2008; Lang, 2008).

2.7.1 Segmentation

In mathematical terms, a segmentation of an image f is a partition of its definition domain D_f into n disjoint nonempty sets X_1, X_2, \dots, X_n called segments such that the union of all segments equals D_f . Usually, an image that has been segmented is represented as a label image where each segment X_i is given a value different from all other segments.

The design of an algorithm for segmenting an image into meaningful features requires some understanding of the image objects and their properties such as shape, size, orientation, grey level distribution and texture. Ideally, these properties allow to discriminate between different image objects (Soille, 2003). Segmentation accuracy determines the eventual success or failure of the further fusion analysis procedures.

Image segmentation can be achieved using data- or user-driven approaches (Wealands *et al.*, 2005). Morphological approaches (Soille, 2003) are gaining importance due to their unique characteristics. In the following sections, a review of different segmentation techniques is presented.

Data-driven segmentation

In data-driven approaches, the image is segmented into regions based on some additional information. Some of the common methods used for data-driven segmentation are edge detection, thresholding, region growing, and split and merge.

Edge detection

Edge-based segmentation attempts to find segment boundaries by detecting edges between image areas with discontinuities in intensity levels. As image objects are assumed to show little variation in grey levels, their edges are characterised by high grey level variations in their neighbourhood. The task of edge detection is to enhance and detect these variations (Soille, 2003). Since edges are local features, they are determined based on local information. A large variety of methods are available in the literature for edge finding (Gonzalez and Woods, 2002; Rosenfeld and Kak, 1982; Hall, 1979). Davis (Fu and Mui, 1981; Davis, 1975) classified edge detection techniques into two categories: sequential and parallel. In the sequential technique the decision whether a pixel is an edge pixel or not is dependent on the result of the detector at some previously examined pixels. On the other hand, in the parallel method the decision whether a point is an edge or not is made based on the point under consideration and some of its neighbouring points. As a result of this the operator can be applied to every point in the image simultaneously. The performance of a sequential edge detection method is dependent on the choice of an appropriate starting point and how the results of previous points influence the selection and result of the next point (Pal and Pal, 1993).

There are different types of parallel differential operators such as Roberts gradient, Sobel gradient, Prewitt gradient and the Laplacian operators (Gonzalez and Woods, 2002; Pal and Pal, 1993). These difference operators respond to changes in grey level or average grey level. The gradient operators, not only respond to edges but also to isolated points.

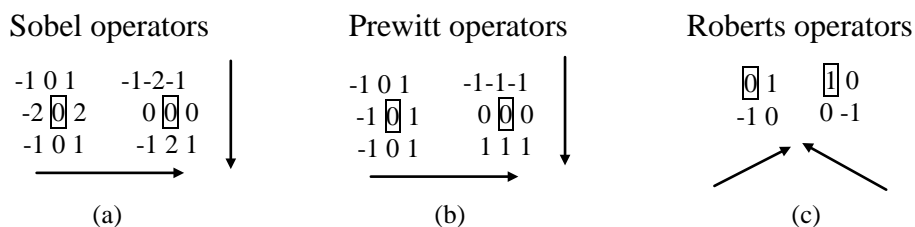


Figure 2.3 Illustration of the gradient operators and their components

The Sobel edge detector uses the Sobel approximation masks in Figure 2.3(a) to approximate the first derivatives. In other words, the Sobel detector computes the

gradient at the centre point in a neighbourhood. The Prewitt edge detector uses the Prewitt approximation masks in Figure 2.3(b) to approximate digitally the first derivatives. The parameters of this function are identical to the Sobel parameters. The Prewitt detector is slightly simpler to implement computationally than the Sobel detector, but it tends to produce somewhat noisier results. The Roberts edge detector uses the Roberts approximation masks in Figure 2.3(c) to approximate digitally the first derivatives. The parameters of this function are identical to the Sobel parameters. The Roberts detector is one of the oldest edge detectors in digital image processing and it is also the simplest (Gonzalez *et al.*, 2004).

The main drawback of these operators is that the resulting edges are seldom connected. Additional processing is then required to obtain closed contours corresponding to object boundaries. As an alternative, the Canny detector (Canny, 1986) links edges by using simple hysteresis during the linking of high-gradient ridges. The Canny method finds edges by looking for local maxima of the gradient of images. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be 'fooled' by noise, and more likely to detect true weak edges.

The disadvantages of all the edge-based segmentations are:

- a. these operators can only work well on a single layer image. For segmentation, they are very good for panchromatic imagery but not useful for colour or multispectral images
- b. small terrain objects can be completely obscured by boundary pixel (Geneletti and Gorte, 2003)
- c. another challenge in edge-based segmentation is to correctly delineate the object of interest from the segmented image. All the edges produced by these operators are, normally, not significant (relevant) edges when viewed by human beings. Therefore, one needs to find out prominent (valid) edges from the output of the edge operators. Kundu and Pal (1986) have suggested a

method of thresholding to extract the prominent edges based on psycho-visual phenomena. Haddon (1988) developed a technique to derive a threshold for any edge operator, based on the noise statistics of the image

Thresholding

Thresholding is one of the old, simple and popular techniques for image segmentation. Thresholding involves defining a threshold value that is then used to distinguish each individual pixel as high or low values. This process can be done based on global information (e.g. the grey level histogram of the entire image) or it can be done using local information (e.g. the co-occurrence matrix (Pal and Pal, 1993)). Taxt *et al.* (1989) refer to the local and global information-based techniques as *contextual* and *non-contextual* methods, respectively. Under each of these schemes if only one threshold is used for the entire image then it is called *global thresholding*. On the other hand, when the image is partitioned into several subregions and a threshold is determined for each of the subregions, it is referred to as *local thresholding* (Taxt *et al.*, 1989). Some authors (Yanowitz and Bruckstein, 1989; Nakagawa and Rosenfeld, 1979; Chow and Kaneko, 1972) call these local thresholding methods *adaptive thresholding* schemes. Weszka (1978) presents a comprehensive review of the thresholding techniques including global, local and dynamic methods.

Global thresholding methods can fail when the image is noisy or the background is uneven and the illumination is poor. In such cases the objects can still be lighter or darker than the background, but any fixed threshold level for the entire image will usually fail to separate the objects from the background. This leads one to methods of local thresholding or adaptive thresholding (Pal and Pal, 1993). In local thresholding, the image is normally partitioned into several non-overlapping blocks of equal area and a threshold for each block is computed independently. Chow and Kaneko (1972) used the sub-histogram of each block to determine local threshold values for the corresponding cell centres. These local thresholds are then interpolated over the entire image to yield a threshold surface. The performance of the algorithm is likely to depend on the choice of the threshold levels.

All of the above thresholding methods have a common drawback; they take into account only the histogram information, ignoring the spatial details. As a result, such an algorithm may fail to detect threshold levels if these are not properly reflected as valleys in the histogram (Pal and Pal, 1993). Dynamic thresholding not only uses grey level information but also topological information and some other local properties. This method is designed to perform well on low quality images in which a single global threshold is inadequate due to contrast differences throughout the image (Weszka, 1978). Weszka and Rosenfeld (1978) use a 'busyness' measure which is dependent on the co-occurrence of adjacent pixels in an image. They minimized the busyness measure in order to arrive at the threshold for segmentation. Deravi and Pal (1983) applied a co-occurrence matrix containing local information to minimize the conditional probability of transition across the boundary between two regions. Since all these methods make use of the spatial details, the results are more meaningful than methods that use only the histogram information.

Over- and under-segmentation is the common problem for the thresholding. In both cases, segmented image unites are not properly represented the features, as the segments are too small or they merging the regions that do not belong to a single feature. Since threshold segmentations are very simple to implement they typically lead to results of a relatively limited quality (Baatz and Schape, 2000). The good control of threshold is the decisive factor for a meaningful segmentation. Local contrasts are not considered or not represented in a consistent way and the resulting regions can differ widely in size. Thresholding works relatively well in a single layer image but thresholding in multispectral image is not easy. Thresholding can only fulfil one criterion very well but thresholding multispectral images needs to fulfil more than one criterion. Thus visually similar looking segments in a multispectral image are not necessarily segmented well using multispectral characteristics. As Haralick and Shapiro (1985) point out, this type of segmentation is most likely to avoid errors through poor region merges; however, it does not produce spatially contiguous regions, and the salt-and-pepper effect can occur.

Region growing segmentation

In region growing segmentation, homogeneous regions are first located. The growth of these regions is based on similarity measurements combining spatial and spectral attributes (Soille, 2003). Region growing proceeds until all pixels of the image are assigned to a region (Chang and Li, 1994). Region boundaries are created when two growing regions meet.

Two factors determine the final segmentation result: the parameters for the region growing, and the location of the seed points. The parameters of this algorithm basically dictate which pixels can be absorbed into the region being grown, and subsequently when the growing must stop. The initial locations of the seed points in the image contribute to determining the size, shape and spatial distribution of the resulting features.

The selection of similarity parameters depends not only on the problem under consideration, but also on the type of image data available. In case of colour imagery, this problem would be significantly more difficult, or even impossible, to handle without the inherent information available in colour images (Gonzalez *et al.*, 2004). For the monochrome image, region growing can be carried out with a set of descriptors based on intensity levels (texture) and spatial properties. Another problem in region growing segmentation is the formulation of a stopping rule. Basically, growing a region should stop when no more pixels satisfy the parameter for inclusion in that region. Parameter such as intensity values, texture, and colour utilise the concept of size, likeness between a candidate pixel and the pixels grown so far and the shape of the region being grown (Gonzalez and Woods, 2002).

In order to ensure that useful results can be generated from many different images, the region growing segmentation can be repeated with multiple sets of seed points, and with multiple region growing parameters. It should be noted that this algorithm is not intended to segment the entire image into regions, but just to extract useful features. For region growing, a rule needs to be describing a growth mechanism and a rule checking the homogeneity of the regions after each growth step.

The initial step of the region-growing algorithm is to select appropriate seed points. Since there is no one method that automatically extracts good seed points for all

images, multiple seed point extraction algorithms have to be used. Here three different seed point generation algorithms are described.

Region growing segmentation is the simplest method of segmentation, but most prone to unwanted region merge errors (Haralick and Shapiro, 1985). This method basically depends on the set of given seed points and often suffers from a lack of control in the break off criterion for the growth of a region (Baatz and Schape, 2000). This method is sensitive to the choice of seeds and therefore is not widely used (Wealands *et al.*, 2005).

Split and merge segmentation

The split method for segmentation begins with the entire image as the initial segment. Then it successively splits each current segment into quarters if the segment is not homogenous enough. Homogeneity can be easily established by determining if the difference between the largest and smallest grey tone intensities is small enough (Haralick and Shapiro, 1985). Region splitting and merging is an alternative to the region growing procedure. Rather than growing regions from seed points, the image is progressively split into smaller and smaller regions until it cannot be split any more. Subsequently, neighbouring regions are merged with each other until no further merging can take place. The result is a segmented image in which the segments should ideally represent real world objects. Unlike the region growing segmentation, the split and merge algorithm segments the entire image. Algorithms of this type were first suggested by Robertson (1973) and Klinger (1973). Kettig and Landgrebe (1975) tried to split all non-uniform 2x2 neighbourhoods before beginning the region merging. Fukada (1980) suggested successively splitting a region into quarters until the sample variance is small. Efficiency of the split and merge method can be increased by arbitrarily partitioning the image into square regions of a user-selected size and then splitting these further if they are not homogeneous (Haralick and Shapiro, 1985).

The split and merge segmentation algorithm works by successively splitting the image into quadrants, and merging those neighbouring quadrants which are similar, until no further splitting or merging can take place. Using a square image with

dimensions of 2^n (n is an integer) means that the image can always be split into quadrants, until those quadrants have a size of 1 pixel. After the image is split into quadrants, each one of the quadrants must be examined to see if it should be split again or not. This decision will be based on the statistics of the grey values of the pixels contained within that quadrant. If those pixels are all similar to each other, then the quadrant does not require further splitting. However, if the pixels are significantly different, then the quadrant must be split into a further four quadrants.

The measure for deciding whether or not the pixels are similar is a statistical one: if all the grey levels fall within m standard deviations, then they are similar. By changing the parameter m the user has some control over the results generated by this algorithm. When it is clear which quadrants will require further splitting, and which ones will not, the merging process can take place. Any neighbouring quadrants that do not require further splitting must be compared to see if they can be combined to form a single region. The same statistics are used to define the merging process: if all the pixels within the two neighbouring regions are within m standard deviations, then they are similar, and can be merged.

Split and merge is the most complex of the above segmentations and requires more processing than the other methods, which is disadvantageous when analysing many spatial fields (Wealands *et al.*, 2005). Because split and merge segmentations successively divide regions into quarters, the boundaries produced by the split technique tend to be squarish and slightly artificial. Sometimes adjacent quadrants coming from adjacent split segments need to be joined rather than remain separate (Haralick and Shapiro, 1985). It is likely that a split and merge strategy could be devised to take care of this problem. The suggestion is to merge a pair of adjacent regions if their grey tone intensity distributions are similar enough.

Morphological watershed segmentation

Other than the above-described data-driven segmentations, a more advanced approach is the morphological segmentation, which combines edge detection and region growing techniques (Soille, 2003). Watershed transformation is one example of morphological segmentations.

Watershed transformation is a very powerful image segmentation tool provided by mathematical morphology (Soille and Vincent, 1990). Watershed segmentation, first proposed by Beucher and Lantuejoul (1979), is a well known image segmentation method that incorporates the region growing and edge detection techniques (Soille, 2003). It groups the image pixels around the regional minima of the image and the boundaries of adjacent groupings are precisely located along the crest lines of the gradient image. The concept behind the watershed algorithm is the idea that the image can be treated as if it were a 3D surface (see Fig 2.4), with lateral dimensions representing the image plane, and the vertical dimension representing the grey values of the pixels. If that surface were then flooded from below, then the regions where the water pools represent individual watersheds. As the water level rises, neighbouring watersheds will begin to merge, unless dams are built between them. The process of merging regions and building boundaries between regions continues until no more growing can take place. The result is a segmented image.

The watershed algorithm produces best results for images that have regions that are homogeneous and low intensity, separated by narrow boundaries of high intensity. Not many images have this particular distribution of grey levels, but application of a first order edge detection algorithm yields this result. Therefore, the first step in the watershed algorithm is to process the image with a first order edge detector to extract the edges. Once the edge strength image has been created, the watershed procedure can begin. If the watershed algorithm is directly applied on the edge strength image, then the result will be over-segmented. In other words, there are too many segments in the resulting image. Therefore it is necessary to use markers to guide the watershed algorithm. Markers are connected components – groups of connected pixels – that belong to an image. There are two types of markers that need to be defined: internal markers and external markers. Internal markers are associated with the objects of interest, and external markers are associated with the background.

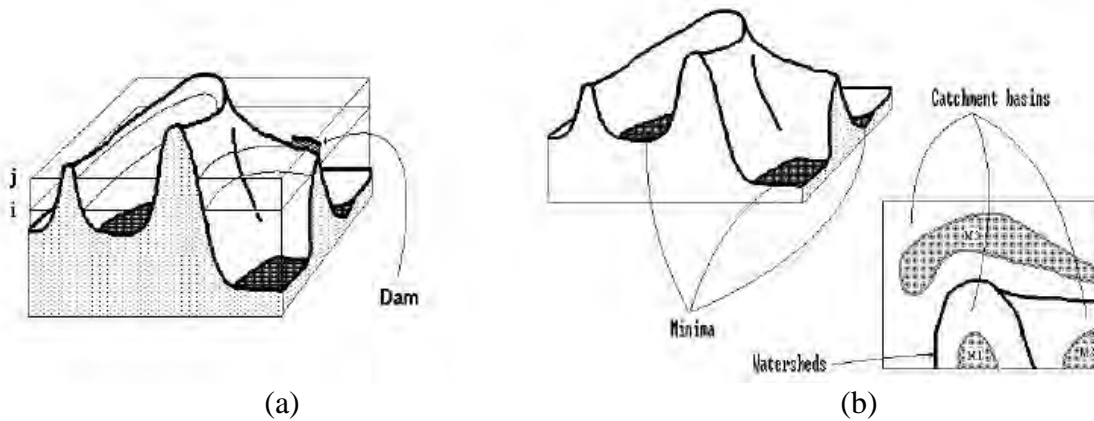


Figure 2.4 Flooding of the relief and dam building (a), catchment basins and divide lines (b), after Beucher (1991).

There are two stages to the marker selection process: pre-processing and marker definition. The pre-processing stage helps eliminate some of the local minima, whilst the marker definition dictates where the internal markers will be located. The pre-processing stage smooths the image with a smoothing algorithm. This is done by passing a mask across the image, averaging all of the values within that mask, and assigning the mean value to the image pixel on which the mask is centred.

A region that is a potential internal marker must satisfy the following criteria: it must be surrounded by points of higher grey values; all the points in the region must be connected; and all the connected points must have the same grey level. The watershed algorithm can now be applied to the image, but region growing can only occur from the internal markers. In other words, the internal markers are the only allowed local minima.

Chen *et al.* (2005) developed a marker-controlled watershed segmentation method for isolating trees from discrete-return LiDAR data. The key for the success of tree isolation is to find the proper treetops from the canopy height model derived from LiDAR data. The treetops were detected by searching local maxima in a canopy maxima model (CMM) with variable window sizes. The variable window sizes were determined by the lower-limit of the prediction intervals of the regression curve between crown size and tree height. The proposed treetop detection method minimised both the commission and omission errors simultaneously. However, this

study only aimed to isolate individual trees not to extract other forest parameters such as tree species, crown size and biomass.

Watershed transformation can only segment one layer of data at a time; therefore segmentation of multi-source data has common features with different shape and size. An easy way to fix this problem is to segment one layer and then mask out the feature boundaries from other data layers.

Watershed transformation is an example of a segmentation procedure that can utilise a set of complex algorithms. The main disadvantage of data-driven segmentation is that the user has less control over the segmentation procedures. A user-driven segmentation can incorporate user-defined parameters into the process. In the following, different user-driven segmentation procedures are presented.

User-driven segmentation

In user-driven approaches, the image is segmented to extract a predetermined type of target object using user-defined parameters. In this segmentation, user defined-parameters initiated the process and user has more control over the segmentation process. Multi-resolution segmentation, or FNEA (Baatz and Schape, 2000) is one example of user-driven segmentations.

Multi-resolution segmentation

Objects of interest typically appear on different scales in an image simultaneously. The extraction of meaningful image objects needs to take into account the scale of the problem to be solved. Therefore the scale of resulting image objects should be free adaptable to fit to the scale of task (Baatz and Schäpe, 1999). Multi-resolution segmentation (Baatz and Schape, 2000) utilises user-defined scale and homogeneity parameters in combination with local and global optimisation techniques for segmenting multi-source data. Segmentation and subsequent object topology generation is controlled by the resolution and the scale of the expected objects (deKok *et al.*, 1999). In this algorithm, a scale parameter is used to control the average image object size and homogeneity criteria are defined by the spectral and spatial information.

Multi-resolution segmentation is a bottom up region-merging technique starting with one-pixel objects. In numerous subsequent steps, smaller image objects are merged into bigger ones. Throughout this pair-wise clustering process, the underlying optimisation procedure minimizes the weighted heterogeneity $n h$ of resulting image objects, where n is the size of a segment and h a parameter of heterogeneity. In each step, that pair of adjacent image objects is merged which stands for the smallest growth of the defined heterogeneity. If the smallest growth exceeds the threshold defined by the scale parameter, the process stops. Consequently, multi-resolution segmentation is a local optimisation procedure (Benz *et al.*, 2004).

Heterogeneity in multi-resolution segmentation is defined by the colour and shape of the features in the image. The increase of heterogeneity f has to be less than a certain threshold.

$$\begin{aligned} f &= w_{colour} \cdot \Delta h_{colour} + w_{shape} \cdot \Delta h_{shape}, w_{colour} \\ &\in [0,1], w_{shape} \in [0,1], w_{colour} + w_{shape} = 1 \end{aligned} \quad (2.12)$$

The weight parameters (w_{colour} , w_{shape}) allow the heterogeneity definition to adapt to the application. The spectral heterogeneity allows multi-variant segmentation by adding a weight w_c to the image channels c . Difference in spectral heterogeneity Δh_{colour} is defined as following:

$$\Delta h_{colour} = \sum_c w_c (n_{merge} \cdot \sigma_{c,merge} - (n_{object1} \cdot \sigma_{c,object1} + n_{object2} \cdot \sigma_{c,object2})) \quad (2.13)$$

with n_{merge} number of pixels within merged object, $n_{object1}$ number of pixels in object 1, $n_{object2}$ number of pixels in object 2, σ_c standard deviation within object of band (or channel) c . Subscripts *merge* refer to the merged object, object 1 and object 2 prior to merge respectively.

The shape heterogeneity Δh_{shape} is a value that describes the improvement of shape with regard to smoothness and compactness.

$$\Delta h_{shape} = w_{compact} \cdot \Delta h_{compact} + w_{smooth} \cdot \Delta h_{smooth} \quad (2.14)$$

with

$$\Delta h_{smooth} = n_{merge} \cdot \frac{l_{merge}}{b_{merge}} - (n_{object1} \cdot \frac{l_{object1}}{b_{object1}} + n_{object2} \cdot \frac{l_{object2}}{b_{object2}}) \quad (2.15)$$

$$\Delta h_{compact} = n_{merge} \cdot \frac{l_{merge}}{\sqrt{n_{merge}}} - (n_{object1} \cdot \frac{l_{object1}}{\sqrt{n_{object1}}} + n_{object2} \cdot \frac{l_{object2}}{\sqrt{n_{object2}}}) \quad (2.16)$$

where n is the object size, l is the object perimeter and b is the perimeter of object's bounding box.

Thus, the *smoothness heterogeneity* equals the ratio of the de facto border length l and the border length b given by the bounding box of an image object parallel to the raster. The *compactness heterogeneity* equals the ratio of the de facto border length l and the square root of the number of pixels forming the image object.

The weights $w_c, w_{colour}, w_{shape}, w_{smooth}, w_{compact}$ are parameters that can be selected in order to get suitable segmentation results for a certain image data stack and a considered application. The scale parameter is the stop criterion for the optimisation process. Prior to the fusion of two adjacent objects, the resulting increase of heterogeneity f is calculated. If this result increase exceeds a threshold t , determined by the scale parameter $t = \psi$ (scale parameter), then no further fusion takes place and the segmentation stops (Benz *et al.*, 2004).

Flanders *et al.* (2003) used multi-resolution segmentation to segment and classify a Landsat ETM+ image for logged forest block delineation and feature extraction. In this process, by adjusting the segmentation scaling parameter to 200 from 25, an upper object level was created to delineate the built-up urban area. A lower image object level was also created using a scaling parameter of 25 to identify cut blocks present outside the urban area. This unit-less scale parameter is not related to the number of pixels, but rather to the maximum allowed heterogeneity within an object. Equal weight was assigned to each of the input ETM+ spectral bands (1-5 and 7). This emphasis was chosen because of the lack of colour homogeneity visually observed within some logged forest blocks in the image. For homogeneity, they applied the ratio of 0.2:0.8 (the sum must total 1.0) for relative weight of spectral

versus shape criteria, emphasizing the importance of logged forest block from over colour. Similarly, a ratio of smoothness to compactness weight was specified 0.1:0.9, emphasizing the discrete, compact nature of logged forest blocks. A higher smoothness emphasis was used to define objects observed to have greater variability between blocks. The compactness weight makes it possible to separate objects that have quite different shapes but not necessarily a great deal of colour contrast, such as logged forest block versus bare patches within forested area.

Van-der-Sande *et al.* (2003) applied a multi-resolution segmentation and classification approach on IKONOS-2 imagery for land cover mapping to assist in flood risk and flood damage assessment. They applied segmentation in four different levels to extract optimum classification results. At each level they extracted specific thematic classes from the image. They included or excluded the IKONOS spectral bands from the segmentation process as a function of their information content. Segmentation at highest level (Level 4) was used to classify the larger objects in the study area such as agricultural fields and water bodies. The red and near-infrared spectral bands were used together with the homogeneity ratio 0.9:0.1 (colour: shape) for the classification of vegetation types and crops as well as water bodies. Small roads were extracted from segmentation at mid-high level (Level 3) using the green, red and near-infrared spectral bands with the scale parameter 30 and the homogeneity ratio 0.5:0.5. Buildings were extracted from mid-low level (Level 2), where a scale parameter of 10 resulted in objects small enough to differentiate individual houses. The lowest level (Level 1) was the aggregation of all levels and it was the final thematic land cover layer. An overall classification accuracy of 74 percent was achieved. In spite of the good results of land cover mapping a number of classes such as residential areas and roads were fairly difficult to identify.

Zhou and Troy (2008) presented an object-oriented approach for analysing and characterizing the urban landscape structure at the parcel level using high-resolution digital aerial imagery and LIDAR data for the Baltimore area. Additional spatial datasets including property parcel boundaries and building footprints were used to both facilitate object segmentation and obtain greater classification accuracy. A three-level hierarchical network of image objects was generated, and objects were classified. At the two lower levels, objects were classified into five classes, building,

pavement, bare soil, fine textured vegetation and coarse textured vegetation, respectively. The object-oriented classification approach proved to be effective for urban land cover classification. The overall accuracy of the classification was 92.3 percent, and the overall Kappa statistic was 0.899. Land cover proportions as well as vegetation characteristics were then summarized by property parcel. This exercise resulted in knowledge base rules for urban land cover classification that could potentially be applied to other urban areas.

In feature-level fusion, the defined features need to be classified after segmentation. A number of segmentation algorithms have been reviewed in the above sections. Now different classification algorithms to fuse the segmented features will be reviewed.

2.8 Classify the Segmented Features

After segmentation, the features from multi-source images are usually classified (or fused) to certain classes according to the class description. A class description is the hierarchical classification system (described in Section 2.2) that guides the fusion process and has the typical properties or conditions of the desired classes. In terms of feature classification, each feature belongs to one definite class or to no class. Traditional pixel-based classifiers such as maximum-likelihood, minimum-distance, or parallelepiped can be used to classify the features into different classes. However, the difference is that they are applied to the features rather than the pixels. A detailed review of these traditional classifiers was presented in Section 2.5. The main disadvantage of these classifiers is that they express the pixel's membership of a class only in a binary manner. Such classifiers are usually called *hard classifiers*. In contrast, *soft classifiers* use a degree of membership, or a probability, to express a pixel's assignment to a class. The membership-probability value must lie between 1.0 and 0.0, where 1.0 expresses full membership-probability to a class and 0.0 expresses absolute nonmembership-probability. Thereby the degree of membership reflects on the degree to which a pixel fulfils the class-description. The main advantage of these soft classifiers lies in the possibility to express uncertainties about class descriptions. With respect to image understanding, these soft classification results are more capable of expressing uncertain human knowledge about the world

and thus lead to classification results which are closer to human language, thinking and mind (Baatz *et al.*, 2004). Fuzzy classification is one of the powerful soft classifiers and can be used in both pixel and feature domain. The following section reviews the fuzzy classifier as a tool to classify segmented features in a feature-level fusion.

2.8.1 Fuzzy classification of segmented features

In feature-level fusion, the feature classes can be defined as fuzzy sets and the features as set elements. Accordingly, each feature is allocated a group of membership grades to indicate the extent to which the feature belongs to predetermined land covers. The allocation of membership grades is termed *fuzzification*. The notion of fuzzy classification is to generate the fuzzified values based on the fuzzy set theory introduced by (Zadeh, 1965). Thus, fuzzification is the single most important consideration in fuzzy classification.

Zadeh (1965) proposed the concept of fuzzy sets, which introduces vagueness by eliminating the sharp boundary dividing members from non-members of a class. Specifically, it assigns a membership value, as a function of its similarity to the individual class, and typically varies between 0 and 1. A membership close to 1 denotes strong relations to a particular class, whereas a membership close to 0 represents a weak relationship.

A fuzzy set can be defined mathematically. Let X denotes a universal set with a generic element of X denoted by x , thus $X =$.

$$A = \{x, \mu_A(x)\}; x \in X \quad (2.17)$$

A fuzzy set A (land cover class) in X is characterized by a membership function $\mu_A(x)$ which associates with each point in X a real number in the interval $[0, 1]$. The value of $\mu_A(x)$ represents the grade of membership of x in A (Zadeh, 1965). There are two possible ways of defining these membership functions: Semantic Import and Similarity Relation model.

The Semantic Import (SI) approach uses an *a priori* membership function with which individuals can be assigned a membership grade. This is useful in situations where users have a very good, qualitative idea of how to group data (Burrough and McDonnell, 1998). The SI approach relies on expert knowledge where boundary values are chosen by custom, law or external taxonomy to generate the fuzzy membership (Burrough, 1986). The selection of class boundaries and class intervals can be an objective or subjective process, depending on the way scientists agree to define classes. According to Burrough and McDonnell (1998), the membership function should ensure that the grade of membership is 1.0 at the centre of the set and that it falls off in an appropriate way through the boundaries to the region outside the set where it takes the value of 0. The location where the fuzzy membership equals 0.5 is called 'crossover point'. These conditions need to be considered in defining fuzzy membership function.

The second approach, named by Robinson (1988), is the Similarity Relation Model, resembles cluster analysis and numerical taxonomy is that the value of the membership function is a function of the classifier used. A common version of this model is the fuzzy c-means, also known as fuzzy k-means, model based on cluster analysis.

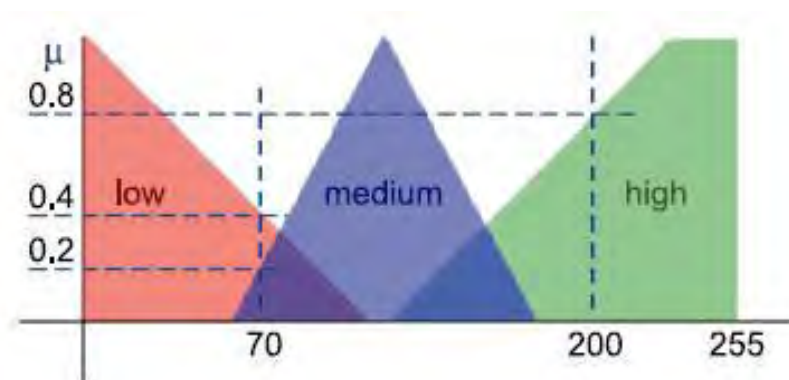


Figure 2.5 Example for three fuzzy sets on feature x . The membership functions on feature x define the fuzzy set *low*, *medium* and *high* for this feature, after Benz *et al.*, (2004).

A fuzzy rule-base is a combination of fuzzy rules, which combine different fuzzy sets. The simplest fuzzy rules are dependent on only one fuzzy set. Fuzzy rules are 'if-then' rules. If a condition is fulfilled, an action takes place. An example is: 'If'

feature x is low, ‘then’ the image object should be assigned to land cover W . According to the example in Figure 2.4, in the case where the feature value $x=70$, the membership to land cover W would be 0.4, in the case where $x=200$, the membership to land cover W would be 0.

Given two or more fuzzy membership functions for the same set, a variety of operators can be employed to combine the membership values together. The logic operators are ‘and’, ‘or’ and ‘not’. There are several possibilities to realize these operators. In most cases, ‘and’ operator implements the smallest fuzzy membership and ‘or’ operator implements the maximum (Benz *et al.*, 2004). It should be noted that, while fuzzy classification gives a possibility for an object to belong to a class, classification based on probability provides a probability to belong to a class. A possibility gives information on a distinct object, whereas probability relies on statistics and provides one value from many objects. Probability of all possible events adds up to one, this is not necessarily true for all possibilities. The non-normalized possibility values provide additional information on the classification reliability for each object (Benz *et al.*, 2004).

De-Kok *et al.*, (1999) demonstrated the usefulness of fuzzy classification in discriminating spectrally similar classes in order to improve classification accuracy from high resolution panchromatic and multispectral imagery of an alpine forest. Fuzzy logic decision rules offer a larger deduction in complexity and a proper aid to group the spatial objects into meaningful classes. In this study, the fuzzy logic decision rules for class membership are the framework in which the expert knowledge has been embedded. The synergy of the spectral properties, the neighbourhood object influences, and the expert knowledge lead to powerful ways of making object membership decision rules. The fuzzy logic rules guarantee the transparency of the decision rules and reduce complexity to a condensed crisp set of membership functions.

It is apparent that fuzzy classification has advantages over conventional classification in:

- (a) expressing the vagueness of spatially distributed categories e.g. land covers derived from remote sensing images;

- (b) offering information on continuum of land cover classes; and
- (c) generating fuzzy boundaries, e.g. uncertain zones among the land cover classes.

The arbitrariness of defining the fuzzy membership is a major constraint in generating final fuzzy land cover maps and subsequent accuracy assessment analysis. Zadeh (1965) proposed the idea of presenting fuzzy membership above a suitable threshold as unions of defuzzified areal classes and fuzzy boundaries. The concept of a defuzzification based on the maximum membership values obtained for individual locations can be applied to generate land cover maps using the outputs of fuzzy classification (Zhang and Goodchild, 2002).

2.9 Evaluation of Fused Results

Pixel- and feature-level fusions have some distinct characteristics and that should be evaluated as a comparative basis for selecting a fusion for a particular application. The following section reviews different mechanisms to compare pixel- and feature-level fusion.

Fusion of remotely sensed data derived from a range of sensors has led to the requirement for increased knowledge of errors and their contribution to the overall quality of the fused product. Error accumulation in remote sensing and GIS is difficult to monitor, for variations in the target, sensor, sensing geometry and ambient environment conditions create specific problems (McGwire and Goodchild, 1997). Both quantitative and qualitative errors are important in remote sensing data fusion. Quantitative errors relate to the positional accuracy of the map data, while qualitative errors are concerned with the correctness of thematic classification of feature within the data (Maling, 1988). In this research, the latter category of errors is explored for the data fusion.

According to Davis and Simonett (1991), classification differences between remotely sensed and reference data arise for a range of reasons:

- (a) misregistration of the sensor-derived data to the cartographic coordinate system;

- (b) misregistration of reference data to the cartographic coordinate system;
- (c) spectral confusion between information classes;
- (d) inappropriate fusion/classification algorithm;
- (e) poor definition of information classes for training and test data;
- (f) information classes containing several classes; and
- (g) sub-pixel variation causing mixed pixel and boundary effects.

Error assessment is required to quantify the accuracy of fusion results and to guide the process of analysis to determine the sources of error. Understanding the above factors can lead to refinement of the fusion process and improvements in the fusion quality. Accuracy of the fusion results can be assessed through site specific and non-site specific methods. Non-site specific methods only provide an assessment of the total area occupied by a specific landscape without performing any location-by-location comparison. If all fusion errors between categories in a non-site specific assessment balance out, it is possible to achieve very high results for the accuracy assessment, however, the results will be misleading. Site-specific approaches make a comparison in one-on-one samples of the fused and reference data, providing an assessment of the fusion error with regard to the location of the fused and reference data. Statistically sound approaches to sample size and sampling design are essential to perform valid assessment of fusion accuracy for landscapes of varying spatial diversity (McCoy, 2004; Congalton, 1991).

2.9.1 Accuracy assessment of fusion results

It is necessary to define *precision* and *accuracy* in error assessment perspective. Precision is defined as the degree of detail in reporting a measurement, which is often determined by the characteristics of the measuring equipment, while accuracy is defined as a measure of the difference between a measured value and a known or true value (McGwire and Goodchild, 1997). From a thematic mapping perspective, precision is related to the level of detail (or generalisation) inherent in the thematic mapping classification system (Janssen and van-der Wel, 1994). On the other hand,

accuracy relates to the agreement of the classified image with a source of reference data giving greater accuracy than the primary remote sensing information.

Accuracy assessment in this research will be directed towards the assessment of the fusion results achieved through pixel- and feature-level fusion of LiDAR and optical imagery. The class hierarchy described in Section 2.2 determines the precision of the fusion process. As the degree of detail increases from lower level to upper level, the possibility of errors also increases, which may lead to more uncertain results and, logically, a lower accuracy of fusion results.

Descriptive techniques

Application of a robust sampling scheme enables acquisition of representative samples of every landscape class, and provides relevant data for population of the error matrix. An example of an error matrix (also known as a confusion matrix or contingency table) is shown in Figure 2.6.

		Reference classification					
		Class 1	Class 2	...	Class N		
Observed classification	Class 1	a₁₁	a ₁₂	...	a _{1N}	$\sum_{k=1}^N a_{1k}$	
	Class 2	a ₂₁	a₂₂	...	a _{2N}	$\sum_{k=1}^N a_{2k}$	
	⋮	⋮	⋮	⋮	⋮	⋮	
	Class M	a _{M1}	a _{M2}	...	a_{MN}	$\sum_{k=1}^N a_{Mk}$	
		$\sum_{k=1}^N a_{k1}$	$\sum_{k=1}^N a_{k2}$...	$\sum_{k=1}^N a_{kN}$	$n = \sum_{i,k=1}^N a_{ik}$	

Figure 2.6 A conceptual model of error matrix.

The *Overall Accuracy (OA)* is computed by dividing the sum of the diagonals by the total of samples checked in the accuracy assessment, and provides a measure of the proportion of all sampled pixels that are classified correctly. Eq. 2.18 illustrates the computation procedure to obtain Overall Accuracy from the confusion matrix.

$$OA = \frac{\sum_{k=1}^N a_{kk}}{\sum_{i,k=1}^N a_{ik}} = \frac{1}{n} \sum_{k=1}^N a_{kk} \quad (2.18)$$

where n is the number of all reference pixels. So OA is the sum of the diagonal entries of the confusion table divided by the number of all reference pixels. Overall accuracy is a very coarse measure. It gives no information about what classes are classified with good accuracy. In fact, a classification with poor overall accuracy may have a certain class with high accuracy, thus may confuse other class results and will be of interest for certain applications only. Therefore, it is important to use other accuracy measures together with Overall Accuracy.

One such measure is the *Producer's accuracy (PA_{class_i})* that estimates the percentage of a particular class is correctly classified against ground truth data. Mathematical representation of the Producer's accuracy is shown in Eq. 2.19. For each $class_i$, the proportion of pixels where classified and reference data agree in $class_i$ and the reference pixels are classified as this class. Total number of the pixels of $class_i$ in the reference classification is obtained as the sum of column i in the confusion table.

$$PA(class_i) = \frac{a_{ii}}{\sum_{i=1}^N a_{ki}} \quad (2.19)$$

Producer's accuracy is actually a measure for the producer of a classification, which indicates how well the classification agrees with the reference classification. However, it gives no information about how well the classification predicts a class, such as, it gives no information about the probability that a pixel classified as $class_i$ is actually of $class_i$. This is the primary interest of a user of a classification and an

estimate of this probability is thus called *User's accuracy* (UA_{class_i}). The estimation of this probability is given by the proportion of pixels for which classification and reference classification agree in $class_i$ compared to the number of all reference pixels classified as $class_i$ by the classification (Eq. 2.20). Now the total number of pixels that are classified as $class_i$ is obtained by the sum of row i of the confusion table.

$$UA(class_i) = \frac{a_{ii}}{\sum_{i=1}^N a_{ki}} \quad (2.20)$$

According to Janssen and van-der Wel (1994) the User's and Producer's accuracy relate to errors of omission and commission as follows:

$$\text{Error of commission} = 100 - \text{User's Accuracy} \quad (2.21)$$

$$\text{Error of omission} = 100 - \text{Producer's Accuracy} \quad (2.22)$$

The User's Accuracy is a measure of the classification reliability and errors of commission express the severity of it (Eq. 2.21). The Producer's Accuracy gauges the proportion of pixels that actually belong to a category but have been classified as other features. Errors of omission express the degree to which this type of error occurs (Eq. 2.22). The User's and Producer's Accuracy also permit a more complete understanding of the between-class confusion for the purpose of fusion parameter refinements.

Analytical techniques

Error matrices form the basis of several analytical statistical techniques developed to evaluate classification accuracy of remotely sensed data. Most approaches utilise discrete multivariate analysis because remotely sensed data are discrete rather than continuous. Most data demonstrate properties of binomial or multinomial distributions, therefore many methods based upon normal probability theory are not appropriate (Congalton, 1991).

Accuracy assessments including all elements of the error matrix may be undertaken using the Kappa Coefficient of Agreement (Cohen, 1960). The Kappa Coefficient was developed for comparison of data grouped by different observers (or interpreters or classification algorithms) according to nominal scales. The overall level of agreement for an error matrix (Kappa Coefficient) is based upon the difference between the actual agreements of the classification compared with the reference data (measured by the matrix diagonal), and the chance agreement, which is indicated by the product of the row and column margin values.

The application of the Kappa Coefficient to the analysis of accuracy of remotely sensed data was first proposed by Congalton *et al.* (1983) and has been widely reported since (Smith *et al.*, 2003; Crosetto and Tarantola, 2001; Hyypya *et al.*, 2000; Jäger and Benz, 2000; Muller *et al.*, 1998b). The method may be used to evaluate an error matrix as a whole or for individual classes, or it may be used to statistically compare error matrices derived from different interpreters or using a variety of fusion techniques. Equation 2.23 shows the computation of Kappa Coefficient (\hat{K}):

$$\hat{K} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} * x_{+i})} \quad (2.23)$$

where r is the number of rows/columns in the error matrix, x_{ii} is the number of observations in row i and column i , x_{i+} is the total of column i and N is the total number of observations. A pair-wise assessment of the significance of the differences between two independent error matrices can be undertaken using the normal curve deviate determined from the corresponding Kappa statistics and their variances (Cohen, 1960). Equation 2.24 shows the computation of variance of Kappa (Bishop *et al.*, 1975):

$$\sigma_{\hat{K}}^2 = \frac{1}{N} \left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4} \right\} \quad (2.24)$$

where

$$\theta_1 = \sum_{i=1} \frac{x_{ii}}{N} \quad \theta_2 = \sum_{i=1} \frac{x_{i+} * x_{+i}}{N^2}$$

$$\theta_3 = \sum_{i=1} \frac{x_{ii}}{N^2} (x_{i+} + x_{+i}) \quad \theta_4 = \sum_{i=1, j=1} \frac{x_{ij}}{N^3} (x_{j+} + x_{+i})^2$$

The test statistic for significance in large samples ($N > 100$) is given by:

$$z \sim \frac{\hat{K}_1 - \hat{K}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \quad (2.25)$$

The Kappa statistic provides statistically valid assessments of the quality of fusion and enables tests of significance between fusion processes for determination of optimum algorithm performance. The result of performing a Kappa analysis is another measure of agreement or accuracy. However, it follows a different idea. Whereas Overall Accuracy checks how many of all pixels are classified correctly, assuming that the reference classification is true, here it is assumed that both fusion and reference data are independent class assignments of equal reliability. The big advantage of the Kappa Coefficient over overall accuracy is that Kappa takes chance agreement into account and corrects for it. According to Landis and Koch (1977) the qualitative descriptors show in Table 2.4 describe the strength of agreement based on Kappa statistics.

Table 2.4 Qualitative descriptions for the strength of agreement for the Kappa statistics, after *Landis and Koch, (1977)*

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.19	Slight
0.20-0.39	Fair
0.40-0.59	Moderate
0.60-0.79	Substantial
0.80-1.00	Excellent

In statistical analysis, the Kappa Coefficient is used to compare the fusion results with reference data. Confidence intervals around the Kappa Coefficient can be computed using the approximate large sample variance and the fact that the Kappa statistic has normal distribution (Congalton and Green, 1999). The Kappa Coefficient provides a means for testing the significance of the Kappa statistic for the error matrix to determine if the agreement between the fusion results and the reference data is significantly greater than 0, i.e. better than a random fusion result.

2.9.2 Sampling design for accuracy assessment

In order to measure the quality of fusion processes and to compare and evaluate the results with respect to their suitability for a specific application, accuracy measures are used. Mostly they are derived on the basis of a comparison of the results in question with other reference data. This latter reference is often obtained using different methods, e.g., by ground truth measurements. Ground truth data collection requires sampling and sampling requires knowledge of the distribution and sampling design (McCoy, 2004). The selection of a proper and efficient sample design to collect valid reference data is one of the most challenging and important components of accuracy assessment because the design will determine both the cost and the statistical rigour of the assessment (McCoy, 2004; Congalton, 1991).

Sampling scheme

Assessment of the quantitative or qualitative aspects of fusion results rely on a sampling scheme with a common set of criteria such that a low accuracy fusion product has low probability of acceptance, a high accuracy fusion product has high probability of acceptance, and a minimum number of reference data samples is required.

A sampling scheme may be developed to estimate the effectiveness of different accuracy parameters, provide information on landscape of limited area extent, and evaluate different fusion procedures. According to Stehman (1999), the important considerations for sampling design are: satisfies sampling design procedure; simple to implement and analyse; low variance for estimates of high priority accuracy

measures; permits adequate variance estimation; provides samples which are spatially well distributed; and cost effectiveness.

Fundamental to the design of a suitable sampling scheme is the selection of reference data independently of the data used to develop the fusion process. This particularly relates to the selection of independent samples selected for image registration and assessment of positional accuracy (McCoy, 2004). Sampling schemes generally follow simple random or systematic selection procedure and utilise population, strata or cluster sampling structures (Congalton, 1991). The most relevant sampling structure and procedure are subject to considerable debate and are reviewed in the following:

- (i) *Simple random sampling*: Simple random sampling may be used to sample pixels in an image and represents a straightforward approach, although location of the sample in the field may be difficult. This method is adaptable to augmenting or reducing the sample size if required, and estimation and standard error formulae are less complex compared to other approaches. The variance in simple random sampling tends to be large where categories have small samples.
- (ii) *Systematic sampling*: Systematic sampling may be undertaken on a rectangular or square grid based on the random location of the starting pixel. The method provides good spatial coverage and is easy to implement. Because it is an equal probability sampling design it share the same advantages and disadvantages as simple random sampling, e.g. small sub-regions will remain under-sampled.
- (iii) *Stratified random sampling*: Each pixel within the population is assigned to a stratum then samples are selected at random within each stratum. Strata may be defined on a category or geographic basis, with each random sample derived independently for each stratum. Stratified random sampling is used to ensure that each stratum is represented within the random sample, but care must be taken to ensure the strata remain valid throughout the analysis. Where analysis objectives change, such as when landscape classes are combined, a new stratified random sample should be extracted in order to retain relevance (Stehman, 1999).

(iv) *Cluster sampling*: Cluster sampling uses two sizes of sampling units. The primary sampling unit which is the cluster itself, and the secondary sampling unit which is represented by the individual pixels within the cluster. The location of each cluster may be defined by any of the above sampling schemes, with each pixel in a cluster forming part of the sample. Cluster sampling is performed mainly for the purposes of convenience and cost, as only the clusters need be evaluated rather than individual pixels (Congalton, 1991).

Concerns regarding the statistical validity of systematic sampling designs have made researchers consider stratified random sampling the most suitable approach for sampling spatial data for the purpose of accuracy assessment (McCoy, 2004; Congalton and Green, 1999; Lo and Watson, 1998). Research indicates that systematic sampling in the absence of data periodicity is the most statistically valid sampling approaches to employ (Stehman, 1999). This applies especially where spatial autocorrelation is present, because it provides for maximum average spatial separation of the sample. Systematic sampling designs also provide a uniform spatial distribution of samples, a factor that is often used to justify stratified random sampling.

Sample size

The major objective of selecting a sample of appropriate size is to provide sufficient data to enable a reliable estimate of accuracy at the required confidence level. Where error matrices are used as part of the evaluation and large numbers of categories are sampled, the requirement to adequately sample each becomes important (McCoy, 2004; Congalton, 1991).

According to Congalton (1991), the rule of thumb is to select at least 50 samples per category and 75 samples where categories occupy large areas. There is considerable variation in the specification of class sample sizes with Richards (1993) suggesting values between 30 and 60 for most situations. Limitations on sample sizes are usually set in order to economise on fieldwork expenditure.

2.10 Comparisons of Pixel- and Feature-level Fusion

Comparisons between pixel- and feature-level fusion results can be carried out in a qualitative or quantitative manner. The qualitative measures mainly depend on visual inspection and vary with personal choice, whereas quantitative approaches are composed of rigorous sampling that provides some numerical values for the application of statistics.

2.10.1 Qualitative comparisons

Kamagata *et al.* (2005) compared pixel- and feature-level fusion of land cover using high resolution satellite data available for an urban fringe area. In the pixel-based analysis the maximum likelihood method and the ISODATA method were applied. The results showed that in both methods misclassification tended to increase in areas of shadow. The pixel-level fusion also experienced difficulty due to factors such as the varied shapes of the forest canopy and mixing of vegetation. The feature-level fusion, in contrast, relied on abstraction of comparatively homogenous areas, and proved capable of extracting the boundaries between forest types. This study employed a high number of minute patches that proved effective even in regions where tree species were mingled together. Some misclassification problems remained, which have to be addressed by future trial and error experiments in the parameter optimisation. In the Kamagata *et al.* (2005) study, the comparison was made solely on visual inspection and therefore the results were not quantifiable. However, they concluded that the feature-level fusion has a high potential for analysing landscape patterns even in highly heterogeneous and rapidly changing urban areas.

2.10.2 Quantitative comparisons

Statistical analysis can be tested using two independent Kappa Coefficients to determine the significance of the two error matrices. With this test it is possible to statistically compare two fusion results and see which produces the higher accuracy. The test of significance relies on the standard normal derivation as explain below:

Let \hat{K}_1 and \hat{K}_2 denote the estimates of the Kappa statistics for the error matrix of fusion result 1 and the error matrix of fusion result 2. The derivation of \hat{K} was shown in Equation 2.23. Let $\hat{\text{var}}(\hat{K}_1)$ and $\hat{\text{var}}(\hat{K}_2)$ be the corresponding estimates of the variance as computed from Equation 2.24. The test statistic for testing the significance of a single error matrix is expressed by

$$Z = \frac{\hat{K}_1}{\sqrt{\hat{\text{var}}(\hat{K}_1)}} \quad (2.26)$$

Where Z is standardised and normally distributed (*i.e.*, standard normal derivation). Given the null hypothesis $H_0 : K_1 = 0$ and the alternative $H_1 : K_1 \neq 0$, H_0 is rejected if $Z \geq Z_{\alpha/2}$, where $\alpha/2$ is the confidence level of the two-tailed Z test and the degrees of freedom are assumed to be ∞ (infinite). The test statistics for testing of two independent error matrices are significantly different from Equation 2.26 and is expressed by

$$Z = \frac{|\hat{K}_1 - \hat{K}_2|}{\sqrt{\hat{\text{var}}(\hat{K}_1) + \hat{\text{var}}(\hat{K}_2)}} \quad (2.27)$$

Here Z is also standardised and normally distributed. Given the null hypothesis $H_0 : (K_1 - K_2) = 0$ and the alternative $H_0 : (K_1 - K_2) \neq 0$, H_0 is rejected if $Z \geq Z_{\alpha/2}$.

The variance of the Kappa statistic and the Z statistic is used for determining if a fusion technique is significantly better or worse than one other. At 95 percent confidence level, the critical value would be 1.96. Therefore, if the absolute value of the test Z statistic is greater than 1.96 the result is significant and we can conclude that the first technique is better than the second one.

Geneletti and Gorte (2003) compared the pixel- and feature-level fusion of high-resolution panchromatic and multispectral images. In this process, they used a high-

resolution black and white orthophoto and a sub-scene of a Landsat Thematic Mapper (TM) image of a study area in northern Italy. Their method consisted of a sequential application of segmentation and classification techniques. First, the TM image was classified using a pixel-based maximum likelihood classifier and additional empirical rules. Subsequently, the orthophoto was segmented by applying a region-based segmentation algorithm. Finally, the classification of the segmented images was performed using as a reference the TM image previously classified. Using the test data, accuracy assessments were performed both on the results of pixel- and feature-level fusions and classifications. According to the confusion matrices, the accuracy and reliability of pixel-level classification was 83.80 percent and feature-level was 86.26 percent. The accuracy results revealed that the feature-level classification performed better in distinguishing practically all the cover classes, improving both average accuracy and average reliability by about 2 percent. According to Geneletti and Gorte (2003), unlike pixel-level fusion, the feature-level approach produces as output a thematic map composed of geographic entities labelled with land cover classes and, as such, can be directly stored on to GIS databases, creating or updating usable geo-information.

The main criticism of the Geneletti and Gorte (2003) approach was of the segmentation process. They used knowledge-based segmentation, which incorporated knowledge derived from training areas or other sources into the segmentation process. This segmentation is mostly specific, not necessarily robust, and does not necessarily deliver homogeneous areas (Baatz and Schape, 2000).

Wang et al. (2004) fused IKONOS 1-m panchromatic and 4-m multispectral images to map mangroves in a study site located at Punta Galeta on the Caribbean coast of Panama. They hypothesised that spectral separability among mangrove species would be enhanced by taking the object (feature) as the basic spatial unit as opposed to the pixel. Three different classification methods were investigated: maximum likelihood classification (MLC) at the pixel-level, nearest neighbour (NN) classification at the feature-level, and a hybrid classification that fuses the pixel- and feature-level methods (MLCNN). Specifically for object segmentation, which is the key step in feature-level classification, they developed a new method to choose the optimal scale parameter with the aid of Bhattacharya Distance (BD), a well-known

index of class separability in traditional pixel-level classification. A comparison of BD values at the pixel level and a series of scale parameters not only supported the initial hypothesis, but also helped to determine an optimal scale at which the segmented objects have the potential to achieve the best classification accuracy. Among the three classification methods, MLCNN achieved the best average accuracy of 91.4 percent. Wang et al. (2004b) also compared the performance of IKONOS and QuickBird images in the classification of mangrove stand composition. In pixel-level classification, multispectral bands were employed to classify seven land cover types. Classification based on the IKONOS image was slightly, but significantly, more accurate than classification based on the multispectral QuickBird image (Kappa Z statistic 1.98). The addition of the panchromatic band to the classification did not significantly change the accuracy of classification for either image type. However, the \hat{K} value (7.75) from IKONOS classification was still significantly higher than that from the equivalent QuickBird bands (Kappa Z statistics 1.98). When texture information in both panchromatic and multispectral bands was used to aid feature-level classification of mangrove species, IKONOS and QuickBird demonstrated almost equal classification effectiveness (73 percent) with \hat{K} values of 0.69. In general, both IKONOS and QuickBird images presented promising results in classifying mangrove species. Spectral information played a more important role in classifying mangrove species than spatial information did.

Walter (2004) investigated change detection based on both pixel- and feature-level classification. In the pixel-level classification airborne four band (RGB & NIR) multispectral imagery was used, whereas in the feature-level 14 bands were included: mean grey value of the 4 spectral bands, vegetation index, texture from blue band, variance of the 4 spectral bands, and percent of forest, green land, settlement and water pixels. All features of the test areas were used as training objects for the classification. In a manual revision, the GIS data were compared with the images. The number of features that were either not collected correctly, or for which it was not possible to decide if they were collected correctly without further information sources constituted more than 6 percent of all features. Walter (2004) proposed refining the process by including LiDAR data. The suggestion was that the results of a pixel-level classification could be improved significantly by the combined use of

multispectral and LiDAR data. LiDAR data can improve the classification result because they have behaviour complementary to multispectral data. Walter (2004) predicted that feature-level classification should also be improved by the combined use of multispectral and LiDAR data. Yan *et al.* (2006) compared pixel-based and object-oriented classifications for land-cover mapping in a coal fire area. Using the object-oriented classification, the overall accuracy was higher than the accuracy obtained using the pixel-based classification by 36.77 percent and the User's and Producer's accuracy of almost all the classes were also improved.

Platt and Rapoza (2008) compared a traditional pixel-based classification using maximum likelihood classification with results from OBIA for a mixed urban-suburban-agricultural landscape surrounding Gettysburg, Pennsylvania. They used 4m spatial resolution and a four spectral bands IKONOS satellite image. They noted that OBIA has at least four components not typically used in per-pixel classification: (1) the segmentation procedure, (2) the nearest neighbour classifier, (3) the integration of expert knowledge, and (4) feature space optimisation. They evaluated each of these components individually and found that the combination of segmentation into image objects, use of the nearest neighbour classifier, and the integration of expert knowledge yielded substantially improved classification accuracy for the scene, compared to a per-pixel method. Specifically, OBIA-derived classification accuracy was 78 percent compared to the pixel-based accuracy of 64 percent.

Zhou *et al.* (2009) presented a comparative study of three methods using high spatial resolution imagery for land cover classification of shaded areas in an urban environment. Method 1 combined spectral information in shaded areas with spatial information for shadow classification. Method 2 applied a shadow restoration technique - the linear-correlation correction method - to create a "shadow-free" image before classification. Method 3 used multi-source data fusion to aid in classification of shadows. The results indicated that Method 3 achieved the best accuracy, with overall accuracy of 88 percent. It provides a significantly better means for shadow classification than the other two methods. The overall accuracy for Method 1 was 81.5 percent, slightly but not significantly higher than the 80.5 percent using Method 2. All three methods applied an object-based classification procedure,

which was critical as it provides an effective way to address the problems of radiometric difference and spatial misregistration associated with multi-source data fusion (Method 3), and to incorporate thematic spatial information (Method 1). Application of the shadow detection and restoration methods help to eliminate the shadow problem in land cover classification when using high spatial resolution images in urban settings.

Riggan and Weih (2009) compared an object-based classification procedure utilising Feature Analyst software (VLS, 2011) with a traditional pixel-based methodology (supervised classification). Medium-spatial resolution multispectral SPOT-5 satellite images were merged with high-spatial resolution colour infrared aerial images. A training-set was produced by selecting and identifying specific land cover class-types using 30cm spatial resolution aerial photos. This training set was used by both of the classification methods (supervised and object-based) to identify the various cover types within the study area. An accuracy assessment was performed on each image utilising error matrices, the Kappa coefficient, and a two-tailed Z-test. Results indicate that the overall accuracy of the object-based classification was 82.0 percent, while the pixel-based classification was 66.9 percent accurate. A Kappa co-efficient and a two-tailed Z test were calculated. These values indicated a significant difference in the overall accuracies of the two classifications.

2.11 Concerns about Per-pixel Based Thematic Accuracy Assessment

Although the basic approaches to accuracy assessment seem relatively straightforward, many problems are often encountered when evaluating an image classification. These range from issues associated with a failure to satisfy basic underpinning assumptions through to the limited amount of information on map quality that is actually conveyed by a basic accuracy assessment (Foody, 2002). A variety of errors are encountered in fusing images. Typically, interest focuses on thematic accuracy, which is the correspondence between the class label assigned by the fusion and that observed in reality. The confusion matrix appears to provide an excellent summary of the two types of thematic error that can occur, namely, omission and commission. However, other sources of error that contribute to the pattern of misclassification are not depicted in the confusion matrix (Husak *et al.*,

1999; Canters, 1997; Congalton and Green, 1993). Non-thematic errors may result in misrepresentation and underestimate the actual accuracy (Congalton and Green, 1993). Unfortunately, non-thematic errors can be large and particular concern focuses on errors due to misregistration of the image classification with the ground data (Muller *et al.*, 1998a; Canters, 1997). This positional uncertainty can have a major detrimental effect on thematic mapping. Significant misregistration problems have often been observed in the mapping of large areas where the problems of obtaining a high positional accuracy have been noted to be a major source of classification error and they are sometimes larger than the actual thematic error (Muller *et al.*, 1998a). It has been assumed that the ground, or reference data, used in the assessment of classification accuracy are themselves an accurate representation of reality. In fact, the ground data are just another classification, which may contain error (Congalton and Green, 1999). These may be thematic errors in which the class labels are erroneous but may also include other errors such as those due to mislocation. Problems with ground data accuracy may be particularly severe if a remotely sensed data set is used as the reference data. Unfortunately, the use of remotely sensed data as reference data is common in the 'validation' of coarse spatial resolution map products depicting very large areas (Justice *et al.*, 2000). A further problem arises as a consequence of the sampling strategy adopted in some ground/reference data collection programs. The size of the sampling units used in ground data collection is often different to the units mapped from the imagery (e.g., pixels or parcels) leading to difficulties in analysing the data sets (Atkinson *et al.*, 2000). The comparison of ground and thematic map labels may, therefore, be based upon differently sized units, which can result in different estimates of classification accuracy (Biging *et al.*, 1999). Irrespective of the misregistration and the problem of mixed pixels, the spatial variability of error can be a major concern, particularly in terms of error propagation. The confusion matrix and the accuracy metrics derived from it provide no information on the spatial distribution of error (Canters, 1997). In classical accuracy assessments all misallocations are equally weighted. Some errors are often more important or damaging than others (Forbes, 1995). In many instances, the errors observed in a classification are between relatively similar classes and so relatively unimportant, while other errors might much more serious (Zhu *et al.*, 2000). An additional source of error associated with the use of a standard (hard)

classifier that allocates each pixel to a single class is the implicit assumption that the image is composed of pure pixels. Unfortunately, remotely sensed data are often dominated by pixels that represent areas containing more than one class and these are a major problem in accuracy assessment (Foody, 1996). In some cases, mixed pixels have been identified as the most important cause of misclassification and a major contributor to the misrepresenting of land cover change (Skole and Tucker, 1993).

Quantitative methods compare reference data (ground-truth) and the fusion results from which error matrices and related measures such as overall, Producer's and User's accuracy, and Kappa coefficient are derived. However, in the case of using high spatial resolution data, some of the general problems related to this procedure are amplified and need even more attention compared with the use of lower resolution data. The underlying reasons are discussed in context of geometric and semantic aspects of the high spatial resolution data (Schiewe and Gahler, 2008).

From a *geometric* point of view, the smaller pixel sizes may lead to a suitable reference with appropriate positional accuracy being difficult to achieve. Furthermore, an adaptation of the number and size of sample units has to take place. In particular, the conventional acquisition on a per pixel basis is not suitable anymore due to excessively small elements and neglect of the neighbouring environment. In analogy to the object-based interpretation approach, a per-object sampling seems to be necessary in order to define training and test elements; however, due to lack of suitable methodology, such an approach is hardly ever applied in practice.

From a *semantic* point of view, high-resolution data allow for the extraction of more thematic details and object classes so that a more complex classification scheme becomes necessary, and that inherits a greater chance of overlapping definitions of attribute value ranges. As a consequence, this may lead to errors or ambiguous assignments during the visual or automatic interpretation process. The greater number of possible classes also makes more sampling units necessary. As with geometric properties, it is also difficult to find a suitable reference with appropriate thematic details and semantic accuracy. It has to be kept in mind that very often a reference data set is nothing other than another classification result based on another, essentially lower resolution, data set.

Finally, the spatial variance within regions containing a topographical object is increased in high-resolution imagery, which leads to more objects, more mixed objects (e.g., forest consists of trees, bare soil and other land covers) and more boundaries. Pertaining to the latter, the number of indeterminate boundaries, in other words the fuzziness effect, is again increased.

Due to the increasing importance of remotely sensed data with high spatial resolution on one hand, and the problems discussed above on the other hand, there is a great necessity to develop uncertainty measures for classification that consider uncertainties in reference data as well as indeterminate boundaries (Schiewe and Gahler, 2008). Most of the studies use site-specific (i.e. point based) accuracy assessment as the appropriate means to judge the classification quality of the outcome (Lang *et al.*, 2010). However, when working with OBIA, point-based method cannot assess the spatial dimension of object delineation (Lang, 2008; Albrecht, 2008). OBIA enables more complex representations of the world but is challenged to evaluate the validity of objects, especially in operational settings. In fact, evaluating the resulting objects with a binary assessment and judging whether these objects are 'correct' or 'incorrect' will fail in terms of capturing the full dimension of object validity (Albrecht, 2008). In this context, Lucieer and Stein (2002) distinguish between existential and extensional uncertainty of delineated objects. Moller *et al.* (2007) developed validation algorithm which quantified localised segmentation inaccuracies and allowed the assessment of segmentation results as a whole. The problem with this process is that it only concentrated on geometric accuracy assessment not the thematic accuracy of the object-based classification results.

2.12 Summary of Reviews

This chapter has reviewed the principles of different data fusion techniques in remote sensing, which provide a basis for selecting an appropriate fusion technique for mapping major landscapes such as urban and forest landscapes. The choice of a particular fusion depends on the appropriate scale or spatial resolution of the data, we desire information extracted from the scene and the spatial structure of the scene itself. A hierarchical landscape classification scheme not only retains the typical

properties or conditions of the desired classes but also guides the overall fusion process. A summary of the literature review is presented in Table 2. 5 highlight the different components of current fusion research.

Table 2.5 A summary review of the literature presenting major components of the current research.

Description		Reference	Characteristics	Limitations		
Pixel-level	General	Pohl and van-Genderen (1998)	Comprehensive review of pixel-level fusion	Does not include other fusion levels		
	Data-driven	Unmixing fusion	Mezneda <i>et al.</i> (2010)	Linear spectral unmixing approach using ETM+, ASTER VNIR and SWIR data for mineral detection.	Only deals with fusion of spectral imagery	
		Principal Component Analysis (PCA)	Richards and Jia (2005)	PCA for indentifying bush fire-damaged areas from multi-temporal Landsat images	No accuracy assessment	
		Unsupervised classification	Haala and Brenner (1999)	Unsupervised classification of multispectral imagery with LiDAR data for the extraction of buildings, trees and grass-covered areas.	Used low spatial resolution of the LiDAR data and had no quantitative accuracy measure	
	User-driven	Maximum likelihood	Schistad-Solberg <i>et al.</i> (1994)	Landsat TM images and ERS-1 SAR images	Various sources cannot be described by a common 'spectral' model	
Comparison		Meyer <i>et al.</i> (1996)	Compare parallelepiped and maximum likelihood classification techniques for identification of forest species	Manually separated the tree cover area		
Feature-level	OBIA general		Blaschke (2010).	Comprehensive review of the OBIA literature	Limited discussion on fusion context	
	Data-driven Segmentation	Region-based	Geneletti and Gorte (2003)	Applied a region-based segmentation	Segmentation does not deliver homogeneous areas	
		Watershed	Soille (2003)	Theory of the morphological watershed segmentation	Not much discussion in fusion context	
			Chen <i>et al.</i> (2005)	Applied marker-controlled watershed segmentation method for isolating trees from LiDAR data	LiDAR data only, no fusion related discussion	
	User-driven	Segmentation	Baatz and Schape (2000)	Multi-resolution segmentation	Only theoretical aspects presented	
		OBIA Classification	Flanders <i>et al.</i> (2003)	Classify Landsat ETM+ image for logged forest block delineation	Narrow focus on cur block delineation	
			De-Kok <i>et al.</i> , (1999)	First demonstrated the usefulness of OBIA-based fuzzy classification	No comparison was made with pixel-based one	
			Van-der-Sande <i>et al.</i> (2003)	Applied IKONOS-2 imagery for flood damage assessment	No comparison with pixel-based fusion	
Fusion	Zhou and Troy (2008)	Object-oriented approach for characterising urban landscape structure using high-resolution digital aerial imagery and LiDAR data	Only applied knowledge base rules no data-driven approach			
Comparison of pixel- and feature-	Classification	Geneletti and Gorte (2003)	High-resolution panchromatic and Landsat TM imagery	The segmentation did not deliver homogeneous areas		
		Kamagata <i>et al.</i> (2005)	Land cover mapping using high resolution satellite data	Comparison made solely on visual inspection		
		Walter (2004)	Investigated a change detection approach based on both pixel- and feature-level classification using airborne multispectral imagery	Did not use LiDAR but mentioned it would improve results		
		Platt and Rapoza (2008)	Used four spectral bands from IKONOS satellite image for mixed urban- landscape classification	Did not included any data-driven component of the classification		
	Fusion	Zhou <i>et al.</i> (2009)	Used multi-sourced data fusion for shadow classification	No broad landscape classification		
		Riggan and Weih (2009)	Multispectral SPOT-5 satellite images with high-spatial resolution aerial imagery using Feature Analyst software for object-based classification	Not much discussion of the segmentation process		
		Wang <i>et al.</i> (2004)	Used pixel-, feature- and hybrid- methods using IKONOS 1-m panchromatic and 4-m multispectral images to map mangroves	Concentrated on forest classification		
Accuracy assessment	General		Foody (2002)	Overview of the error matrix-based accuracy assessment	Discussion paper only	
			Congalton and Green (1999)	Detailed discussion of sample design, sample size and overall accuracy assessment	Discussion limited to error matrix-based assessment	
	Point or pixel-based accuracy using error matrix			Zhou <i>et al.</i> (2009)	Used a total 200 random points for assessment with 30 points per class	Very low sample size per class
				Platt and Rapoza (2008)	300 random points were generated then 249 points were supplemented by stratified samples	Inconsistent sample size-highest 138 and lowest 31
	Object-based only			Albrecht, 2008	Object boundary validation using Object Fate Analysis (OFA)	Interpretation of the accuracy was difficult
				Moller <i>et al.</i> (2007)	Validated localised segmentation inaccuracies	Concentrated on geometric accuracy assessment did not cover thematic accuracy
				Schiewe and Gahler, 2008	Considered uncertainties in reference data as well as indeterminate boundaries	Did not include sensitivity to uncertainties

The investigation of the fusion process included pixel-level fusion, like colour-related methods, numerical methods, statistical methods, signal processing-based methods and classification methods, as well as the feature-level fusions including different segmentation and feature classifications methods.

Colour-related fusion methods only enhance the colour composition therefore fusion results are only suitable for visual interpretation. Red-Green-Blue (RGB) and Intensity-Hue-Saturation (IHS) colour transformations are the most common colour-related fusion techniques. Numerical methods produce fused images by combining pixel values from multiple input images using some function or formula to give a new output pixel value. Statistical methods such as principal component analysis (PCA) can maximise the effect of the high-resolution data in the fused image. Signal-processing based fusion includes different high-pass filters and wavelets techniques that utilise different multi-scale approaches in pixel fusion.

Classification is one of the pixel-level fusion techniques that uses the multi-dimensional data vector to include different data sources. Some pre-processing is required to implement this stacked-vector approach. Standard pixel-level classifiers can be divided into two groups: data- and user-driven approaches. The data-driven approach includes *unsupervised* technique, which detects clusters of pixels in feature space and categorise the pixels into the clusters based on the statistical patterns inherent in the data. The most problematic task when using this technique is to separate different classes for which *a priori* knowledge is required. *Supervised* approaches fall into the user-driven category and overcome this problem by using *a priori* knowledge from training information for classifying multi-source images. A set of reliable signatures is created using the training data to analyse each pixel independently. There are many classifiers found in the literature to carry out the remainder of the classification process. The maximum likelihood classifier is one of the most popular classifiers. However, some data are unable to be represented by normal class models, and computational cost may increase with data dimensionality in maximum likelihood classifier. A nonparametric classification technique such as parallelepiped classifier could overcome these problems, since it depends only on the application of threshold components of the multi-source data vector.

Feature-level fusion is gaining attention due to the availability of robust segmentation algorithms and the reduction of computational expense. Feature-level fusion starts with the crucial initial step of grouping neighbouring multi-source pixels into meaningful areas or segments. A huge number of segmentation algorithms can be found in the literature but not all of them are robust and operational enough to use in feature-level fusion. Edge-based segmentation algorithms are easy to apply on grey images but struggle to delineate objects of interest from segmented images. Although thresholding is a very old, simple and popular segmentation technique, typically it leads to results of a relatively limited quality. Region growing segmentation is a method of feature extraction in which regions (features) in an image are grown from seed points. Two factors determine the final result: the parameters for region growing, and the location of the seed points. This method basically depends on the set of given seed points and often suffers from a lack of control in the break off criterion for the growth of a region. Region splitting and merging is an alternative to the region growing procedure. The split and merge segmentation algorithm works by successively splitting the image into quadrants, and merging those neighbouring quadrants which are similar, until no further splitting or merging can take place.

The watershed transformation is a segmentation tool provided by mathematical morphology. It is a combination of region growing and edge detection techniques. It groups the pixels around regional minima, and their location are along the crest lines of the gradient image. Marker-controlled segmentation is one of the most convenient, fast and powerful techniques in watershed transformation. However, noise and strong contrasts can lead this method to achieve inappropriate results.

Multi-resolution segmentation is a bottom up region-merging technique starting with one-pixel objects. In numerous subsequent steps, smaller image objects are merged into bigger ones. In this algorithm, a scale parameter is used to control the average image object size and homogeneity criteria are defined by the spectral and spatial information. Heterogeneity in multi-resolution segmentation is defined by the colour and shape of the features in the image.

Fuzzy set theory provides useful concepts and tools for classifying the features from multi-source segmented images. It also deals with the uncertainty of derived land cover maps from remotely sensed data. The key to the derivation of fuzzy land cover maps relies on defining appropriate fuzzy membership functions. Two broad approaches, namely the semantic import (SI) model and the similarity relation model, can be employed to define the fuzzy membership. For classifying the segmented features, a typical method of defining fuzzy memberships is the SI model. In this algorithm, fuzzy membership values for each feature belonging to all candidate classes are generated by expert knowledge. The selection of class boundaries and class intervals can be an objective or subjective process, depending on the way classes are defined.

Pixel- and feature-level approaches can be compared in a qualitative or quantitative manner. The qualitative measures are mainly depended on the visual inspection and vary with personal choice. Quantitative approaches are composed of rigorous sampling and statistical methods that provide and utilise numerical values by which to make comparisons. The concerns of the pixel-based accuracy assessment are reviewed as a precaution for further improvement in the implementation phase.

2.13 Chapter Conclusions

The purpose of this chapter was to introduce the subject of image fusion in order to highlight some of the aspects of LiDAR and optical image fusion. It is accepted that a knowledge of image fusion should lead to a better understanding of how to select and apply appropriate fusion technique specifically to fuse passive and active imageries. A full description of pixel-level image fusion can be found in Pohl and van-Genderen (1998), but unfortunately there is no similar single reference which fully describes the feature-level fusion. However, a comprehensive review of the OBIA development can be found in Blaschke (2010). After considering different fusion models, it was decided that this study would concentrate on the data- and user-driven models using aerial imagery with LiDAR data. This study also includes a comparison of pixel- and feature-level fusion techniques in order to select an appropriate fusion technique for mapping a range of landscape features. The error

matrix derived from field and fused data helps to evaluate the accuracy of fusion results in a quantitative manner.

The following chapter of this thesis introduces the study area and the datasets that are used to implement different pixel- and feature-level fusions.

CHAPTER 3

STUDY AREA AND DATASETS

This chapter describes the location and characteristics of the major landscape types in the study region. The reasoning behind fusing aerial imagery with LiDAR data is discussed, then the datasets of LiDAR, colour and multispectral imagery are introduced. In addition to imagery, the use of topographic maps and field survey data are discussed.

3.1 Introduction

It is very difficult to make general statements on the suitability of fusion techniques. The decision on which fusion technique to use is very much driven by the application proposed. The choice of datasets is also application dependent, as is the technique used to fuse data (Pohl and van-Genderen, 1998). In this research, the study area and the selected landscape features to delineate govern the selection of both datasets and fusion techniques. The review of scale effect and standardised landscape classification schemes in Section 2.2 and 2.3 revealed that there is a direct relationship between the spatial resolution of the data and the identification of the landscape objects. In this context, a hierarchical landscape classification scheme was selected to guide the fusion process for extracting landscape objects from multi-source data. The following sections describe the study area and the datasets required for delineating a range of landscape objects using different fusion techniques.

3.2 Selection of the Study Area and Fusion Model

A range of landscape objects was investigated when fusing aerial imagery with LiDAR data for mapping landscape. Not all landscape objects were found in the same location and some fusion techniques were not suitable for mapping some landscapes. In this context, the most dominant landscape types - forest and semi-urban areas - were selected from the study region. The selected study areas were situated in and around Mathoura, a country town in southern New South Wales. The

forest study area was a part of Moira State Forest and semi-urban study area was in Mathoura township. The location of the study areas is shown in figure 3.1.

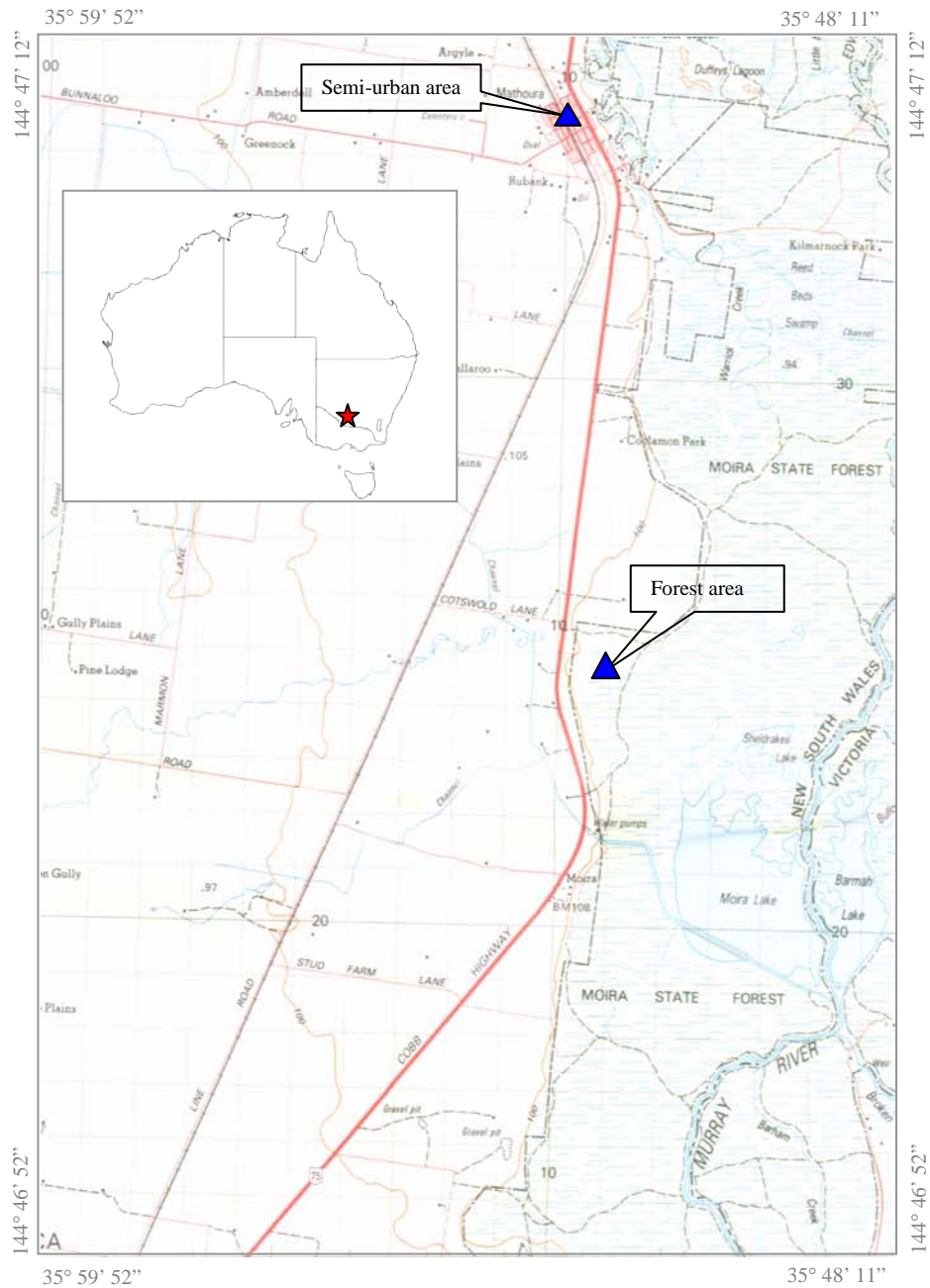


Figure 3.1 Location of the study areas

The forest study area is predominantly red gum (*Eucalyptus camaldulensis*) forest but it is also wetland because of the mosaic of open water bodies, swamps, meadows, and marshes that occur within it. The study area covers 1.25km x 1km or 1.25 square

kilometres of the forest comprising: river red gum (*Eucalyptus camaldulensis* ssp. *obtusata* Dehnh), black box (*Eucalyptus largiflorens*), and grey box (*Eucalyptus microcarpa*). A detailed review of the forest species identifying from field surveying and remotely sensed data is presented in Section 3.7.

The 1.25 km x 1 km semi-urban study area is the representative a typical Australian regional town with buildings ranging from very large to small. The area has a mixture of vegetation cover, open space, and road networks. A comprehensive description is given in Section 3.8.

Selection of representative landscape domains to be identified through image fusion in any study area requires consideration of a range of factors, including:

- (i) access to an archive of multi-source remotely sensed data for analysing attributes of different fusion processes,
- (ii) convenient access to the study area for compilation of field reference data for verification of fusion results,
- (iii) preferably a rich history of study by previous researchers in order to compare results,
- (iv) availability of ancillary data such as analogue and digital maps, and high quality aerial photographs for compilation of additional reference data.

A data-driven fusion model is dependent on data itself and used when little else is known about an area before attempting classification. This model was suitable in the forest study area because there were only three dominant *Eucalyptus* species for use to classify with high spatial resolution spectral, texture, and height data. On the other hand, users closely control a user-driven fusion model by inputting knowledge rules into the process. This model was suitable for the semi-urban study area that contained a large number of classes (six classes) with complex landscape structures therefore; only data is not sufficient enough to separate semi-urban classes.

Compatible field reference data are essential for comprehensive analysis of primary remote sensor information. The extent of the study areas and the selected landscape

classes were sufficient for testing different fusion models using multi-source remotely sensed data. High spatial resolution digital aerial imagery and field survey data provided detailed reference data for assessing thematic accuracy of the imagery-LiDAR fusion results.

3.3 Geography of the Study Region

Mathoura is located in the Riverina region 768 km southwest of Sydney, 34 km south of Deniliquin and 97 m above sea level. The area depicted in Figure 3.1 lies between 35° 48' 11" and 35° 59' 52" and 144° 46' 52" and 144° 47' 12". Mathoura's main claim to fame is that it is the gateway to the world's largest red gum forest (The Barmah-Millewa forest). The major landscape types within the town are open spaces, native vegetation, residential, commercial and industrial properties and surrounded mainly by agricultural and forest lands.

3.4 Major Landforms of the Study Region

Approximately 25,000 years ago, an uplift of land in the southern Murray-Darling Basin created what is now known as the Cadell Tilt (Cadell Fault). The edge of the 12m high block runs north/south near Deniliquin and Echuca (Chong, 2003). The Tilt is visible on the eastern side of the Cobb Highway where the ground slopes down to the forest. It influenced the course, pattern and character of about 500km of the River Murray.

Following the uplift, a large shallow lake was created by the dammed Murray and Goulburn rivers. The Murray took a new course around the northern side of the Fault (now referred to as Wakool channel), the bed of which is today occupied by the Edwards River. For thousands of years the Goulburn River continued to feed the lake but it eventually also broke out to the west.

Around 8000 years ago, the Murray turned south, breaking through the section between Picnic Point and Barmah (taking over the Goulburn channel downstream of Echuca). This section is today known as the Barmah Choke. During major floods, large volumes of water bank up behind the Barmah Choke, flooding the former lake area. This flooding has created a wetland surrounding the Barmah-Millewa Forests,

which contains flora and fauna that are typical of a region which receives three times more rainfall than the region actually does (MDBC, 2006).

3.5 Climate of the Study Region

According to DCE (1992), the climate of the Barmah-Millewa Forest region is 'temperate hot summer'. Rainfall across the central Murray Valley decreases from southeast to northwest as the influence of the eastern highlands diminishes. Winter rainfall is typically of low intensity, whilst summer falls are usually the result of heavy thunderstorms. Mean annual rainfall is about 400-450mm (Chong, 2003). Average rainfall statistics for Echuca, Barmah and Tocumwal for the past 10 years are presented in Figure 3.2.

Annual evaporation for the region is approximately 1400mm, almost half of which occurs between December and February, and two thirds between November and March (DCE, 1992). The hot season commences in November and extends until March, with mean maxima and minima 31°C and 15°C in the hottest month of January and mean maximum and minimum temperatures in the coldest month of July 13°C and 4°C (BOMa, 2006; BOMB, 2006; BOMc, 2006). The average frost-free period lasts about 7 months from late October to mid May. Severe frosts are limited to a period of eight weeks in June, July and August (DCE, 1992).

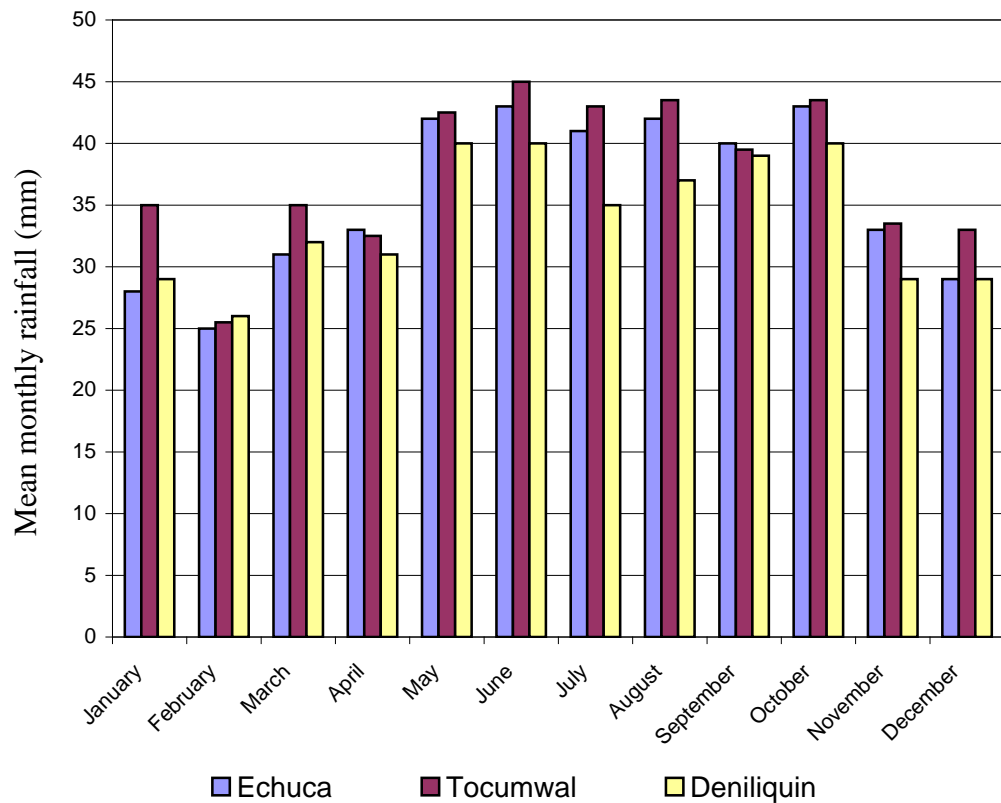


Figure 3.2 Rainfall statistics in the study region, Source: BOM (2006)

3.6 Characterisation of the Landscape Types

As explained in Section 3.2, two landscape types (forest and semi-urban areas) were selected for the analysis due to their dominant characteristics being representative of the whole Mathoura landscape. According to the geological landforms (see Section 3.4), the semi-urban study area in Mathoura is on the high land of the Cadell Tilt while the forest study area is on the down slope of this Tilt. The natural landscape has been extensively cleared particularly the native vegetation of the high land and replaced by agricultural and semi-urban land uses. However, scattered pine plantations, grazing, horticulture and significant conservation reserves also exist in this landscape. The down slope of the Cadell Tilt retains the bushland, which is dominated by native forest. In addition to bushland and water catchment reserves, some cleared corridors to the Barmah-Millewa forest exist in this landform.

Table 3.1 Major landscape types in the study region

Study area	Major landscape types	Description
Forest area	black box (BB)	a slow growing tree with rough bark on trunk and branches
	grey box (GB)	a tree with fine, pale, fibrous bark on trunk and large branches
	river red gum (RRG)	a large fast growing, spreading tree with pale grey to white smooth bark
Semi-urban area	open space/road	open yards, sealed and unsealed roads
	grey/colour roofs	commercial and Residential buildings
	tree/grass	recreational areas characterised by grasses, shrubs and occasional native plants
	shadows	building and tree shadows

Secondary data source (topographic maps, land use maps) are employed and random field visits are made to the study areas to characterise the dominant landscapes before selecting part of Moira State Forest and Mathoura Township to represent the forest and semi-urban landscapes. A summary of the major landscape types of the study areas is presented in Table 3.1 and discussed in detail in the following sections.

3.7 The Forest Study Area

In Australia's southeast riverine native vegetation survives only in remnant patches. River red gum (*Eucalyptus camaldulensis*), black box (*Eucalyptus largiflorens*), and grey box (*Eucalyptus microcarpa*) are common tree species found in the Moira State Forest. Figure 3.3, presents a composite view of these tree species using aerial and terrestrial photographs. The RGB composite of the aerial photographs was derived from infrared, red and green channel of the multispectral aerial photographs and the terrestrial photographs were taken with a digital camera. A detailed description of the aerial data source is given in the later sections. The characteristics of these three tree species from the taxonomic perspective are given below though aerial and terrestrial photographs captured the tree species on a macro-scale comparatively on a continental scale.

3.7.1 River red gum (RRG), *Eucalyptus camaldulensis*

River red gum is the most widely distributed of all eucalypts being found in all Australian mainland States along watercourses and on flood plains. Figure 3.2 Plate: 3, shows a typical RRG tree in the forest study area. RRG is a fast-growing tree, commonly 20m tall and occasionally reaching 45 m, with a diameter of 1-2 m or more (Wilson, 1995). It has smooth bark, ranging in colour from white and grey to red-brown, shed in long ribbons (Plantnet, 2009a). In adult trees, the leaves are lanceolate (length 4 times its width and broadest towards the tip), 8-30 cm long, 7-20 mm wide, green or grey green; with a leaf stalk 12-15 mm long. The buds are 4-6 mm long, 3-6 mm wide, with a hemispherical, pointed or conical cap (operculum). The 7-11 flower buds form an umbel (individual flower stalks of similar length forming a cluster) (Holliday, 1969).

RRG grows in warm to hot, sub-humid to semi-arid climates. Up to 20 frosts annually may occur in southern and inland areas. Mean annual rainfall is mainly in the range 250-600 mm. In low rainfall areas it relies on seasonal flooding and/or a high watertable. Rainfall has a winter maximum in the south and is monsoonal in northern Australia. *Eucalyptus camaldulensis* occurs mainly on heavy clays in southern Australia and sandy alluvial soils in the north. It is found on salt lake margins but is not adapted to calcareous soils. RRG is typically a riverine species and in arid areas has a ribbon-like distribution across the landscape. It also occurs on flood plains in open-forest or woodlands dominated by other eucalypts (Wilson, 1995).

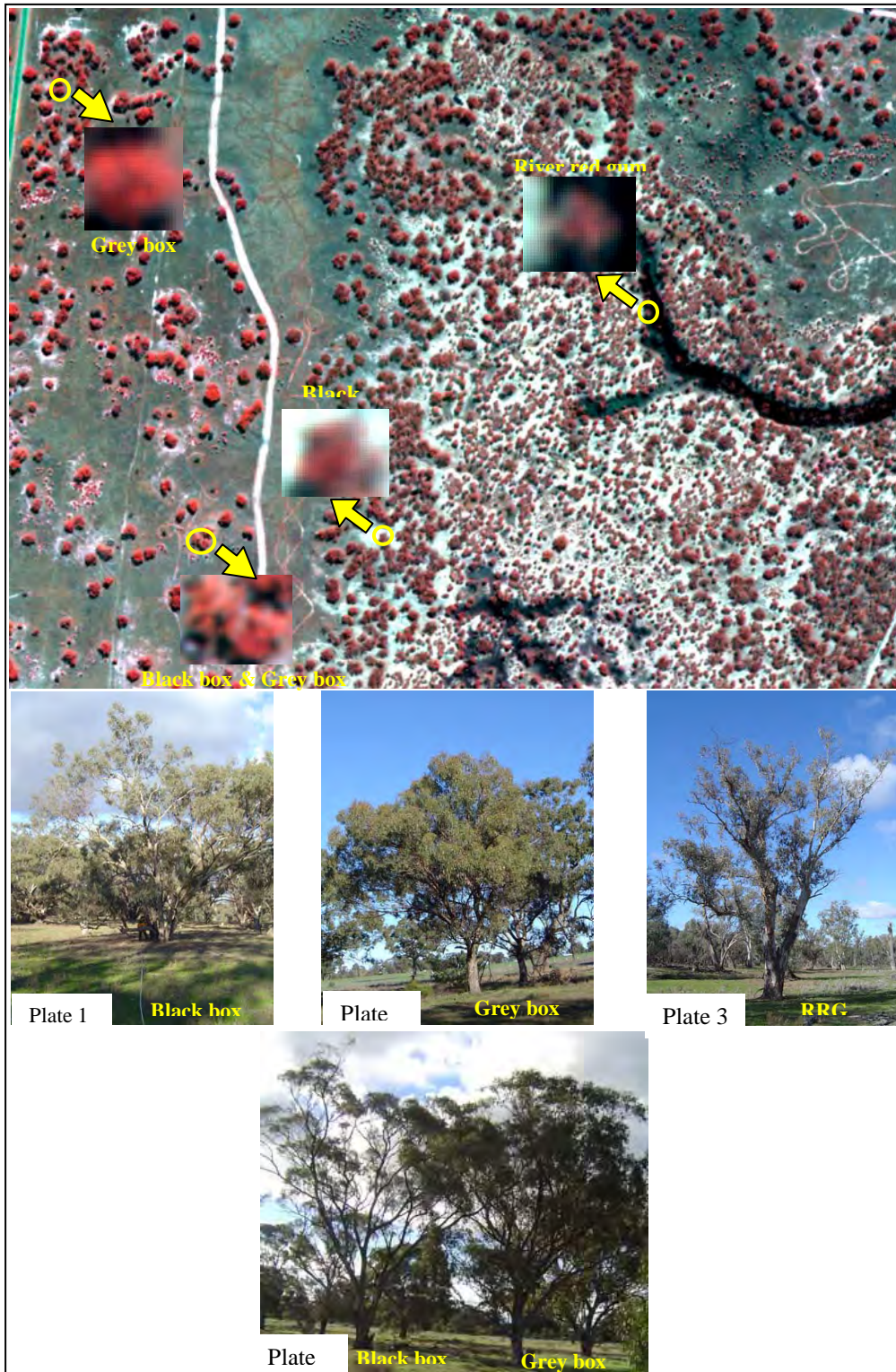


Figure 3.3 Characterisation of the dominant trees in forest study area

3.7.2 Black box (BB), *Eucalyptus largiflorens*

Black box is a slow growing tree with rough bark on the trunk and branches. An example of a black box (*Eucalyptus largiflorens*) tree in the forest study area is shown in Figure 3.3, Plate 1. The adult leaves of Black box (*Eucalyptus largiflorens*) are lanceolate, petiolate, and more or less symmetric with size of 18 cm x 18 mm, alternate, and dull green. The juvenile leaves are linear, petiolate, to 15 x 1 cm, dull blue-green, alternate. The buds are club-shaped, to 5 x 3 mm, pedicellate, in groups of 7-11, on a slim peduncle. The operculum is rounded and conical shape. The fruits are wineglass-shaped with size of 6 x 5 mm, pedicellate. The natural habitat is in river flats that are dry in summer and inundated in the winter (Ewart, 1931). Usually on clay or clay loams and found in poor drainage areas with low rainfall and high summer temperatures (Holliday and Hill, 1974). Mature height is normally 18 metres and is considered a low-branched tree. Diameter at breast height is up to 1 metre (Muell, 2007).

3.7.3 Grey box (GB), *Eucalyptus microcarpa*

A medium sized grey box tree attains height of 20 to 30 metres and 1 m in stem diameter. The trunk is generally straight and has good form. A finely tessellated grey coloured box-type bark is persistent to the base of the branches, where it changes to a smooth light grey bark which is often shed in ribbons. An example of a grey box (*Eucalyptus microcarpa*) tree is shown in Figure 3.3, Plate 2.

GB has a fine, pale, fibrous bark on the trunk and large branches. In adult trees, the leaves are lanceolate, petiolate, slightly asymmetric, 13 cm long, 25 mm wide. In juvenile trees, the leaves are ovate to broadly lanceolate, petiolate, to 15 x 5 cm, dull green. Buds are narrowly ovoid, to 9 x 4 mm, pedicellate, in groups of 7, on a slim peduncle; operculum rounded, conical. The fruits are narrow wineglass-shaped, to 7 x 5 mm, pedicellate; valves 3-4, at level of rim (Plantnet, 2009b).

GB occurs throughout the central and northern coastal areas of New South Wales and Eastern Queensland, from Jervis Bay in the south to the Atherton Tablelands in the north. It is common in the wheat belt areas of Victoria, New South Wales and Queensland. It also has a limited occurrence in the Flinders and Mt. Lofty Ranges of

South Australia. This tree is normally found in sheltered sites near rivers or in open, poorly drained, relatively flat country (Holliday and Hill, 1974).

3.8 The Semi-urban Study Area

The semi-urban study area is within the regional town of Mathoura, gateway to the world's largest red gum forest. It covers an area of approximately 1 square kilometre, which includes the town centre and its mainly commercial and residential buildings. As shown in Figure 3.4 the dominant landscape types of Mathoura township are characterised by semi-urban areas, which includes commercial buildings with large corrugated iron roofs (Fig 3.4 Plate 1); office buildings with tile roofs surrounded by asphalt/concrete road (Fig 3.4 Plate 2); and residential buildings surrounded by trees and grass land (Fig 3.4 Plate 3 & 4). The western part of the study area is comprised of open space and large commercial buildings while residential areas dominate the eastern part of the study area. Vegetated landscapes (trees and grass) exist in several locations within the study area. These landscape objects were identified during field visits and correlated with their spatial patterns in aerial photographs before collecting the reference data for the evaluating of thematic accuracies derived from different fusions. Brief descriptions of the semi-urban landscape objects are presented in the following sections.

3.8.1 Open space/road

Open space and roads are characterised by their own spatial pattern. Large commercial buildings mainly surround the open space. These spaces are bare ground due to the heavy use by commercial vehicles. Open spaces also include unsealed parking areas and yards. The main roads are sealed but minor roads and lanes are unsealed. An example of sealed road and open space is shown in Figure 3.4, Plate 3.

3.8.2 Roofs

The roofs of the study area such as those on commercial, office and residential buildings were identified and further categorised by their colour. The commercial buildings are built with corrugated iron sheets, which are normally grey in colour. An example of a grey colour roof in the semi-urban study area is shown in Figure 3.4,

Plate 1. Residential building roofs are made of varieties of material including corrugated iron and tiles. The colours of these roofs vary with the different roofing materials. However, the majority of residential roofs are grey and there are a few of red ones.



Figure 3.4 Dominant landscapes in the semi-urban area

3.8.3 Tree/grass

Vegetation in the semi-urban study area is dominated by grass, shrub and tree cover in that order, mainly in recreational areas such as playgrounds and front and backyards of residential houses. Plate 4 in Figure 3.4, shows the vegetation composition on part of the semi-urban study area.

3.8.4 Shadows

In Section 2.2.1, the inclusion of Shadow classes into urban landscape classification scheme was described. The delineation of shadows is critical for interpreting high spatial resolution imagery. Shadows normally obscure the natural illumination of areas and create darker areas surrounding elevated objects. Creation of shadow is dependent on an object's position relative to the sun angle and sun rotation. As a result, acquisition time of the aerial photography is crucial to minimise the shadow effects. In this research, shadows were delineated as part of the semi-urban landscape description. Shadows were not only detected using spectral signature but also identified using their association with the object. Buildings and trees can create large shadows due to their elevated position in the landscape.

The review of selected landscape features reveals that a single source data will not be able to delineate all landscape features. There is need to consider multi-source data fusion methods due to the complex nature and geometry of selected landscape objects. The following sections describe the justification for data selection procedures and later introduce the data properties for better understanding of their role in fusion processes.

3.9 Data Selection for the Fusion Models

The complementary nature of aerial imagery and LiDAR data was the reason for their selection in this data fusion study. LiDAR data provides an accurate measurement of landscape structure in the vertical plane. However, LiDAR sensors have limited coverage in the colour spectrum. By contrast, aerial imagery provides extensive coverage of landscape classes in the colour spectrum but is relatively insensitive to variation in object height. As a result, the fusion of aerial imagery with LiDAR data has the potential to significantly improve the identification and measurement of landscape objects. The strong argument in favour of fusion is that the LiDAR measurements are not distributed homogeneously and usually have gaps between them. As a result, the three-dimensional structure of the objects might not be very well defined (Baltsavias, 1999). It thus becomes fairly complex to obtain a good 3D model of the landscape objects with a low density of LiDAR returns. The idea of exploiting the complementary properties of LiDAR and aerial imagery is to extract

semantically meaningful information from the aggregated data for a more complete surface description. The following sections introduce the LiDAR and aerial sensors and describe their properties.

3.10 Acquisition and Properties of LiDAR Data

LiDAR is a relatively recent development in remote sensing with great potential when used to measure range (distance) to and reflectance from objects on the earth surface. Like radar it is an active remote sensing technique, which means the sensor provides its own source of energy or illumination. This leads to a number of advantages over conventional passive remote sensing systems, the main advantage being the ability of the sensor to acquire images at any time of the day or night.

LiDAR system development goes back to the 1970s and 1980s, with an early NASA system and other attempts in USA and Canada. Then, the GPS solution of the critical positioning problem made high accuracy performance feasible. Thorough investigations at Stuttgart University from 1988-1993 with a laser profiler proved the high geometric accuracy potential, especially for DTM generation, and clarified the essential system parameters (Ackermann, 1999). The way was clear for new scanning systems, which then followed in quick succession. The method has successfully established itself in the last few years, and quickly spread into various practical applications.

3.10.1 Principles of LiDAR remote sensing

In contrast to microwave radar techniques, lasers have two principal advantages for range measurements. Firstly, high-energy pulses can be realised in short intervals and secondly, at their comparatively short wavelength electromagnetic pulse can be highly collimated using a small aperture. For these reasons, shortly after the advent of lasers, very precise ranging was carried out with this new tool. As soon as lasers with high pulse repetition rates were available on the market, laser scanning systems could be realised with the ability to obtain range images (Wehr and Lohr, 1999).

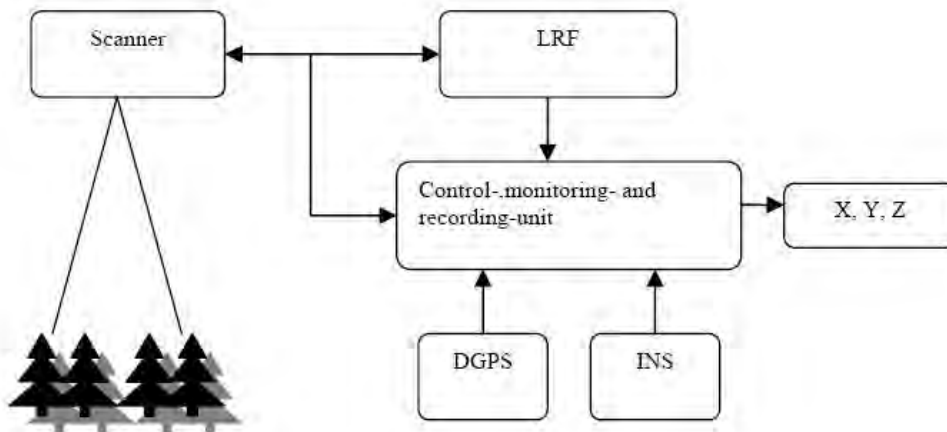


Figure 3.5 A typical laser scanner system producing three dimensional data (XYZ coordinates), based on Wehr and Lohr (1999)

The development of LiDAR remote sensing has been technology driven. It becomes initially possible by pulse lasers operating in the near infrared, which give clearly recordable return signals after diffusion and reflection from the ground. The travel times are recorded to nearly 10^{-10} second and converted to distance. Recently, continuous wave (CW) lasers have emerged, which obtain range by phase measurements. Precise kinematic positioning of the platform by differential GPS and inertial attitude determination by an IMU (Inertial Measurement Unit) or an INS (Inertial Navigation System) provide the accurate reference to an external coordinate system. Laser scanning systems furnish geometric results in terms of distance, position and attitude and coordinates. For each pulse, the spatial vector from the laser platform to the point of reflection is established, thus providing the XYZ coordinates of the laser footprint.

The laser scanning system is a combination of the positioning and orientation system, the scanner, and the Laser Range Finder (LRF) units (Figure 3.5). The LRF measures the distance from the laser scanner aperture to the reflection surface. The pulse direction relative to the scanner (scan angle) is known for each laser measurement. The position and orientation of the laser scanner system are known at any time during the mission and are determined by combining a Differential GPS and an INS. Data from all units are time marked and stored in the control-, monitoring-, and recording-unit. The data from the different sources are linked using the time mark during the post-processing, and XYZ coordinates are produced.

3.10.2 LiDAR range measurement

The time history of the reflected energy pulse is fully digitised and converted to units of distance through knowing the speed of light in the atmosphere. The *first pulse* return above a threshold is used to derive the distance to the object top and the *last pulse* return is used to find the range to ground: subtraction then yields the laser-derived object height.

The return waveform pulse gives a record of the vertical distribution of nadir-intercepted surface. At any particular height, the amplitude of the waveform measures the strength of the return. Thus, for surfaces with similar reflectance and geometry within a footprint, a larger amplitude indicates more object material and a smaller amplitude less (Drake *et al.*, 2002). The waveform provides only an apparent object profile because of attenuation of the pulse through the object so it must be adjusted to approximate the true object profile (Lefsky *et al.*, 1999).

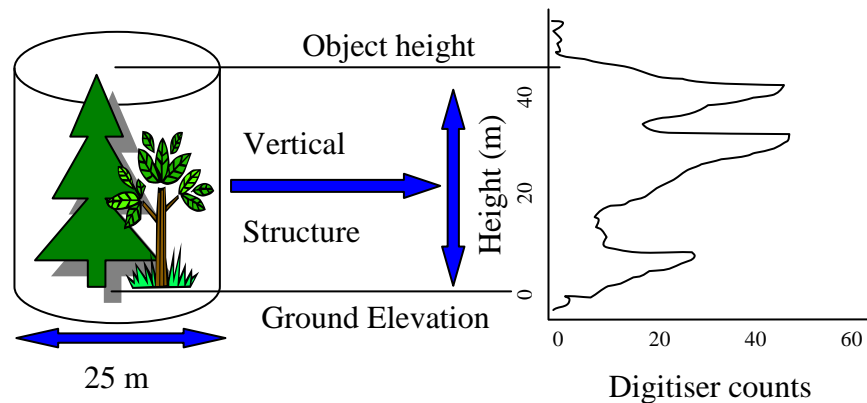


Figure 3.6 Conceptual basis of LiDAR range measurement.

The conceptual basis for LiDAR range measurement is illustrated in figure 3.6. The incident Gaussian-distributed pulse of laser energy from airborne instruments reflects off various portions of the objects, resulting in a return waveform where the amplitude of the pulse is a function of the area reflecting surfaces at that height. The entire waveform gives the vertical distribution of surfaces intercepted by the incident pulse. Some of the incident laser penetrates all the way through the canopy to produce the last amplitude Gaussian-shaped spike in the waveform known as the ground return. LiDAR systems do not measure canopy height, but rather a target

range determined by measuring the travel time of the pulse. The object height is determined by subtracting the range to the ground from that to the first detectable return or some threshold above that return (Drake *et al.*, 2002).

A laser pulse that is fired over an object usually has multiple reflections. Some of the laser pulse may be reflected by the top of the object, resulting in the returned signal being registered by the sensor as the *first pulse*. The remainder is likely to penetrate the object and eventually be reflected by the ground. The *last pulse* registered by the sensor corresponds to the lowest point from where the signal was reflected. In certain cases, the difference in elevation between the first and last pulses can be assumed to be a measure of the height of the object.

Along with the time of transmission of the signal from the sensor to the object and back to the sensor, the intensity of the returned laser pulse may also be registered by LiDAR systems. Since LiDAR systems typically operate in the near infrared (NIR) part of the electromagnetic spectrum, the intensity can be interpreted as an NIR image. However, the quality of the intensity image depends on the sensor specification and the flying height. Generally intensity image is under-sampled and thus very noisy, because the footprint is of the order of 0.2m, while the average sampling point distance is typically 1-2m.

3.10.3 The properties of a LiDAR system

The LiDAR data used for this research was acquired by AAMGeoScan (now AAMHatch) in July 2001. The LiDAR system was the ALTM 1225, which operates with a sampling intensity of 11 000 Hz at a wavelength of 1.047 μm . Approximate flying height of this sensor was 1100m and the laser swath width was 800m. Vertical accuracy was 0.15m (1σ), the internal precision was 0.05m, and the original laser footprint was 22cm in diameter and point spacing in the order of 16 points per m^2 .

The original LiDAR dataset was processed by AAMHatch and provided to the Victorian Department of Sustainability and Environment (DSE). In the processing phase, the data was thinned to remove superfluous points not adding to the terrain definition. It also included the conversion of the original LiDAR point heights into a regular 1m grid. The data was provided as ASCII format with two separate files

representing the first (non-ground) and last (ground) return point data. The data format consisted of Easting, Northing, Elevation and Intensity fields. The data was provided in 2km by 2km tiles and each tile had approximately 11million points. The supplied LiDAR data was compliant with the GDA (Geocentric Datum of Australia) system. The horizontal datum was GDA94 and projection was MGA (Map Grid of Australia) Zone 55. The vertical datum was AHD (Australia Height Datum) and geoid model was Ausgeoid98. This research-specific LiDAR processing is further described in the Chapter 4.

3.11 Acquisition and Properties of Colour Imagery

In this research, different fusion techniques were developed and tested for the fusion of aerial imagery with LiDAR data for landscape mapping. The aerial colour and multispectral images were collected with different sensors and platform. The colour imagery was acquired with a Zeiss LMK152 aerial camera system by AEROMETREX (www.aerometrex.com.au). The focal length for the colour imagery was 152.261mm. The platform was flying at ~1000m above ground level with a heading of ~188°. These images were accompanied with camera calibration certificate and the exterior orientation parameters ($X_o, Y_o, Z_o, \omega, \phi, \kappa$), captured using onboard GPS and IMU sensors for each photo. The analogue colour images were scanned with a flatbed scanner at 15 μ m to provide a pixel size of ~8.25cm. The radiometric resolution of the colour image was 8-bit. The orthorectification process of the colour imagery is presented in Chapter 4.

3.12 Acquisition and Properties of Multispectral Imagery

Digital multispectral imagery was captured by IFMS Germany (www.arcforest.com) over the study area using an Ultracam-D with a calibrated focal length of 101.4mm. Ultracam uses a set of 8 optical cones to assemble a large format digital image in natural colour (red, green and blue) with false colour infrared (IR). The panchromatic images were collected with a 9.00 μ m pixel size and multispectral images were collected with a 28.125 μ m pixel size. The multispectral images were collected as a medium format image and they were later upscale to panchromatic image format. The radiometric resolution of the digital images was 16-bit. This increased

radiometric range of digital image relative to analogue captures more detailed information of the landscape features. As a result, in extreme bright and dark areas we still manage to get redundant information beyond what is visible in images with lower radiometric resolution (Leberl and Gruber, 2005).

Table 3.2 Summary descriptions of the datasets

Image Type	Year of Acquisition	Bands	Spatial resolution	Radiometric resolution
LiDAR data	2001	First & last return	16 point/m ²	8-bit
Colour image (Scanned)	2001	3 (R, G, B)	0.08m	8-bit
Multispectral image	2004	4 (R, G, B, IR)	0.88m	16-bit

The UltracamD datasets also came with camera calibration certificate and the exterior orientation parameters ($X_o, Y_o, Z_o, \omega, \phi, \kappa$), captured using onboard GPS and IMU sensors. The geometric correction procedure of the multispectral imagery is presented Chapter 4. Table 3.2 lists the acquired datasets.

3.13 Other Data Sources

In addition to aerial remote sensing data, digital orthophotographs and topographic maps were used to select the reference data for fusions themselves and were later used to evaluate the thematic accuracies derived from different fusions. The ancillary data were used for generating descriptive statistics of the multi-source data fusion. Extensive field data was also collected to validate the accuracy of the fused dataset.

3.13.1 Supporting digital aerial photography collections

In April 2006, small format aerial photography was collected throughout the study region to enhance the datasets for fusion process. The RedLake MS4100 high-resolution 3CCD multispectral camera was used with a calibrated focal length of 13.20mm. The aerial photographs were collected at three different scales by varying the flight heights according to the spatial resolution of the images. The flight height

was ~650m for the 50cm pixel images. Four band (CIR) digital aerial photographs were used as an additional source to improve the orthorectification of the colour and multispectral imagery.

3.13.2 Field data collection

Field visits were undertaken during July 2005 and April 2006. The training sample sites were surveyed in 2005 and sample data for assessing the accuracy of fusion results were collected in 2006. In addition to these surveys, in July 2003, a group of experts from RMIT University and NSW National Parks and Wildlife visited this region to assess the tree condition. The 2003 study spanned 6km x 4km area and forty test sites were initially selected using a stratified random sampling method and subdivided into trips and points. The location of individual trees was identified using handheld GPS. In areas where the identification or singling out of individual trees was not possible stem density was attributed in a plot size of ~50m². This study recorded detailed descriptions of the individual tree structures as well as vertical strata attributes at each sample site. Out of these forty sample sites seven sites were located in the forest study area (Figure 3.9).

In April 2006, an extensive field survey was conducted to measure the position of the GCPs for orthorectification of the aerial imagery. Twenty temporary GCPs were erected in the forest study area and their positions using DGPS were recorded. Thirty-two GCPs locations were recorded for the semi-urban study area. Detail description of the GCPs location and orthorectification of the aerial imagery are presented in the following Chapter.

During the 2005 survey, an additional seven sample sites were selected from the forest study area using a stratified random sampling technique. Detailed landscape surveys were conducted at each of the sites for selecting training data and assessing the thematic accuracy of the fusion results. Each site consisted of a 50m x 50m (2500 m²) plot. Within each sites all individual tree position, height, and species were recorded into a structured survey form. A copy of the forest study area survey form is shown in Appendix 1. A total of 76-tree was photographed and their heights were measured using a clinometer for LiDAR-derived height validation purpose.

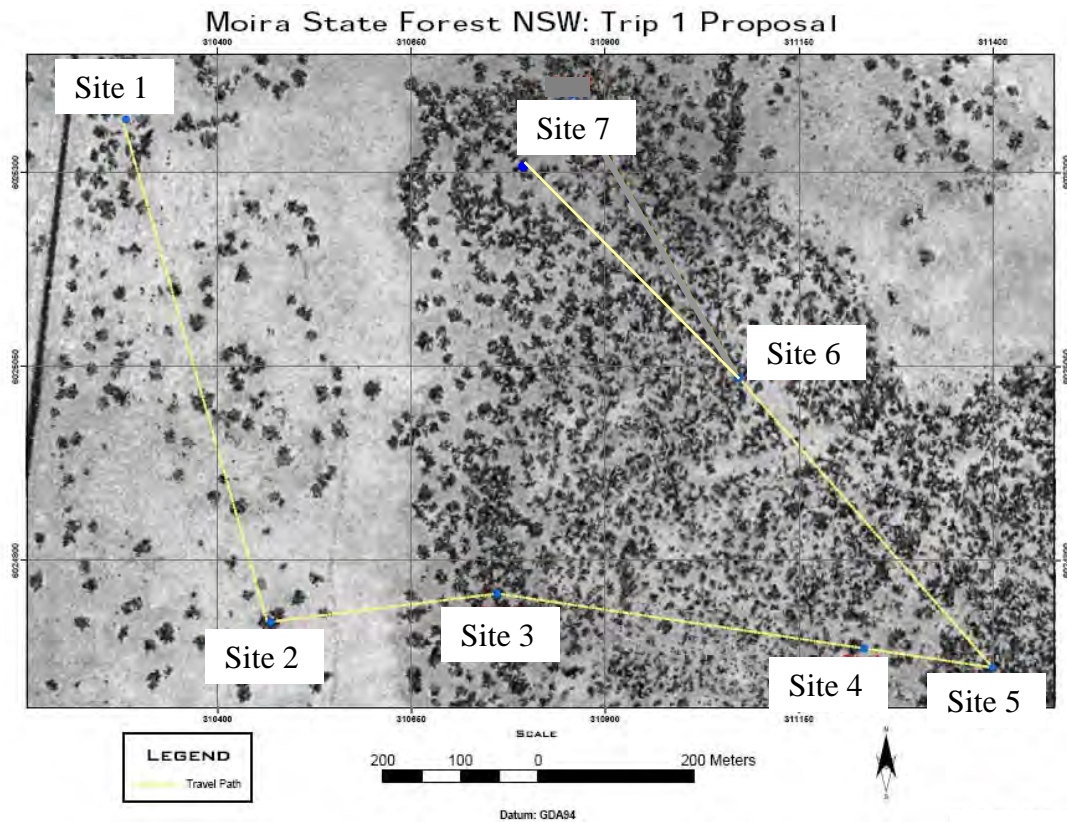


Figure 3.7 Sample locations for the forest site

Location of the trees was recorded using a GPS receiver (Pathfinder Pro XRS, Trimble Navigation). The GPS receiver was configured using Universal Transverse Mercator (UTM) coordinates system with GDA94 datum and three-dimensional (3D) mode of operation. In 3D operational mode, the Pathfinder GPS triangulates a position when four or more satellites are connected. Positional accuracy was checked using the Position Dilution of Precision (PDOP). It provides the possible errors related to the geometry of the satellites that are used to triangulate a position on the Earth. Good accuracy is obtained when the PDOP ranges between 1 and 4, implying field error location between 1 to 10m (Kennedy, 2002). Accuracy lower than $\pm 2\text{m}$ was found for the Pathfinder GPS in an accuracy assessment between the measured easting and northing of four survey marks in the surrounding study area. NSW Land Survey provided true locations of the survey marks. Average PDOP values ranging between 2 and 4 were obtained, which was considered acceptable.

A field data collection form was designed which included site identification number, calendar date, X, Y, Z coordinates, feature height and landscape type. A copy of the

semi-urban study area survey form is shown in Appendix 2. An optimal number of high quality natural colour terrestrial images were acquired using the digital camera for each sample site in addition to documentation of the visual identification of the landscape composition using the field data collection form.

3.14 Summary

A study region was chosen, which contained a variety of landscape objects particularly the forest and semi-urban landscapes. The active and passive airborne sensor-derived data were acquired and fused for delineating these landscape objects from different fusion models. The complementary nature of aerial imagery and LiDAR data was the main reason for the selection of these two remote sensing devices.

Detailed descriptions of the study region including physical attributes, geographic location and climate conditions were collected. The study region consisted of two main landscape types: forest and semi-urban. The forest landscape consists mainly of natural objects, river red gum, grey box (*Eucalyptus microcarpa*), and black box (*Eucalyptus largiflorens*) trees. The semi-urban study area consists of different natural and man-made objects. Open spaces, vegetation, different types building roofs and their shadows were the main objects in this landscape.

Airborne colour and multispectral images are to be fused with LiDAR data using different fusion models. LiDAR data was acquired by AAMHatch with the ALTM1225 system. This LiDAR system collects *first* and *last* returns data with intensity values. The initial processed LiDAR data was provided, as ASCII format with a 1m spatial resolution. Three-band colour (RGB) aerial imagery was collected by AEROMETREX with Zeiss LMK152 camera. The four-band (RGB & IR) aerial multispectral imagery was collected by IFMS with the UltracamD system. Other data sources, such as small format aerial photograph and field survey data were also collected as a reference for fusion validation purposes.

Geometric alignment is a prerequisite for any fusion model using multi-source data. Geometric corrections of the aerial imagery and LiDAR data are presented into the following Chapter.

CHAPTER 4

GEOMETRIC CORRECTION OF THE MULTI-SOURCE DATA

In the previous chapter the selected study areas and the datasets were introduced; now the geometric correction procedures are reviewed and the best-suited procedure used with these datasets. Geometric processing consists of two steps: orthorectification of the aerial imagery and derivation of a normalised digital surface model (nDSM) from the LiDAR data. These stages are essential precursors for the fusion of aerial imagery with LiDAR data. Orthorectification provides a common planimetric base for the fusion of different remote sensing data (Zitova and Jan, 2003). The orthorectification technique is directly affected by the quality and the selection of GCPs (Yastikli and Jacobsen, 2005). Image resampling is critical to the maintenance of the spectral and spatial quality of the datasets (Ehlers, 1997). The quality of the orthorectification and resampling processes underpin the subsequent data fusion. Although the LiDAR data for this study did not require orthorectification, it needed further processing for generating absolute height of the landscape objects.

4.1 Introduction

Analysis of multiple sources of remotely sensed data is dependent on the ability to accurately relate corresponding locations in each image through spatial referencing. The basic requirement is for a geometric match to be established between data sources and for relevant pixel values to be transferred according to the derived geometric relationships. Where data sources comprise multiple images, the geometric process is termed registration or, alternatively, the process for combining image and planimetric data utilising a reference coordinate system and DTM is termed orthorectification. Transferring individual pixel values to the registered or rectified image is called image resampling (Ehlers, 1997).

A comprehensive survey of image registration methods was published in 1992 by Brown (1992). According to the database of the Institute of Scientific Information (ISI), in the last 10 years more than 1000 papers have been published on the topic of image registration (Zitova and Jan, 2003). Brown (1992) covered all the registration

related classic or introductory key ideas that are still in use. Zitova and Jan (2003) cover relevant approaches introduced later and give a complete view of research in image registration.

The requirement for orthorectification arises for a number of reasons such as multi-source data fusion, change detection, and integration of remotely sensed data with GIS (Buiten and van Putten, 1997; Fonseca and Manjunath, 1996; Kardoulas *et al.*, 1996; Dowman and Dare, 1999). Refinements in orthorectification accuracy also have potential for improving results achieved from some standard information extraction approaches (Wolter *et al.*, 1995).

The influence of orthorectification accuracy on data fusion is well documented. Ehlers (1991) indicates that improving rectification accuracy allows depiction of cartographic detail and enhances spatial resolution in multi-sensor image datasets. In vegetation monitoring, only a 0.2 pixel rectification error causes a 10 percent change in Normalised Difference Vegetation Index (NDVI) between epochs (Townshend *et al.*, 1992). Martin (1989) indicates that displacements between images of only 0.5 pixel can introduce unacceptable levels of error. A focus on rectification methods and quality is therefore significant in the development of appropriate multi-source fusion process. The following sections review different image rectification and resampling methods and later sections present the implementation of these processes for the multi-source datasets.

4.2 Review of the Image Rectification Methods

Image rectification models may be categorised as either *parametric* or *non-parametric* methods (Richards and Jia, 2005). Parametric methods are designed to model the nature and magnitude of distortions inherent within the image and to devise specific correction formulae. In contrast, non-parametric approaches rely on establishing an analytical relationship between the image pixels and the corresponding coordinates on the ground (Richards, 1993).

4.2.1 Non-parametric rectification methods

Non-parametric approaches to geometric correction, such as polynomial transformations, are based on generic functions not directly related to specific distortions or error sources. They can be successful when dealing with low resolution, narrow field of view imagery, such as some satellite imagery (Yang, 1997). Non-parametric functions are very simple to implement and provide a reasonable alternative to geometric modelling when little is known about the geometric nature of the image. However, non-parametric techniques generally process images one at a time. They cannot provide an integrated solution for multiple images. It is very difficult for non-parametric techniques to achieve reasonable accuracy without a great number of GCPs. Misalignment is more likely to occur when mosaicking separately rectified images.

4.2.2 Parametric rectification

Parametric models overcome all the problems mentioned above by using least squares bundle block adjustment, and create the most reliable ortho-images from the raw imagery. They are unique in terms of considering the image-forming geometry, utilising information between overlapping images, and explicitly dealing with the elevation dimension. Parametric models can process multiple images with very few GCPs, while at the same time eliminating the misalignment problem associated with image mosaics.

The processes of parametric rectification are based upon a thorough knowledge of the physical characteristics and magnitude of the errors and distortions inherent in the remotely sensed data. Sources of error comprise elements of the sensing system (interior orientation), platform parameters (exterior orientation) and target (representation of the ground). Some satellite systems cannot provide the detailed data required for implementation of the parametric approach, however aerial photogrammetry systems can provide the required data by combining global positioning system (GPS) and internal navigation system (INS) onboard the platform (Ehlers, 1997). The main advantage of the parametric approach is high precision and robust rectification (Pala and Pons, 1995).

Well-known obstacles in digital photogrammetry include defining the interior and exterior orientation parameters for each image using a minimum number of GCPs. Due to the costs and labour intensive procedures associated with collecting ground control points, a limited number of GCPs should be used. Additionally, airborne GPS and INS techniques provide an initial approximation to exterior orientation, but the final values for these parameters need to be adjusted to attain higher accuracies.

Interior orientation defines the internal geometry of a camera or sensor, as it existed at the time of data capture. The variables associated with image space are defined during the process of interior orientation. Interior orientation is primarily used to transform the image pixel coordinate system to the image space coordinate system. The internal geometry of a camera is defined by the principal point, principal distance and lens distortions. The principal point is mathematically defined as the intersection of the perpendicular line through the perspective centre of the image plane. The distance from the principle point to the perspective centre is called the principal distance (Wang, 1990). Lens distortion deteriorates the positional accuracy of image points located on the image plane. Two types of lens distortion exist: *radial* and *tangential* lens distortion. *Radial* lens distortion causes imaged points to distort along radial lines from principle point. *Tangential* lens distortion occurs at right angles to the radial lines from the principal point. Since tangential lens distortion is much smaller in magnitude than radial lens distortion, it is considered negligible. The principal distance of a camera and the effects of lens distortion are commonly determined in a laboratory during the camera calibration procedure.

Exterior orientation defines the position and angular orientation associated with an image at the time of exposure or capture. Figure 4.1 illustrates the elements of exterior orientation. The positional elements of exterior orientation include X_o , Y_o , and Z_o . They define the position of the perspective centre (O) with respect to the ground space coordinate system (X , Y and Z). Z_o is commonly referred to as the height of the camera above sea level, which is commonly defined by a datum.

The angular or rotational elements of exterior orientation describe the relationship between the ground space coordinate system (X , Y and Z) and the image space coordinate system (x , y and z). Three rotation angles omega (ω), phi (ϕ), and kappa (κ) are

commonly used to define angular orientation. Omega is a rotation about the photographic x -axis, phi is a rotation about the photographic y -axis, and kappa is a rotation about the photographic z -axis, which are defined as being positive if they are counter clockwise when viewed from the positive end of their respective axis. The photographic z -axis is equivalent to the optical axis (f , principal distance). The x' , y' and z' coordinates are parallel to the ground space coordinate system.

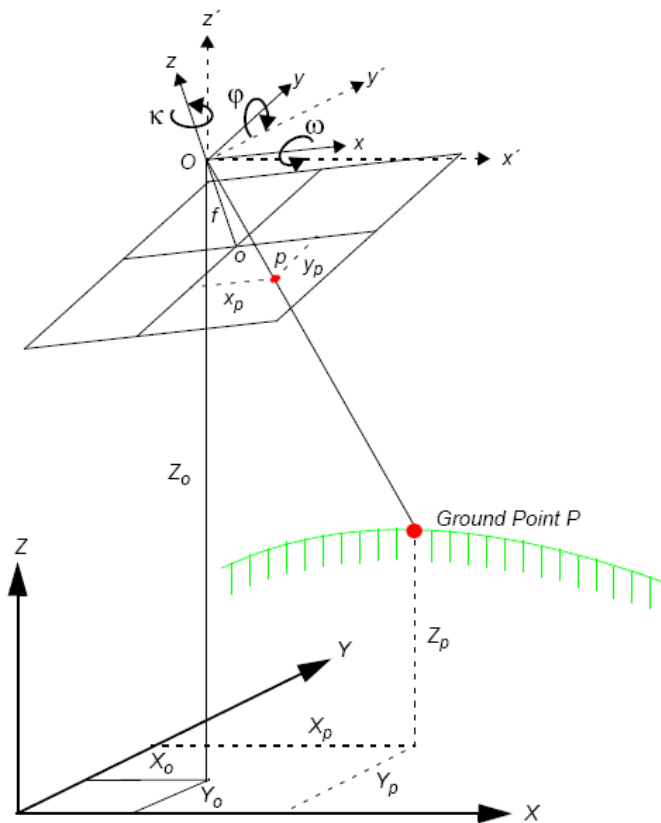


Figure 4.1 Elements of exterior orientation, after Mikhail *et al.*, (2001)

Using the three rotation angles, the relationship between the image space coordinate system (x , y and z) and ground space coordinate system (X , Y and Z or x' , y' and z') can be determined. A 3x3 rotation matrix is used to define the relationship between the two systems. The rotation matrix, M is denoted as:

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \quad (4.1)$$

The rotation matrix is derived by applying a sequential rotation of ω about the x -axis, ϕ about the y -axis, and κ about the z -axis.

Using the collinearity equation, the relationship between the camera/sensor, the image, and the ground can be defined. With reference to Figure 4.1, an image vector a can be defined as the vector from the exposure station o to the image point p . A ground space or object vector A can be defined as the vector from the exposure station O to the ground point P . The collinearity defines the relationship between image vector and ground vector if a line extending from the exposure station to the image point and to the ground is linear. The image vector and ground vector are only collinear if one is a scalar multiple of the other. The equation is as follows:

$$\begin{bmatrix} x_p - x_o \\ y_p - y_o \\ -f \end{bmatrix} = kM \begin{bmatrix} X_p - X_o \\ Y_p - Y_o \\ Z_p - Z_o \end{bmatrix} \quad (4.2)$$

where k is a scalar multiple. The image and ground vectors must be within the same coordinate system. As a result, image vector a is comprised of the following components:

$$a = \begin{bmatrix} x_p - x_o \\ y_p - y_o \\ -f \end{bmatrix} \quad (4.3)$$

Where x_o and y_o represent the principal point and x_p and y_p represent the point p in the image coordinates. Similarly, the ground vector can be formulated as follows:

$$A = \begin{bmatrix} X_p - X_o \\ Y_p - Y_o \\ Z_p - Z_o \end{bmatrix} \quad (4.4)$$

The ground vector is multiplied by the rotation matrix M , in order to bring the image and ground vector in the same coordinate system. The following equation can be formulated:

$$a = kMA \quad (4.5)$$

where:

$$\begin{bmatrix} x_p - x_o \\ y_p - y_o \\ -f \end{bmatrix} = kM \begin{bmatrix} X_p - X_o \\ Y_p - Y_o \\ Z_p - Z_o \end{bmatrix} \quad (4.6)$$

The above equation defines the relationship between the perspective centre of the camera exposure station and ground point P appearing on an image with an image point location of p . This equation forms the basis of the collinearity condition that is used in the photogrammetric project. The collinearity condition specifies that the exposure station, ground point, and its corresponding image point location must all lie along a straight line, thereby being collinear. Two equations represent the collinearity condition.

$$x_p - x_o = -f \left[\frac{m_{11}(X_p - X_{o1}) + m_{12}(Y_p - Y_{o1}) + m_{13}(Z_p - Z_{o1})}{m_{31}(X_p - X_{o1}) + m_{32}(Y_p - Y_{o1}) + m_{33}(Z_p - Z_{o1})} \right] \quad (4.7)$$

$$y_p - y_o = -f \left[\frac{m_{21}(X_p - X_{o1}) + m_{22}(Y_p - Y_{o1}) + m_{23}(Z_p - Z_{o1})}{m_{31}(X_p - X_{o1}) + m_{32}(Y_p - Y_{o1}) + m_{33}(Z_p - Z_{o1})} \right] \quad (4.8)$$

One set of equations can be formulated for each ground point appearing on an image. Therefore, the collinearity condition defines the relationship between the camera, the image and the ground.

It has already been mentioned that airborne GPS and INS techniques only provide initial approximations of the exterior orientation. The final values of these parameters need to be adjusted to attain higher accuracy. Using the GCPs, it can be achieved, but the cost associated with the collection of GCPs is high. To minimise the costs fewer GCPs are used in conjunction with a bundle block adjustment. A bundle solution computes exterior orientation parameters of each image in a block and the X, Y, and Z coordinates of tie points and adjusted GCPs. In this approach, a block of images is simultaneously processed in one solution. A least squares adjustment is used to estimate the bundle solution for the entire block while minimizing and distributing the error. For each of the GCPs two collinearity equations can be formulated as shown in Equation 4.7 and 4.8. If a GCP has been measured on the overlapping area of two images, four equations can be written. The

overall quality of a bundle block adjustment is largely a function of the quality and redundancy in the input data. The redundancy can be computed by subtracting the number of unknown from the number of known. The resulting redundancy is commonly referred to as the degree of freedom in a solution. Once each observation equation is formulated, the least squares adjustment can be used to solve collinearity condition in a bundle block adjustment.

4.3 Review of Image Resampling Methods

Image resampling is a process which is used to determine the pixel values for the output image after registration (Lillesand and Kiefer, 1994). This process is based on the rectification parameters, extent of the image and the output pixel dimension. Each new pixel is defined by real number coordinates, which do not necessarily coincide with the pixel locations of the input image. Consequently, an interpolation procedure is required to allocate brightness values for each of the output pixels. Ehlers (1997) refers to this as a pixel-filling approach that ensures every output pixel is addressed only once during the process, and that no gaps occur in the output image.

The selection of the resampling algorithm and the specification of the output pixel resolution are the most important components of the rectification process. Selection of an inappropriate resampling algorithm can have a deleterious effect on the intensity values of the output pixels of the rectified image, and specification of an inappropriate pixel resolution can lead to degradation of the spatial quality of the data. The following sections review the issues related to the resampling process.

4.3.1 Resampling methods

Resampling techniques that can be applied for intensity interpolation include the nearest neighbour, bilinear and cubic convolution interpolation algorithms. In the following, these three different resampling techniques are reviewed.

Nearest neighbour resampling assigns the brightness value of the nearest input pixel to the output pixel located in the image. This pixel value is then transferred to the corresponding display grid location. This procedure is computationally simple,

however it introduces pixel level geometric discontinuities (up to a maximum of $\sqrt{2}/2$ pixel), making the image appear visually disjointed or *blocky*. Dikshit and Roy (1996) indicate that nearest neighbour resampling has the significant effect on the extraction of texture (spatial) features. However, others (Ehlers, 1997; Richards, 1993) argue that retention of the original values is advantageous for subsequent classification or other spectral-based processing. Both points of view need to be considered when choosing a resampling algorithm for a specific application.

In bilinear interpolation the brightness value of the output pixel is interpolated using three linear interpolations over the four pixels surrounding the input pixel. The bilinear interpolator also acts as a spatial filter that subdues extreme brightness values throughout the output image. It may provide a much more visually appealing result, but may also degrade some image detail (Ehlers, 1997).

The cubic convolution interpolation uses the nine (3x3) neighbouring pixels to interpolate the output pixels location. Interpolation is first undertaken in the y-direction for each of the four vertical lines in the matrix to determine the brightness value equivalent to the x coordinate of the output pixel. Interpolation is then undertaken in the x direction to determine the equivalent brightness value of the output pixel.

4.3.2 Resampling interval

The final dimension of pixels in the rectified image is determined by the resampling interval, which has an impact on subsequent processing. The spatial resolution at the time of image acquisition is independent of the resampling interval, and resampling to a finer pixel size does not improve the spatial resolution compared to the original data. However, there may be implications for the spatial distribution of pixels in the resampled image when the resampling interval is significantly different to that of the original data.

For comparison of the different levels of fusion processes and subsequent classification it is necessary to resample all the multi source data to a common geometric datum and pixel dimension. A suitable resampling interval must be determined, but at the same time any detrimental spatial or radiometric effects on the

data must be minimised. The following knowledge is important for the investigation of the resampling interval of the datasets in this study:

- (a) the information content of remotely sensed images is dependent on the measurement scale determined by the spatial resolution of the sensor
- (b) neglecting the resolution and aggregation level, the fusion of imagery can produce unpredictable results having little correspondence in the scene
- (c) there is no unique spatial resolution for the detection and discrimination of all targets comprised of multi-scale data.

These comments are particularly applicable to the relationship between the scene and the original pixel dimension of all data sources. However, large resampling intervals potentially aggregates and reduces the spatial resolution, which will degrade the radiometric and spatial information content of the data. The resolution of remotely sensed data should be a balance between a pixel size sufficiently large to acquire the desired information with the minimum possible data, but be fine enough to capture the variations of interest within the target (Atkinson and Curran, 1997). Spatial resolution of the sensor should be much finer than the resolution at which the maximum local variance in the target occurs because, if it is not, the spatial variation of interest in the target may be lost. In order to retain the radiometric and spatial qualities of each image it is necessary to resample at least to the same spatial resolution as the original data. The question whether the spatial distribution of the pixel values is retained during the resampling process is nevertheless important, and a pixel resolution finer than that of the original data should be selected. Conversely, resampling to large pixel size will degrade the spatial resolution and certainly alter the radiometric distribution of the pixels in the image.

4.4 Selection of the Image Rectification Method

The selection of an appropriate rectification technique is somewhat dependent on spatial resolution, where the objective is to achieve residual rectification errors less than 0.5 pixel (Labovits and Marvin, 1986). Data from aircraft and satellite platforms are subject to similar geometric distortions, however their magnitude and

significance varies (Ehlers, 1997), affecting the selection of the rectification technique. Satellite sensors are affected by the systematic effects of earth rotation and earth curvature, but are generally very stable in their altitude, attitude and velocity. In contrast, aircraft sensors suffer very little or no systematic effects from earth rotation and earth curvature, but are generally unstable in their altitude, attitude and velocity.

Limitations of the non-parametric approaches are evident due to the requirement for extensive GCPs and the lack of a physical interpretation model. In this context, the current research therefore relies on the application of the parametric model, which is suitable for the rectification of airborne datasets. While the parametric model was established as a routine approach for analog aerial photography, its application to the rectification of digital aerial photography also enables the rectification of high spatial resolution aerial data.

4.5 Selection of the Image Resampling Method

In this research, all resampling techniques are evaluated and the best-performing one selected. Since multi-source images are to be used for fusion and subsequent classification, maintenance of the original brightness values of the pixels is important. With both bilinear interpolation and cubic convolution the output brightness value of the single pixels are interpolated to different values from the original data. In addition, the nearest neighbour resampling often causes the distortion of straight edges in images and makes the image look coarser than an image which is resampled using bilinear and cubic convolution interpolation (Gonzalez and Woods, 2002). Forester and Trinder (1984) indicate that the specific resampling algorithm applied does not appear to affect subsequent percentage classification accuracy, however boundary pixels were seen to move between spatially adjacent classes as a result of change in the resampling strategy. From the computational and quality perspective, nearest neighbour resampling is the quickest to perform but generally gives the coarsest result. Cubic convolution usually gives the smoothest result, but is computationally most expensive. In the past, bilinear interpolation was seen as a good balance between quality of result and computing time; however, rapidly increasing computer power means that CPU time is not really

a major processing problem, so cubic convolution can be seen as the best resampling method. Cubic convolution resampling works well when there is a requirement for high visual quality, such as for photo interpretation, or when the image is to be viewed under magnification. This method was utilised in this research to accurately fuse the multi-source high spatial resolution data for landscape mapping. This enabled the proper integration of reference data with fused data for further thematic accuracy assessment, but at the same time minimised radiometric distortion of the data.

4.6 Orthorectification of the Aerial Imagery

The earlier sections reviewed different image rectification and resampling methods. Now, the implementation of the orthorectification process is presented. The review presented in Section 4.2 showed that parametric methods are suitable for the aerial imagery as they are designed to model the nature and magnitude of distortions inherent within the image and to devise specific correction formulae. LiDAR-derived DEM used also as a source to remove any terrain distortions.

Figure 4.2 illustrates the orthorectification process that is used for rectification of the airborne optical images for each study area. The major steps of the implemented rectification process are described in the following sections.

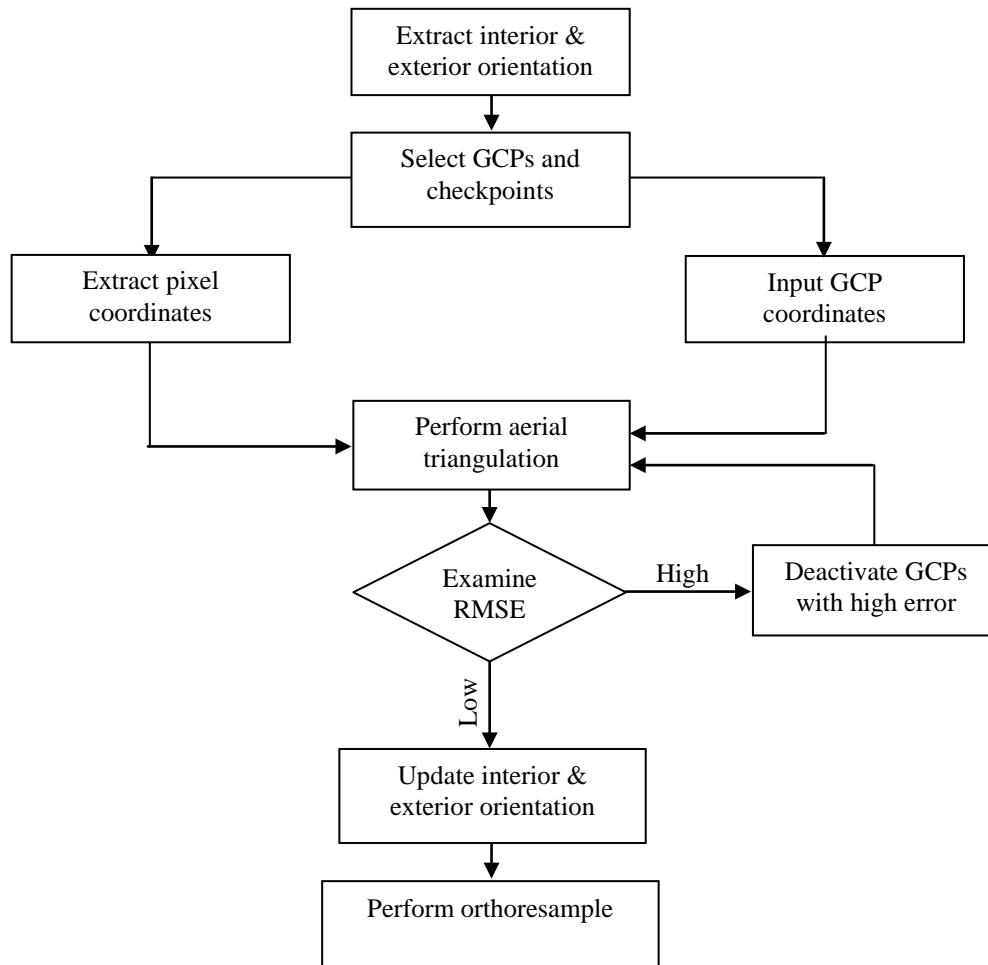


Figure 4.2 Flowchart showing the orthorectification of the aerial imagery

4.6.1 Extraction of interior and exterior orientation parameters

The interior orientation of the aerial imagery sensors came with the camera calibration certificate. Table 4.1 presents the summary of the information, which was used for this research.

Table 4.1 Summary of the camera models

Camera Name	Company	Focal length (mm)	Flying height (m)
Zeiss LMK 152	AEROMETREX	152.261	~1000
Ultracam-D	IFMS Germany	101.400	~3000

The exterior orientation parameters of the images were collected using on-flight GPS and IMU systems. The initial approximation of the exterior orientation parameters (x, y, z, omega, phi, kappa) of each image was supplied by the acquisition companies.

However, the positional (x, y, z) accuracy was within 5 meters and the attitude parameters (ω , ϕ , κ) were within 0.2 degree. With this positional accuracy, the orthorectified aerial images were not properly aligned with the LiDAR layers. An optimal number of GCPs were collected to improve the positional accuracy of the images and they were incorporated into the orthorectification process.

4.6.2 Selection of GCPs and checkpoints

The critical component in establishing an accurate transformation relationship between the camera/sensor and the ground is the GCPs. GCPs are identifiable objects located on the Earth's surface that have known ground coordinates in X, Y and Z. Horizontal control only specifies the X, Y, while vertical control only specifies the Z. Normally, intersection of roads, utility infrastructure (e.g. fire hydrants and manhole covers), intersection of agriculture plots and survey benchmarks are used as GCPs. The coordinates of the GCPs are collected using ground GPS, total station survey, and orthorectified images.

In this study, OmniSTAR differential GPS was used to collect GCP coordinates. This is a single-frequency, L1 DGPS that was particularly useful to achieve sub-meter precision for the position (X, Y) but for elevation measurement (Z value) was not as precise. In the GPS survey PDoP (Percent Dilution of Position) is used as an indicator for precision.

GCP requirements

The minimum GCP requirements for an accurate mapping project vary with respect to the size of the project. With respect to establishing a relationship between image space and ground space, the theoretical minimum number of GCPs is two GCPs having X, Y, and Z coordinates and one GCP having a Z coordinate associated with it. According to Fonseca and Manjunath (1996) the most difficult step in image rectification is the establishment of the transformation relationship between objects on the ground and in the image, whereas the computation of the mapping function is relatively straightforward. The need for careful selection of GCPs of adequate precision is highlighted by enumerating sources of GCP-related errors that can affect

the quality of image rectification as well as fusion. Important sources of error include the spatial resolution of the image, and the number and distribution of GCPs (Kardoulas *et al.*, 1996; Labovits and Marvin, 1986; Welch *et al.*, 1985). The spatial resolution of the image influences the ability to discern objects that are used as GCPs, and affects image qualities such as shape and tone. For identification purposes, control points must be bigger than the dimension of a pixel and have sufficient contrast with the background.

GCPs should be well distributed around the edges and generally over the extent of the image to ensure adequate control of the interpolation throughout the image (Richards, 1993). The success of interpolation also varies with the resampling techniques and the number of control points. These two issues were discussed in Section 4.3. According to Labovits and Marvin (1986), the orthorectification result does not vary when the spatial location of the GCPs and their cluster of pixels is not even. It is common practice for GCPs to be evenly spread throughout the image and this approach is followed in this research.

Collection of GCPs for forest area

Figure 4.3 shows the distribution of the GCPs for the forest study area. All these GCPs had to be properly located due to the lack of man-made identifiable landmarks. Their good distribution within the forest study area insured good quality orthorectification of the images. A total of 20 GCPs were collected for the forest study area.

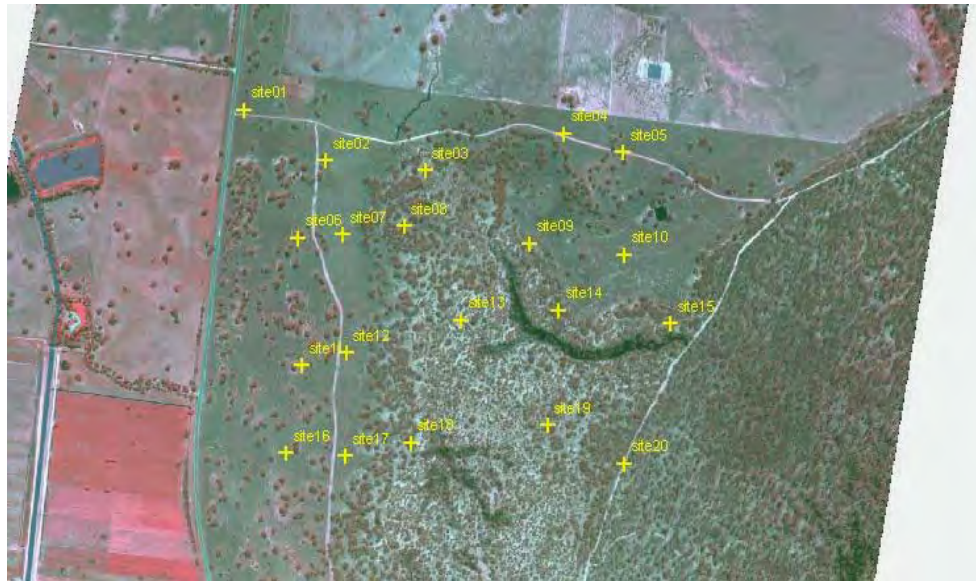


Figure 4.3 GCPs location for the forest area orthorectified images

Collection of GCPs for semi-urban area

Figure 4.4 illustrates the distribution of the GCPs for the semi-urban study area. A total of 32 GCPs were collected for the semi-urban study area. An adequate distribution of GCPs was provided in the vicinity of the study area indicated on the diagram, although the distribution was not uniform throughout the whole area. While the quality of the orthorectification cannot be guaranteed in areas not covered by GCPs, analysis shows that the number and distribution of GCPs did not affect the quality of the orthorectification.



Figure 4.4 Distribution of GCPs for orthorectification of the semi-urban area images

4.6.3 Evaluating GCPs location precision

The positional accuracy of collected GCPs using different instruments is related to the data collection and compilation standard, map scale, and the instruments' internal accuracy. Using differential or real-time kinematic GPS, one metre to centimetre level positional accuracy can be achieved whereas locating GCPs from maps depends on map accuracy as well as the digitising accuracy. In GPS survey, PDoP (Positional Dilution of Precision) is a good indicator of positional accuracy. It is a unitless figure of merit expressing the relationship between the error in user position and the error in satellite position, which is a function of the configuration of satellites from which signals are derived for positioning. Generally a PDoP value less than 4 indicates sub-meter accuracy. After collecting all GCP coordinates for each of the study areas, data were processed and the GCP precisions were measured.

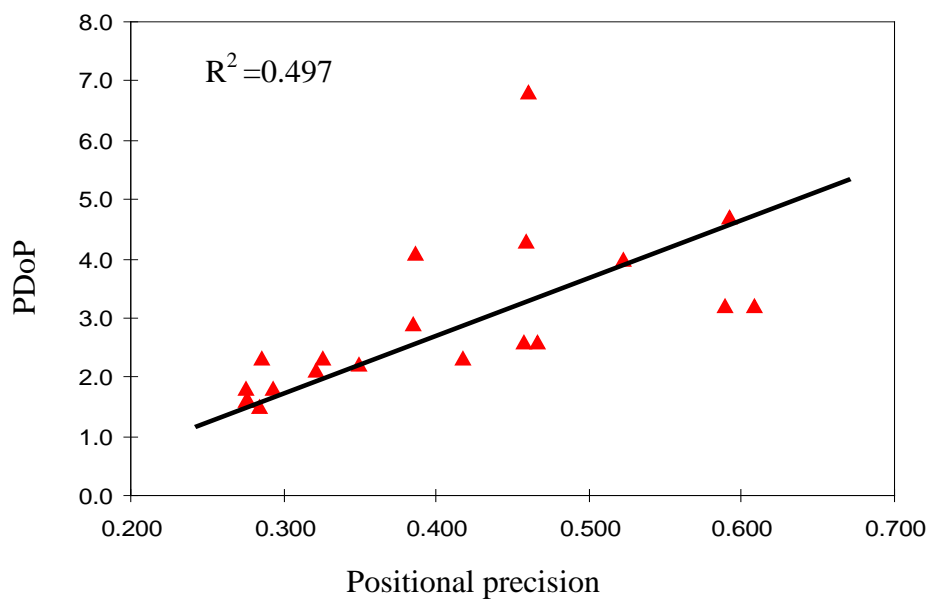


Figure 4.5 Positional precisions compared with the PDoP for the GCPs in the forest study area

A total of 20 GCPs were collected for the forest study area with 2 of them is being eliminated due to poor signal reception in the high density forest. The average PDoP value was 2.9, with the highest being 7.1 and the lowest 1.5. Figure 4.5 illustrates results of the accuracy analysis of the GCPs in the forest study area. The regression analysis utilises the positional precision and PDoP. The R^2 -test indicates the

agreement between the positional precision and the PDoP. When R^2 is close to 1, this indicates perfect agreement and close to 0 indicates disagreement. For the forest study area, the overall R^2 value was 0.497, which indicated the positional precision and PDoP had reasonable agreement and therefore GCPs had moderate positional accuracy.

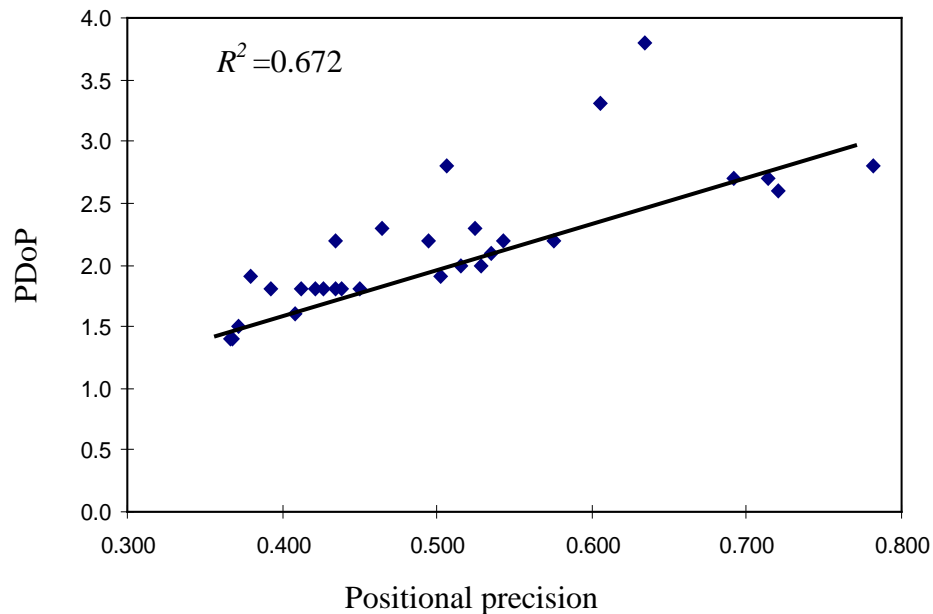


Figure 4.6 Positional precisions compared with the PDoP for the GCPs in the semi-urban study area

For the semi-urban area, a total of 32 GCPs were collected with the differential GPS. Figure 4.6 shows the scatter plot of the PDoP and the positional precision of the GCPs. The R^2 value was 0.672, which implies these two accuracy indicators had good agreement and the precision of the GCPs were higher than expected.

The comparative GCPs positional accuracy for both of the study areas indicated that GCPs in semi-urban areas were more stable than the GCPs in forest study area. This result was expected due to the GCPs physical condition and the availability of the GPS signal. The GCPs in the semi-urban study area were permanent structures and clearly visible in the collected aerial imagery, whereas GCPs in the forest study area were temporarily erected and not all of them were clearly visible in images. In addition, GPS signal was consistently strong in the semi-urban area and had less signal-to-noise ratio and multi-path problem due to the absence of high structures

within the GCPs location. In the forest study area, finding a clean opening for a GCP was the big challenge. Most of the GCPs location had some sort of multi-path problem due to the close proximity of forest trees. As a result, the signal-to-noise ratio and multi-path problem were high in the forest study area.

4.6.4 Performing aerial triangulation

Aerial triangulation starts with the initial approximation of the supplied exterior orientation parameters (x , y , z , ω , ϕ , κ) of each of the aerial images. Then using GCPs with underline DEM the aerial triangulation was refined and finally performed. The Root Mean Square Error (RMSE) value in the triangulation report should be reviewed as a global quality indicator. Any GCPs that had large residual values would be deactivated and generated new triangulation report for refinement. The RMSE value should be around 1 micron or less than one-fourth of a pixel. After successful triangulation the exterior orientation was updated. Finally, the resampling method and the output pixel size were defined and the final orthorectification was executed. Aerial triangulation of the images from each study area was thus performed and an independent evaluation of the quality of the orthorectification undertaken.

Table 4.2 Summary of the orthorectification results with quality evaluation (RMSE in microns)

Study area	Colour image			Multispectral image		
	GCPs	Checkpoints	RMSE	GCPs	Check points	RMSE
Forest	13	5	1.1124	11	9	1.1728
Semi-urban	19	13	0.7965	22	10	0.8825

Table 4.2 summarised the RMSE values of the GCPs used to derive aerial triangulation. The RMSE values in this table indicated that images of the semi-urban study area had more positional accuracy than the images of the forest study area. However, the accuracies of the orthorectified images were within the one-quarter of the pixel limit and had good alignment with the LiDAR data for further fusion analysis.

Sufficient GCPs must be selected in order to provide some for aerial triangulation and others as checkpoints for *a posteriori* consideration of the quality of the aerial triangulation (Kardoulas *et al.*, 1996). This permits not only evaluation of the interpolating techniques, but also assessment of the complete aerial triangulation including any effects of relief displacement and other error sources which are not considered. The major disadvantage of this assessment is that it is not possible to isolate specific sources of error contributing to the overall RMSE. In this research, RMSE from the independent checkpoints is calculated for each study area to evaluate the aerial triangulation processes. Note that some of the surveyed points were used as GCPs while others were used as checkpoints. Colour images in the semi-urban site used 19 GCPs and 13 checkpoints for aerial triangulation.

Table 4.3 Summary of residuals of exterior orientation parameters for the multispectral images of forest study area

Image ID	rXs	rYs	rZs	rOMEGA	rPHI	rKAPPA
0059	0.1862	-0.3361	-1.1197	0.0082	0.0223	-0.0039
0060	-0.4750	0.0498	0.1884	0.0063	0.0269	-0.0103
0061	0.2888	0.2863	0.9314	0.0040	0.0545	-0.0154

Tables 4.3 and 4.4 list details of the residuals of the exterior orientation parameters for the multispectral and colour images of both study sites. The residual ground coordinates (rXs, rYs and rZs) are in metres and orientation residuals (rOMEGA, rPHI and rKAPPA) are in degrees. After each iteration of the least squares adjustment, the exterior orientation parameters of each camera station are estimated. The newly estimated exterior orientation parameters are then subtracted from the original exterior orientation parameters values. The differences are the residuals of the exterior orientation parameters.

Table 4.4 Summary of residuals of exterior orientation parameters for the colour image of the semi-urban study area

Image ID	rXs	rYs	rZs	rOMEGA	rPHI	rKAPPA
1-164	0.1562	0.3368	0.1197	0.0052	0.0469	-0.0143

All the residuals shown in the tables are derived from aerial triangulation by bundle block adjustment. The objective of this residual calculation is to indicate any systematic pattern in the magnitude and distribution of residual values. Consistently large values indicate a poor fit to control or insufficient GCPs in the triangulation.

4.6.5 Performance of the resampling

In this study, the spatial resolution of aerial colour imagery was 0.08m and 0.88m for the multispectral imagery. The 16 points per square metre LiDAR return was resampled into 1.0m LiDAR imagery. It is a critical task to bring all these disparate datasets into the same pixel spacing without changing much of the thematic contents. The multispectral and the LiDAR images were coarser than the colour image; therefore, the optimal pixel size should be in the middle of these two extremes. Resampling of the multispectral and LiDAR images into a finer pixel size does not affect the original spatial details but pixel size degradation of the colour image has some impacts on radiometric and thematic properties. Reviewing the scaled-down colour image, it was found that degraded images still retained similar spatial and radiometric properties of landscape objects of interest. The cubic convolution resampling technique had been applied on the colour images to reduce the spatial resolution from 0.08m to 0.5m. Multispectral and LiDAR images had been scaled-up and also resampled into 0.50m since bringing all images into a common spatial resolution is an essential prerequisite for further fusion applications. Any changes of the spatial and radiometric quality of the data are most likely to occur when changing pixel dimension during the resampling process. In this case, colour imagery was severely affected as it was resampled from 0.08m to 0.50m.

The images under consideration contain a range of targets (see Section 3.4), which all contain unique radiometric and spatial characteristics. According to Marceau *et al.* (1994), these characteristics may be visible at different spatial resolutions. In order to maintain the maximum radiometric content and spatial integrity of the observed data from all sensors, a uniform resampling interval of 0.50m was selected. This sampling interval was smaller than that of the original multispectral image and LiDAR data but larger than that of the original colour image.

4.7 LiDAR data Processing

The processing of LiDAR data often aims at removing unwanted erroneous measurements or modelling data to fit into a specific model (Axelsson, 1999). The unwanted measurements are characterised as noise, outliers or gross errors. The most common processing of LiDAR data is removal of these unwanted measurements to find the ground surface (such as DTM) from a mixture of ground and vegetation measurements. DTM not only represents terrain but also includes other terrain parameters such as slope and aspects, terrain features such as ridges and valleys and other geographical/environmental characteristics (Qiming *et al.*, 2008). In LiDAR processing, these terrain features need to be considered for better delineation of the landscape objects and can also utilise these characteristics for further objects classification (Dragut and Blaschke, 2008). The object isolation from LiDAR data is typically based on the object height model, which is the difference between object height and a digital elevation model (DEM) of the earth surface. Chen *et al.* (2006) applied a method for filtering LiDAR data into terrain and non-terrain return and the extracted terrain pulses are used to generate a DEM by interpolation (Hyypä *et al.*, 2001; Brandtberg *et al.*, 2003).

It has already been mentioned that AMMGeoScan, the supplier of the LiDAR data, processed and provided data in two separate files representing the first and last return point clouds. The last return of the LiDAR normally represents the digital terrain model (DTM) and the first return the digital surface model (DSM). In this research, the LiDAR data was further processed to generate a normalised DSM from these two data sets. This processing determined the mean height information of the landscape objects, which is later used in the fusion process as an additional information layer.

4.7.1 Normalised DSM Generation from LiDAR Data

It was inevitable that the frequency of ground or last returns were low compared to the first returns, meaning the DTM might degrade the accuracy of the object heights (Popescu *et al.*, 2002). Additionally, the ellipsoid height accuracy of differentially corrected and processed Optech ALTM 1225 data is about 15cm. These two error sources need to be considered when measuring the mean object height. Næsset (1997) determined mean tree height as the difference between tree canopy hits and

the corresponding DTM values, though details of the algorithm used to compute the DTM were not provided. In this research, the same approach was adopted to determine the mean height of the landscape objects. A height threshold was applied to remove any outliers close to the terrain surface.

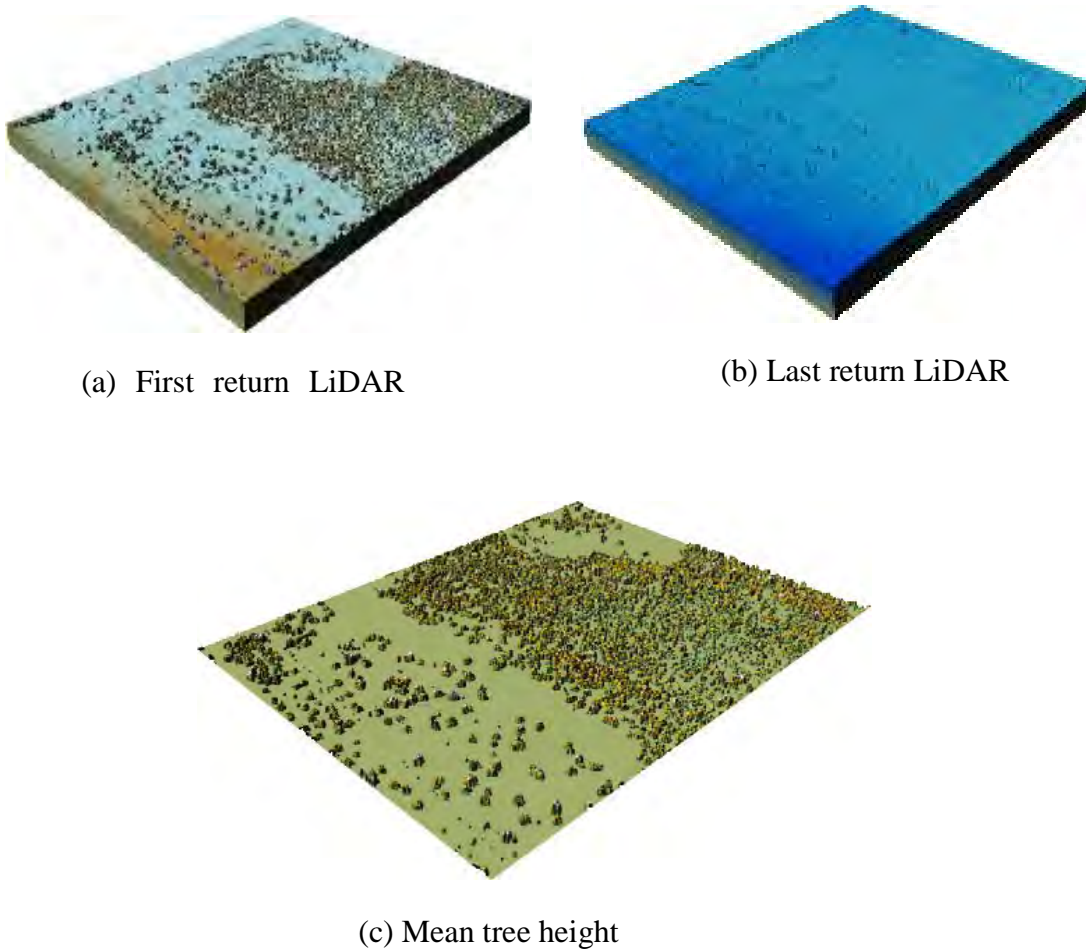


Figure 4.7 Mean tree height derived by subtracting the last return from first return of the LiDAR data

For the forest study area, the mean tree canopy height was computed by subtracting the last LiDAR return from the first LiDAR return. Observations with a height value of less than 1.5m were excluded from the dataset in order to eliminate the effect of high slope and shrubs in the forest study area. Figure 4.7 illustrates the processing of the datasets. The interpolated first and last LiDAR returns are shown in figures 4.7(a) and (b). Figure 4.7(c) shows the subtracted image giving mean tree height.

The DSM of the semi-urban area was obtained by interpolating the first LiDAR return to a regular grid. In a similar way, a DTM was generated with last LiDAR return. Figure 4.8(a) and (b) show the 3D views of the interpolated first and last return LiDAR data. The mean feature heights were computed as the difference between the first and last LiDAR return layers. A height threshold less than 1.5m was used to eliminate the effect of low-lying objects such as fences and shrubs. Figure 4.8(c) shows the final mean height layer for the semi-urban area.

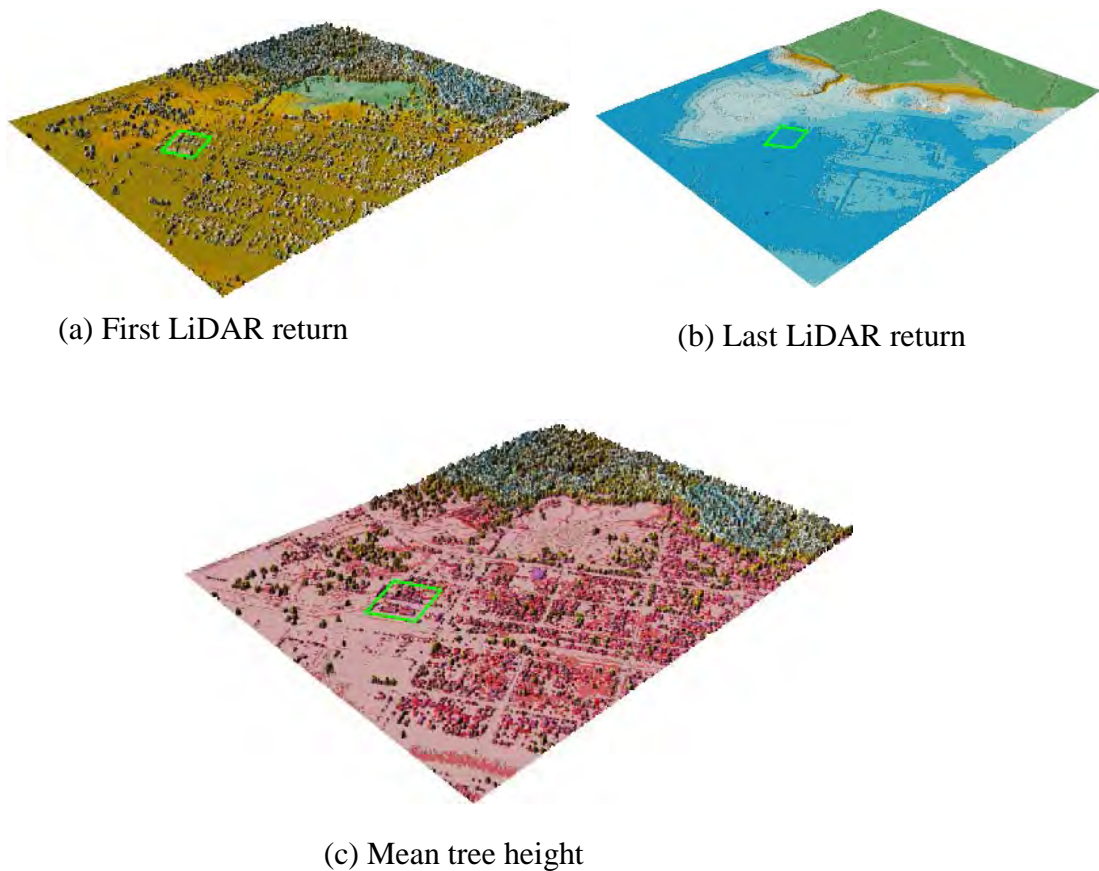


Figure 4.8 Mean semi-urban objects height derived by subtracting the last return from first return of the LiDAR data

4.7.2 Normalised DSM accuracy analysis

The mean of the LiDAR heights were compared to the ground collected height for nDSM accuracy analysis. A description of the samples and their height derivation was presented in Section 3.13.2. Regression analysis was used to assess height accuracy in the normalised DSM.

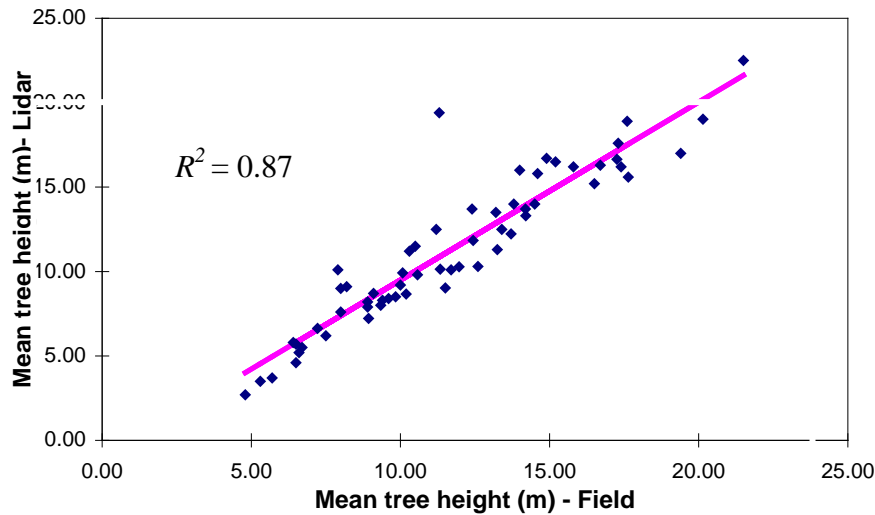


Figure 4.9 Relationship between tree heights as estimated in the field and from LiDAR data for the forest study area

In the forest study area, LiDAR-derived tree heights are compared with the field-surveyed data to validate LiDAR-derived tree heights accuracy. Figure 4.9 shows the comparison of mean tree heights derived from LiDAR data and from field measurements derived from 76 trees in the 7 plots. This analysis confirmed close correspondence ($r^2 = 0.87$, 95 percent confidence level, standard error = 0.67m). Mean height was more reliably estimated for trees with large and relatively flat crowns than for those with small and pointed crowns. The estimates of height from LiDAR and field measurement, were within ± 1 m of each other, although discrepancies as high as 3m were observed.

Similarly, a linear regression was derived for the semi-urban area using 32 randomly selected field-measured heights and corresponding with LiDAR-derived mean heights (Figure 4.10). The R^2 value of 0.91 (close to 1) indicates that for these feature heights there was very good correlation between field-measured height and LiDAR estimates.

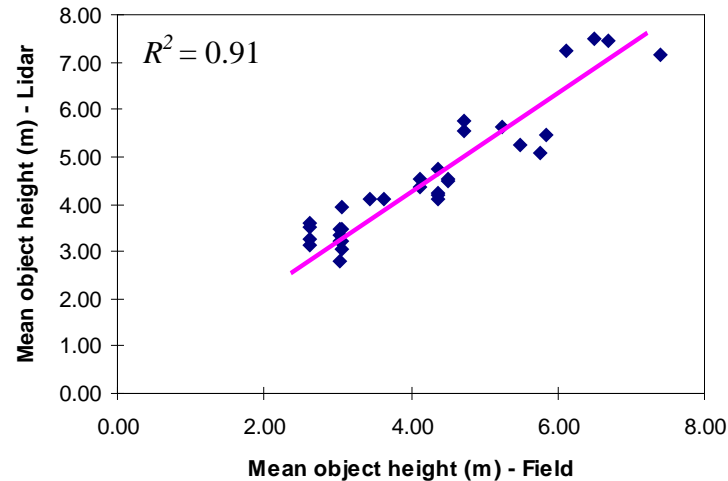


Figure 4.10 Relationship between field- measured and LiDAR-measured heights for the semi-urban study area

4.8 Discussion

Previous research indicates that for image rectification RMSE values should be less than 0.5 pixels. Table 4.2 indicates these values had been achieved in this study with all RMSE values substantially less than 0.5 pixels. From an operational viewpoint, the RMSE is of more interest because it provides a global estimate of rectification quality. The RMSE values for the semi-urban study area were particularly low at 0.7965 microns for the colour images and 0.8825 microns for the multispectral images. However, the RMSEs for forest study area were high for both colour and multispectral images. RMSEs close to 1 micron indicated that rectification accuracy was higher than one-quarter of a pixel, which was better than the standard 0.5 pixel accuracy.

The multispectral residual results of the forest study area (Table 4.3) indicated that no gross errors in GCP identification or exterior orientation determination were present. These results indicate the transformation process was geometrically sound. The relatively high RMSE for the forest study area was therefore due to the limitations inherent in the identification and positioning of GCPs in forest study area.

It has already been mentioned that all the GCPs for orthorectification were derived with the aid of differential GPS. Estimation of coordinates with a GPS depends on the precision level of the GPS receiver as well as the satellites, but identification of

GCPs on the imagery is much more difficult compared to the other data sources. Additionally, fewer GCPs were located for the forest study area; therefore higher residuals could be expected compared to other sites. GCPs were selected with a uniform distribution across the images and were based upon well-defined points identifiable in both the images and the planimetric reference data. GCPs positional accuracy was derived from GPS positional precision and PDoP attributes. The regression analysis of the positional precision and PDoP attributes indicated positional accuracy. The R^2 value for the forest and semi-urban study areas were 0.497 and 0.672 respectively. The lower R^2 value for the forest area indicated that its GCPs were less accurate than the semi-urban area. This was expected because of the lack of identifiable GCPs in the forest study area and the GPS reading had low signal-to-noise ratio.

The relatively high RMSE values for the forest study area were due to lower geometric consistency of the data and the availability of fewer GCPs for determination of the aerial triangulation. The forest site had very limited well-established GCPs and the existing maps had only provided generalised detail. As a result, these maps gave limited opportunity for selecting GCPs. To overcome this, during the data collection period, a number of temporary GCPs were built. However, these were not sufficient to cover the whole study area adequately, a compromise necessary to reduce the project cost. In the semi-urban study area hundreds of potential GCP locations could be identified in the high spatial resolution imagery. However, for the same reason not all of them were surveyed. Only handful of GCPs were collected with high signal-to-noise ratio of GPS receiver.

The accuracy analysis of the normalised DSM revealed that the object height of the semi-urban study area was more accurate than the object height in forest study area. The R^2 value for the forest and semi-urban study areas were 0.87 and 0.96 respectively. This result also suggested that the forest tree height measurement was more error prone than for semi-urban objects. Because of the random nature of tree heights, it is difficult to accurately measure them. Particularly in the forest areas, it is very difficult to find the tops of individual trees both in field survey and in LiDAR data. However, the sampled tree heights were still close to the LiDAR-derived tree heights and maintained consistent agreement with LiDAR measurements. It can be

said that normalised DSMs derived from LiDAR data truly represent the height of the landscape objects in both of these study areas.

4.9 Summary

Fusion of multi-sensor data requires geometrically coherent datasets derived through rectification and pixel resampling. With the development of digital aerial photography, orthorectification of the high spatial resolution imagery with sub-pixel precision required investigation of appropriate algorithms. All data, including the high spatial resolution multispectral and colour imagery, were orthorectified using bundle block adjustment. Investigation of this technique applied to high spatial resolution imagery indicated that a suitable precision and acceptable orthorectification results could be achieved with a minimum number of GCPs.

Bundle block adjustment uses interior and exterior orientation parameters to rectify the images. Airborne GPS and IMU systems normally provide initial approximations to the exterior orientation, but the final values for these parameters need to be adjusted using GCPs. An optimum number of GCPs were surveyed with differential GPS. For all images, aerial triangulation produced RMSE values close to 1 micron, which can be interpreted as an error equal to, or less than, one-quarter of a pixel, and in all cases, orthorectification achieved RMSE values of less than 0.5 pixel without the application of specialised processing approaches. In this research, the bundle block adjustment had produced geometrically consistent data that were suitable for fusion.

Maintenance of pixel brightness values during orthorectification and resampling is vital where spectral analysis is to be undertaken following geometric correction of the data. Nearest neighbour and bilinear interpolation were considered, but cubic convolution resampling was utilized to achieve the highest pixel quality.

Multi-sensor fusion was facilitated through resampling all data to a common pixel dimension. To minimise the loss of radiometric information and to facilitate multi-sensor fusion, a 0.5m pixel grid was selected so that the spatial integrity of the observed data from all the sensors could be maintained. This sampling interval is

smaller than the multispectral image and LiDAR data but larger than the colour image.

It has already been mentioned that 18 GCPs for the forest study area and 32 GCPs for the semi-urban study area were collected with the aid of differential GPS. Each set of images for each area was treated independently due to the wide range of spatial resolutions and the orthorectification results. LiDAR-derived DEM was introduced to remove terrain distortions and kept accuracy levels within the desirable limit. The cubic convolution resampling technique was used for final orthorectification with a spatial resolution of 0.5 m.

The processing of the LiDAR data was different from the processing of the aerial imagery. LiDAR data did not require orthorectification, however a normalised DSM needed to be generated from the LiDAR source for use in the fusion process. Normalised DSMs represent the mean top surface of objects whose height can be computed as the difference between the first and last LiDAR returns. A height threshold was used in order to eliminate the effect of low-lying objects such as shrubs and fences. Regression analysis of object mean height measured in the field suggested that the mean heights generated by the normalised DSM were as nearly accurate as real mean heights.

The following chapters present the implementation procedures and the results of aerial imagery and LiDAR data fusion for the forest and semi-urban study areas.

CHAPTER 5

IMPLEMENTATION OF DATA-DRIVEN FUSION MODELS FOR TREE ATTRIBUTE DELINEATION

This chapter focuses on implementation of the data-driven pixel- and feature-level fusion models for the delineation of individual tree attributes from the forest study area. Detailed descriptions of the forest study area, and the characteristics of the tree structures, were presented in Chapter 3. The geometric-correction of the datasets was presented in Chapter 4. Data-driven fusion models are implemented at pixel- and feature-level using geometrically corrected aerial imagery and LiDAR data. In pixel-level fusion, an unsupervised classification scheme is implemented for the fusion of multispectral imagery with LiDAR data. In feature-level fusion, a watershed algorithm is used for delineating individual tree crowns, a masking technique is used for collecting tree feature attributes, and finally an unsupervised classification scheme is incorporated for fusing tree features for tree species discrimination.

5.1 Introduction

A forest ecosystem not only functions as a biological production and environmental services but also considers as a research object to understand inner structure of the forest landscape (Hsiaofei *et al.*, 2006). Individual tree components (both in the

*The contents of this chapter have been published in the following peer reviewed publications:

- **Ali, S.,** Dare, P. and S. Jones (2009). *A new object-based fusion model using LIDAR and multispectral imagery for forest structure assessment at the tree level.* In: Ostendorf, B., Baldock, P., Bruce, D., Burdett, M. and P. Corcoran (eds.), Proceedings of the Surveying & Spatial Sciences Institute Biennial International Conference, Adelaide 2009, Surveying & Spatial Sciences Institute, pp. 943-952. ISBN: 978-0-9581366-8-6.
- **Ali, S.,** Dare, P. and Jones, S. (2008) *Fusion of remotely sensed multispectral imagery and LiDAR data for forest structure assessment at the tree level,* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVII, part B7, Beijing 2008, pp.1089-1094.

horizontal and vertical plane) are part of the inner forest landscape and need close attention to understand the forest ecosystem as a whole. LiDAR data provide accurate measurements of forest structure in the vertical plane; however, current LiDAR sensors have limited coverage in the horizontal plane. Conversely, high spatial resolution multispectral imagery provides extensive coverage of forest structure in the horizontal plane, but is relatively insensitive to variation in the vertical plane. Therefore, it is desirable to synergistically use both sensors for mapping forest parameters at the tree level. Delineating individual trees and extracting relevant tree structure information from fused remotely sensed data has significant implication in a variety of applications such as reducing fieldwork required for forest inventory (Gong *et al.*, 1999), assessing forest damage (Kelly *et al.*, 2004) and monitoring forest regeneration (Clark *et al.*, 2004).

In this research, data-driven pixel- and feature-fusions are implemented for delineating individual tree attributes from the forest study area. In these processes, users play no part until the computational aspects are completed. An unsupervised classification scheme is implemented for pixel-level fusion of multispectral imagery with LiDAR data. The feature-level fusion incorporates a data-driven watershed algorithm for delineating individual tree features and uses an unsupervised classification for final fusion.

5.2 Tree Property Analysis

The primary objective of this research is to investigate the data-driven fusion models for delineating tree species from the forest study area. Tree species have to be identified both in spatial and spectral domains. The former may be satisfied by a well-structured generic classification scheme, such as was introduced in Section 2.2.2. The latter is satisfied by maximising the inter-class variance and minimising the intra-class variance. In this research, inter- and intra-class variances are considered part of attributes derived from multispectral imagery and LiDAR data. Fusion of multispectral imagery with LiDAR data uses the spatial and spectral properties to separate tree species. The location of the forest study area was given in Section 3.5 together with the description of the dominant tree species.

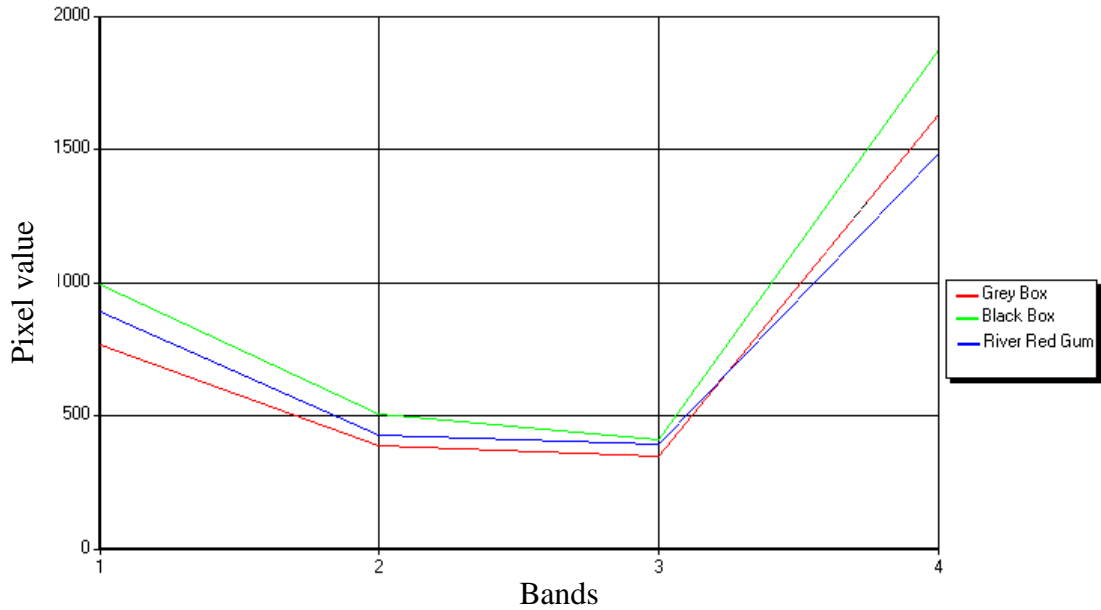


Figure 5.1 Tree species spectral profile of the forest site

Common tree species in the forest study area are the river red gum (*Eucalyptus camaldulensis*), black box (*Eucalyptus largiflorens*) and grey box (*Eucalyptus microcarpa*). Different *Eucalyptus* species can be differentiated by visual inspection of their vertical profiles, such as their bark, leaf shape, and branch arrangement. However, these characteristics cannot be used in the interpretation of aerial imagery, only spectral differences in different trees can be picked by multispectral imagery. Spectral profile of the tree species in Figure 5.1 supports that argument. The different tree species exhibit some separation in the blue and green bands; however, they are inseparable in red and near infrared bands. These are the most critical bands to distinguish tree species according to their spectral response. The inclusion of LiDAR-derived tree height and texture data along with multispectral imagery is critical for tree species identification process.

5.3 Methodology for tree delineation

In a data-driven scenario, the fusion of multispectral imagery with LiDAR data is illustrated through Figure 5.2.

The methodology started with the data processing stage, where geometric corrected multispectral imagery and LiDAR data were prepared as the input of the fusion

processes. A detailed description of the geometric corrections of the multispectral imagery and the normalised DSM generation from LiDAR data was presented in Chapter 4. The pixel- and feature-level fusions are independently implemented using these processed data. Finally, comparison of the pixel- and feature-level fusions is made for delineating individual tree species and evaluating their accuracies. Detailed descriptions of the steps are given in the following sections.

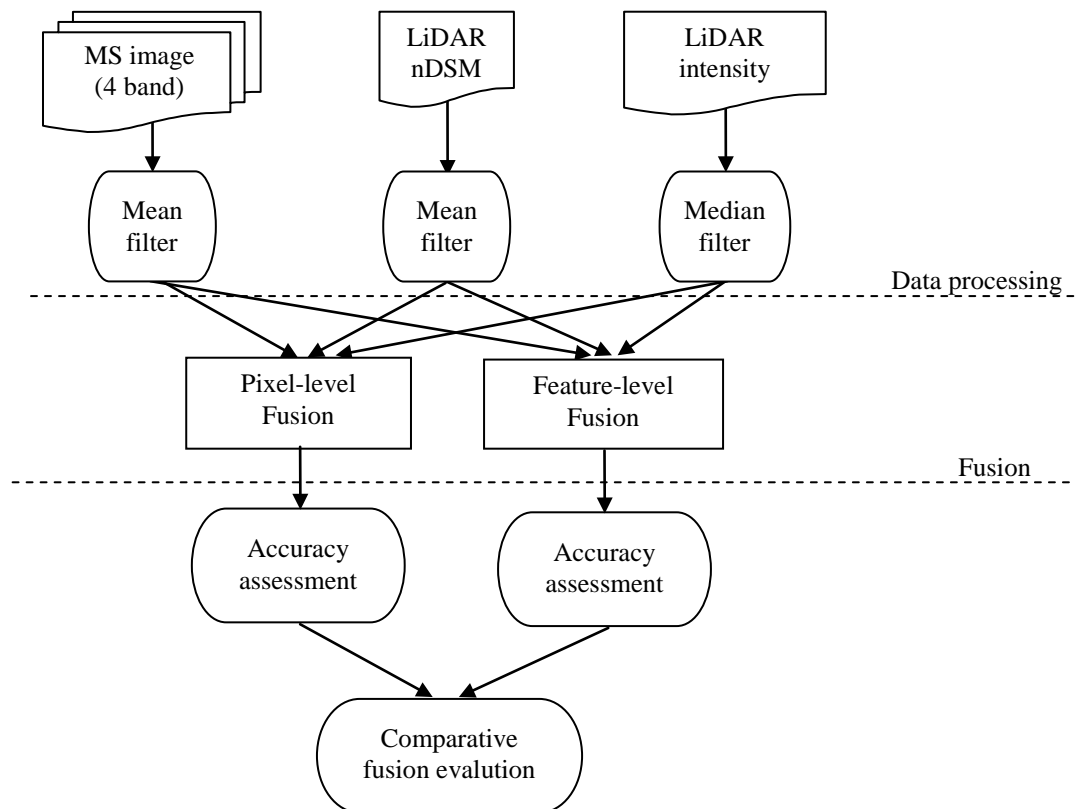


Figure 5.2 A conceptual model for comparing different data-driven fusion levels for tree species delineation

5.4 Data processing

The geometric corrected 4-band multispectral imagery and LiDAR-derived height and texture data are used in this study as the input layers. A description of these layers is given in Table 5.1. Different filtering techniques are applied to these datasets to enhance the spectral and spatial properties of the tree objects. A simple 3x3 mean filter is used on the multispectral imagery to impose a degree of homogeneity among the brightness values of adjacent pixels thereby increasing the

chance that neighbouring pixels may be given the same label. This filtering procedure also led to the suppression of shadow effects within the sunlit area of the tree crown. The coefficient and the dimensionality of the filter are primarily dependent on the solar direction at the time of over flight, the tree size, and the illumination conditions within the tree crown.

Table 5.1 Original and derived layers used in the proposed fusion procedures

Layer/Band	Description
1	Blue layer of original image
2	Green layer of original image
3	Red layer of original image
4	Infrared layer of original image
5	Height layer: LiDAR-derived nDSM
6	Texture layer: LiDAR first return intensity

A 3x3 mean filter was applied to the LiDAR-derived nDSM layer to reduce the dynamic range of tree heights. A 3x3 median filter was applied to the texture layer that is derived from LiDAR first return intensity. A median filter helps in reducing salt and pepper noise that would lead to inconsistent class labelling.

5.5 Pixel-level Fusion Model for Tree Species Delineation

An ISODATA unsupervised classification procedure is implemented for fusing multispectral imagery with LiDAR-derived height and texture data for the forest study area. A detailed review of the ISODATA fusion technique was presented in Section 2.5.5.1.

Figure 5.3(a) shows the study region where trees are highlighted in green and background soils are shown with blue to purple colours. Figure 5.3(b) shows a scatter diagram of the image in feature space. The red band versus infrared band brightness has been plotted. This is a subspace of the full six dimensional feature spaces of the data sources and illustrates how the tree species are distributed over this particular spectral space. The close mean distribution of the river red gum (*Eucalyptus camaldulensis*), black box (*Eucalyptus largiflorens*) and grey box (*Eucalyptus microcarpa*) on the scatter plot confirm that using only red and infrared bands

Eucalyptus species cannot be separated. The mean points are somewhat bit spread over the infrared band but they are very close together in red band. This interoperation very much supports the analysis in Section 5.2.

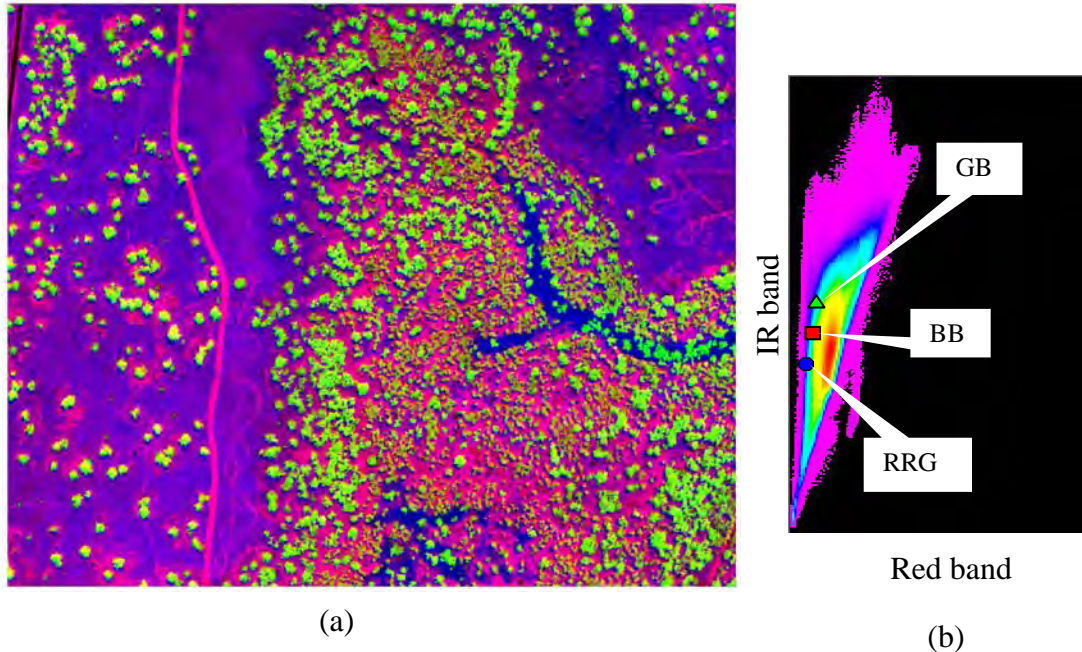


Figure 5.3 (a) A RGB composite image of the forest study area (R: IR, G: nDSM and B: texture); (b) Red band versus IR band scatter diagram with the tree species cluster centre

The ISODATA algorithm is implemented with 24 iterations and is required to determine 10 classes. The convergence threshold is set at 0.95 to prevent the ISODATA utility from running indefinitely. Appendix 3 shows the summary statistics of the initial 10 clusters. Merging and splitting options are employed at the end of the iterations, leading ultimately to the backgrounds and tree species clusters. In this process, the summary statistics and the covariance matrix of the clusters play a vital role. Generally, the trees and backgrounds are discriminated because the LiDAR-derived nDSM layer has a lower height for background clusters than for individual tree clusters. The availability of a cluster map allows a tree classification to be made through overlaying aerial photographs and field survey data. If some pixels with a given label are identified with particular tree species then all pixels with the same label are associated with that class. This method of fusion depends on *a posteriori* recognition of tree species since the user plays no part until the computational aspects are completed.

Visual interpretation of the clusters with the help of aerial photographs is used for final tree species separation. The statistical details for the tree species are presented in Table 5.2. These statistics represent the final properties for each tree species. The mean and standard deviation for each shows the tree clusters are separable using both spectral and spatial properties derived from combined multispectral imagery and LiDAR data.

Table 5.2 Mean values and standard deviation for each of the tree species generated from pixel-level fusion

Tree type	Layer	Mean	St. Dev.
Black box (BB)	Blue layer of original image	854.39	65.27
	Green layer of original image	431.19	26.54
	Red layer of original image	388.44	20.70
	Infrared layer of original image	1501.17	64.19
	Height layer: LiDAR-derived nDSM	8.18	3.80
	Texture layer: LiDAR 1 st return intensity	79.02	39.02
Grey box (GB)	Blue layer of original image	908.05	81.03
	Green layer of original image	453.59	32.47
	Red layer of original image	391.15	27.50
	Infrared layer of original image	1750.81	94.26
	Height layer: LiDAR-derived nDSM	8.24	4.12
	Texture layer: LiDAR 1 st return intensity	88.20	39.93
River Red Gum (RRG)	Blue layer of original image	734.28	56.69
	Green layer of original image	383.10	23.97
	Red layer of original image	363.39	17.70
	Infrared layer of original image	1297.91	74.43
	Height layer: LiDAR-derived nDSM	8.42	3.90
	Texture layer: LiDAR 1 st return intensity	78.05	40.10

An important aspect of this approach is that the whole process of fusion is data-driven and the user has very little influence. The visual interpretation largely comes after the processing, and is used for quality control purposes. Thus an unsupervised classification employing a pixel-level fusion of multispectral imagery with LiDAR data can successfully map tree species.

5.6 Feature-level Fusion Model for Tree Species Delineation

In the previous section, the data-driven pixel-level fusion model was implemented for tree species delineation. Now, the implementation of the data-driven feature-level fusion is presented.

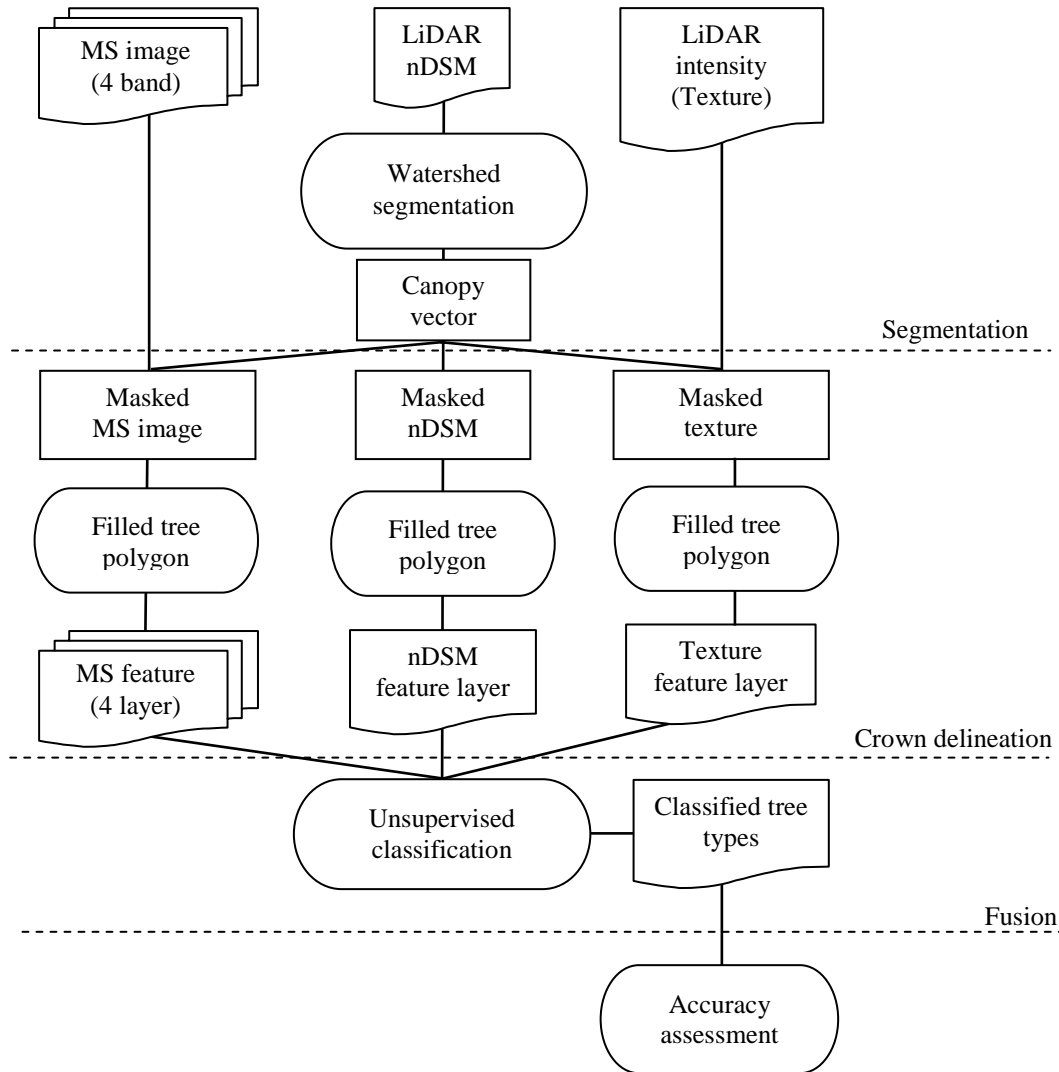


Figure 5.4 Feature-level fusion of multispectral imagery with LiDAR data for tree species delineation

The proposed feature-level fusion model consists of four parts: (1) watershed segmentation, (2) tree feature delineation, (3) tree feature fusion, and (4) accuracy assessment. The flowchart in Figure 5.4 illustrates the major steps performed within this feature-level fusion model.

5.6.1 Watershed segmentation for tree feature delineation

The LiDAR-derived nDSM represents the tree canopies of the forest study area in vertical as well as horizontal extents. Single and disjoint tree canopies can easily be delineated in this dataset. However, a segmentation procedure is needed to isolate individual trees in a group. This feature-level fusion model aims to use the watershed segmentation in tree feature isolation. A theoretical background of the watershed segmentation was given in the review of data-driven segmentation section (Section 2.6.1.1.5). This section describes implementation of watershed segmentation for tree feature delineation.

The review of the watershed segmentation algorithms concluded that the direct application of watershed transformation on the raw image generally leads to *over-segmentation* due to noise and other local irregularities. The literature suggests that over-segmentation can be serious enough to render the result of the algorithm virtually useless. To avoid this problem, Meyer and Beucher (1990) introduced marker-controlled watershed segmentation. The idea is to perform watershed segmentation around a specified marker. A *marker* is a connected component belonging to an image. The *internal* markers are associated with the object of interest, and *external* markers are associated with the background.

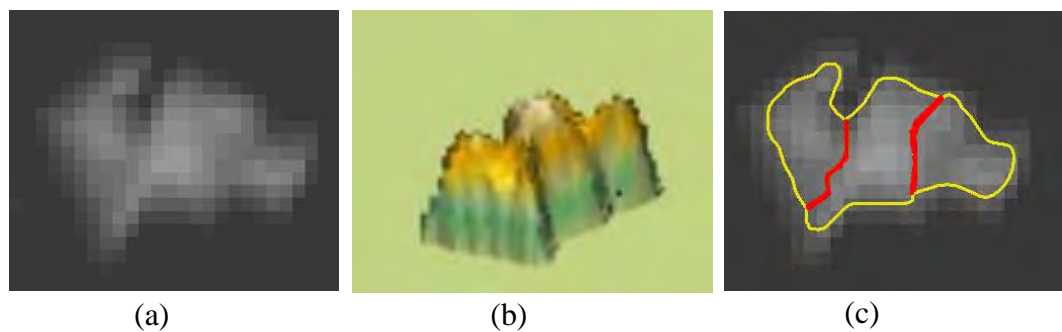


Figure 5.5 An illustration of watershed segmentation. (a) A canopy model derived from nDSM, (b) 3D view of the canopies, and (c) Segmentation results with dams built at the divide line

In the marker-controlled watershed segmentation process, the image is treated as a three-dimensional surface, with lateral dimensions representing the image plane, and the vertical dimension representing the grey values (Figure 5.5a). The watershed transform finds *catchment basins* and *water ridgelines* in the image by treating it as a

surface where high pixels are light grey values and low pixels are dark grey values. Internal markers are used to locate the local minima, which are associated with high grey values (such as selected tree crowns), and external markers are pointed to the local maxima, which are associated with the background. The watershed segmentation is performed through flooding from the local minima. Neighbouring watersheds are merged unless boundaries are built to isolate individual features (Figure 5.5c). The process of merging regions and building boundaries continues until no more region growing can take place.

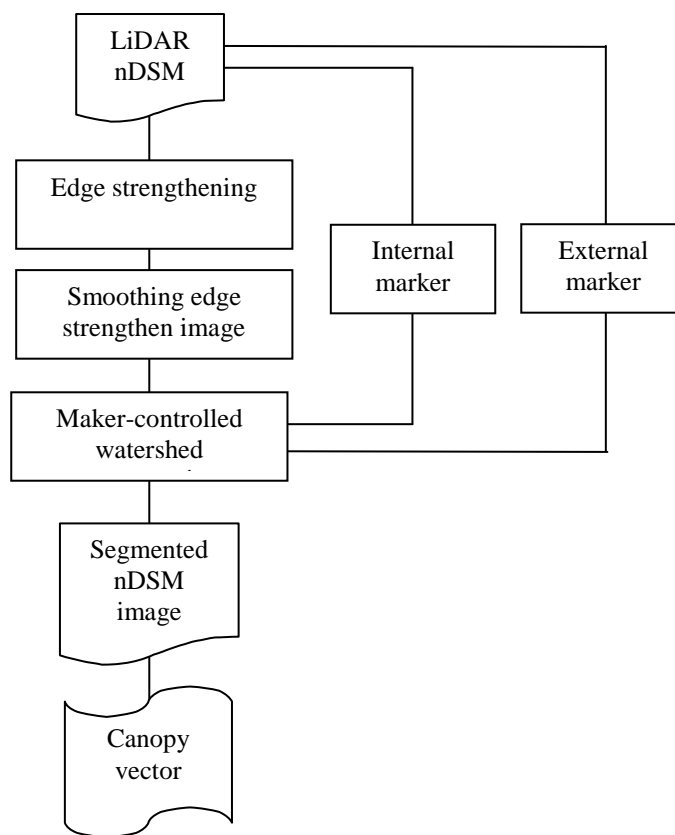


Figure 5.6 Flow chart of marker-controlled watershed transformation for tree feature delineation

Figure 5.6 illustrates the implementation procedure of the marker-controlled watershed algorithm using a LiDAR-derived nDSM for tree crown delineation. Marker-controlled watershed segmentation produces the best results for images, which have regions that are homogeneous and low intensity, separated by narrow boundaries of high intensity. The LiDAR-derived nDSM image does not have the low intensity with high frequency grey levels, but application of a first order edge

detection algorithm can yield this result. So, this is the first step in the watershed algorithm to use the 3x3 Sobel edge detector operation to compute the edge strength image. The gradient is high at the borders of the trees and low inside the tree canopy.

Definition of internal markers

Once the edge strength image has been created, the watershed procedure is begun with a variety of procedures to find the internal markers, which connect cluster of pixels inside each of the tree foregrounds. In this process, morphological techniques called *opening-by-reconstruction* and *closing-by-reconstruction* are used to pre-process the image (Gonzalez *et al.*, 2004; Soille, 2003). A regular morphological opening is an erosion followed by a dilation, while opening-by-reconstruction is an erosion followed by a morphological reconstruction. First, the opening-by-reconstruction is computed for a morphological reconstruction. Then closing-by-reconstruction is computed for removing dark spots. The reconstruction-based opening and closing are more effective than the standard opening and closing for removing small blemishes without affecting the overall shapes of the objects (Soille, 2003). The region maxima are calculated to obtain good internal markers. This procedure tends to leave some isolated pixels that must be removed. This is done by removing all clusters that have fewer than 10 pixels.

Definition of external markers

After defining the internal markers, it is necessary to define the external markers, which represent the background. In the nDSM image, the dark pixels or low height objects belong to the background, so initially a thresholding operation is used to separate them. The background pixels are close to the ground, however, preventing the external marker having too close to the edges of the trees. The background is thinned by computing the *skeleton by influence zone*. This operation is done by computing the watershed transform of the distance transform and then looking for the watershed ridgelines (Soille, 2003). Subsequently the nDSM image needs modification so that it has regional minima only in certain desired locations. The gradient magnitude image of nDSM is modified so that it only has regional minima occurring at internal and external marker pixels.

Application of watershed transformation

The external markers effectively partition the image into regions, with each region containing a single internal marker and part of the background. After that the watershed segmentation algorithm is applied each of the individual regions. In other words, the algorithm simply takes the gradient of the smoothed image and then restricts the operation to a single watershed that contains the marker in that particular region.

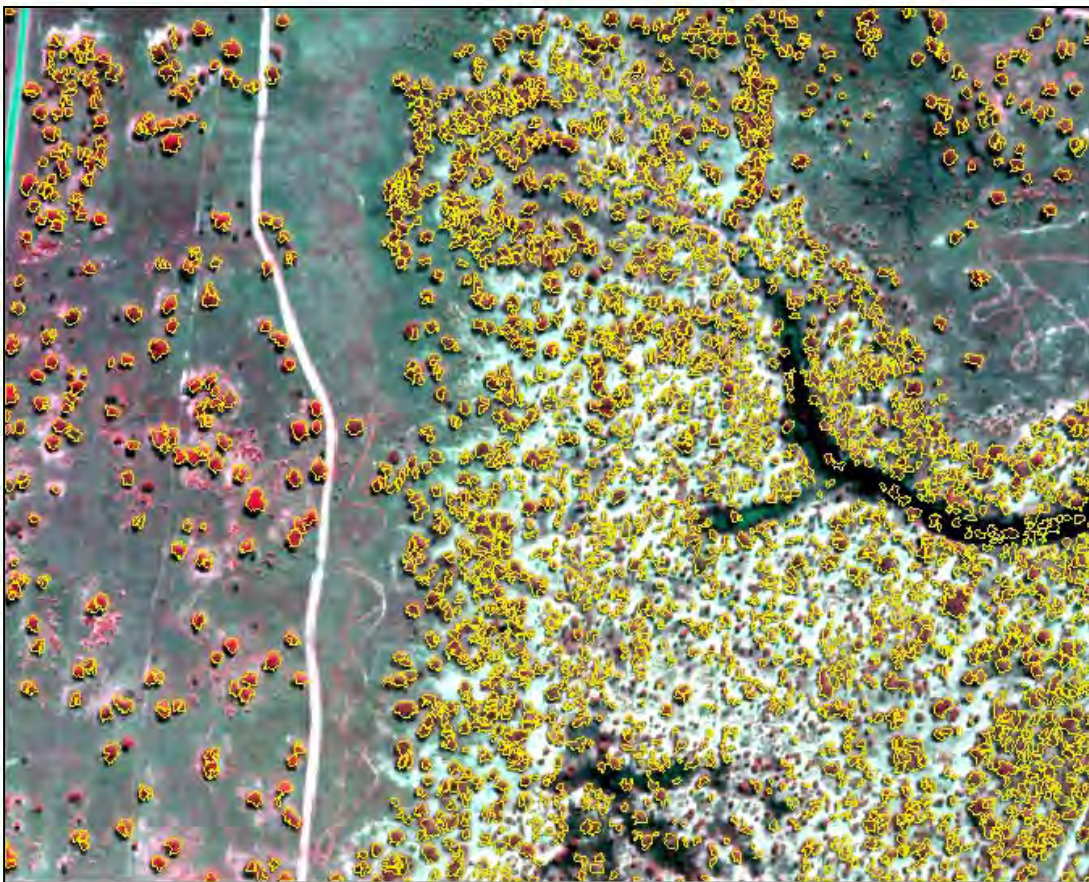


Figure 5.7 Tree crown map in the forest study area

After segmentation, the raster segments are converted into vector polygons for proper delineation of the tree crown features. The vector segments are post-processed to eliminate dead trunks and electric pylons present in the study site. All segments with a diameter less than 2m are removed so juvenile trees are ignored. Figure 5.7 shows the initial tree crown map derived from watershed segmentation using LiDAR nDSM data.

5.6.2 Attributes delineation for tree features

After segmentation, the resulting tree feature polygons are overlaid on the multispectral imagery and LiDAR data to mask out the spectral signatures and texture information of the tree crown features. Masked 4-band multispectral imagery provides the spectral contents of the tree features. Masked LiDAR intensity provides the texture information of the tree features and masked LiDAR nDSM provides height information of the tree features. These six masked layers are the basis for the feature-level fusion.

5.6.3 Filling tree features with attributes

In high spatial resolution data fusion, the class variability within the tree crown is caused mainly by the variability in crown structure (shadow effects), crown density (background material) and different tree components (bark and leaves) (Meyer *et al.*, 1996). In order to increase the significance of the fusion results, the entire tree polygon feature is filled with average digital number (DN) of each of the masked layers. This is achieved by extracting zonal statistics from six masked layers and saving them as the polygon attributes for each tree crown. Then, tree polygons are converted into raster layers and the mean value is used as the DN for each tree feature. In this way, only one DN of each masked layer occupied the entire polygon feature despite the initial tree feature having multiple DN. The six layers of tree features are created and they are ready for the input in the fusion process.

5.6.4 Unsupervised classification of tree features

The tree crown features derived from six masked layers are fused using the ISODATA unsupervised classifier. In this process, only three groups of tree species are determined using six iterations with convergence threshold of 0.95, since the main task is to delineate the three most prominent tree species. The same pixel-level ISODATA classifier is implemented for feature-level fusion. However the main difference is that it is used for delineating tree features rather than the single pixels. A detailed review of this algorithm was presented in Section 2.5.5.1 and pixel-level implementation was presented in Section 5.5.

Table 5.3 presents the summary statistics of the tree species derived from feature-level fusion using multispectral imagery with LiDAR data. Each tree species has its own set of mean values with their standard deviations, which helps to delineate the tree features of the different species.

Table 5.3 Mean value and standard deviation for each of the tree species generated from feature-level fusion

Tree types	Layer	Mean	St. Dev.
Black box (BB)	Blue layer of original image	922.79	66.86
	Green layer of original image	455.11	26.13
	Red layer of original image	401.85	21.35
	Infrared layer of original image	1505.95	87.25
	Height layer: LiDAR-derived nDSM	8.48	1.97
	Texture layer: LiDAR 1 st return intensity	81.29	17.69
Grey box (GB)	Blue layer of original image	1126.72	124.59
	Green layer of original image	532.49	46.24
	Red layer of original image	450.16	34.96
	Infrared layer of original image	1663.60	101.70
	Height layer: LiDAR-derived nDSM	7.26	1.85
	Texture layer: LiDAR 1 st return intensity	84.47	20.76
River Red Gum (RRG)	Blue layer of original image	798.79	76.185
	Green layer of original image	406.18	31.22
	Red layer of original image	376.05	20.00
	Infrared layer of original image	1333.24	100.09
	Height layer: LiDAR-derived nDSM	8.88	1.89
	Texture layer: LiDAR 1 st return intensity	74.86	16.66

5.7 Summary

This chapter presents the implementation of the pixel- and feature-level fusion models for forest species identification at an individual tree level using high spatial resolution multispectral imagery and LiDAR data. In data-driven pixel-level fusion the unsupervised classification is employed for the fusion of multispectral imagery with LiDAR-derived nDSM data. A total 10 classes were determined after 24

iterations in ISODATA algorithm. Then, merging and splitting options were employed to separate the backgrounds and tree species clusters. This method of pixel-level fusion depends on *a posterior* recognition of tree species, since the user plays no part until the computational aspects are completed.

The data-driven feature-level fusion consists of four steps: watershed segmentation, tree feature delineation, tree feature fusion, and accuracy assessment. Individual tree features were isolated using marker-control watershed segmentation on LiDAR-derived nDSM data. Spectral signatures and texture information of the tree crown features were masked out from multispectral imagery and LiDAR intensity layer. Then, the individual tree polygon feature was filled with average digital number of each masked layer. Finally, fusion of tree features was carried out using same pixel-level ISODATA unsupervised classifier. The main difference was that it was used to delineate tree features rather than the single pixels.

The following chapter is going to present the implementation of user-driven fusion models for identifying landscape objects from semi-urban study area.

CHAPTER 6

IMPLEMENTATION OF USER-DRIVEN FUSION MODELS FOR SEMI-URBAN LANDSCAPE MAPPING

This chapter focuses on the implementation of user-driven fusion models for mapping a range of landscape features within a semi-urban study area. A detailed description of the semi-urban study area was presented in Chapter 3 and the geometric correction of the datasets was described in Chapter 4. Successful fusion of aerial imagery with LiDAR data is dependent on the clear recognition of landscape objects from analysis of spectral and spatial characteristics. A review of the urban landscape classification scheme was presented in Section 2.2.1. In pixel-level fusion, supervised classification algorithms are implemented for the fusion of the aerial imagery with LiDAR-derived nDSM data. In feature-level fusion, multi-resolution segmentation is used to define landscape features and subsequently utilise rule-based fuzzy classification for features fusion.

6.1 Introduction

Successful fusion and subsequent identification of landscape features is dependent on the image enhancement and information extraction procedures applied to the

*The research presented in this chapter has been published in the following peer reviewed publications:

- **Ali, S. S.**, Dare, P. and Jones, S. (2009) *A comparison of pixel- and object-level data fusion using LiDAR and high-resolution imagery for improved classification*, In: Jones, Simon, Reinke, Karin (Eds.), *Lecture notes in Geoinformation and Cartography*, Springer, pp. 3-18.
- **Ali, S. S.**, Dare, P. and Jones, S. (2006) *A comparison of pixel- and object-level data fusion using LiDAR and high-resolution imagery for improved classification*, *Proceedings of the 13th Australasian Remote Sensing and Photogrammetry Conference*, November 2006, Canberra.
- **Ali, S. S.**, Dare, P. and Jones, S. (2005) *Automatic classification of land cover features with high resolution imagery and LiDAR data: an object-oriented approach*, *Proceedings of SSC2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute*, September 2005, Melbourne, pp. 512-522.

remotely sensed data. In this research, image enhancement techniques need only be investigated to the extent necessary to provide data of consistent quality suitable for fusion. Data that occupies relatively small sites with minimal topographic variation are assumed to exhibit consistent atmospheric and topographic effects (Gong and Howarth, 1990). The fusion algorithms utilised in this research are applied to independent multi-source datasets and are statistically invariant to linear transformations, consequently no radiometric corrections are made to the data. Contrast stretching and formation of colour composite images are performed as an aid in the delineation of training samples for pixel-level fusion.

The overall objective of this research is to evaluate whether the fusion of high spatial resolution aerial imagery with LiDAR data can be used as a tool for semi-urban landscape mapping. The research aims specifically to evaluate the results of the pixel- and object-level fusions used to integrate the information content of multispectral imagery with LiDAR-derived DSM data.

6.2 Properties of the Semi-urban Landscape Objects

The design of an effective landscape classification system relies on the recognition of the target and sensor characteristics in conjunction with the interpretation approach to be applied. A review of the standardised landscape classification system was presented in the Section 2.2. For this research, the landscape classification system is developed as part of the semi-urban area mapping using a range of low altitude aerial imagery. Consequently, semi-urban landscape classes are defined in terms of spectral separability, minimal interclass variability and contextual information. The characteristics of the objects in the semi-urban study area were presented in Section 3.6.

Table 6.1 Landscape classification system developed for the semi-urban study area

Level I	Level II	Level III
1 Natural objects	11 Vegetation	111 Tree 112 Grass
2 Manmade objects	21 House 22 Infrastructure	211 Roof 212 Road/open area
3 Obscured objects	31 Shadow	311 Roof shadow 312 Tree shadow

Table 6.1 illustrates the landscape classification system develops for the semi-urban study area. This is a hierarchical system and the class level expands from Level I to III. The class level accommodates the detail of the objects identified from various airborne data sources. Level I and Level II consist of the most generic classes and can be identified from a single data source. However, Level III needs to utilise multi-source data fusion for delineating semi-urban landscape objects.

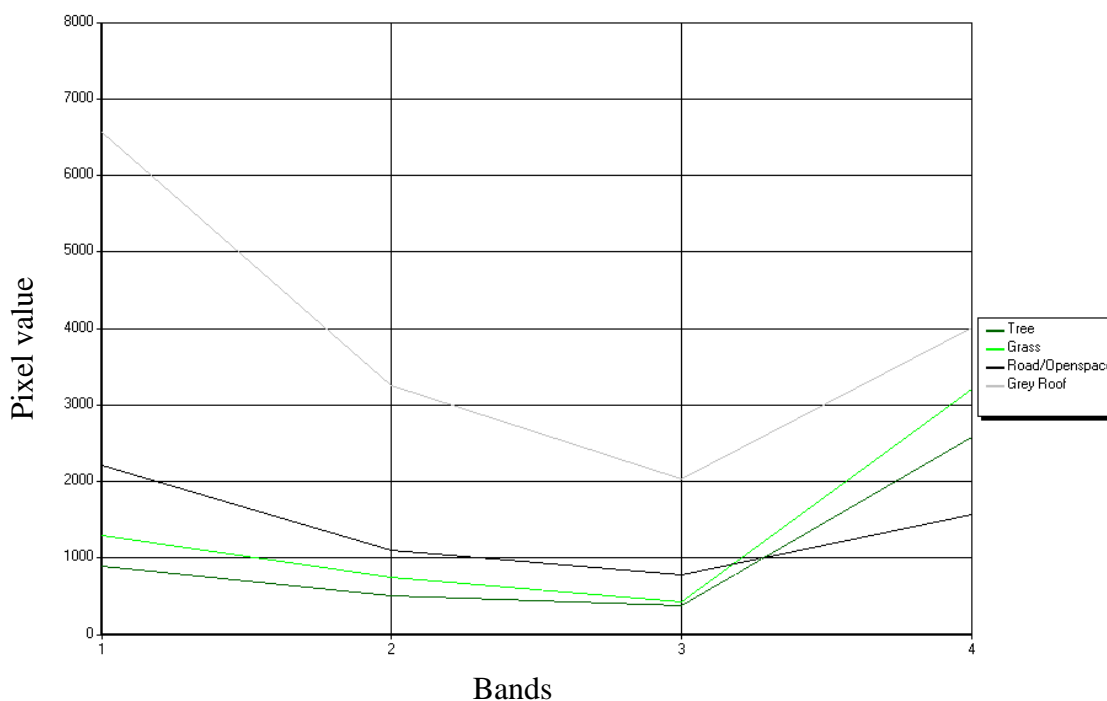


Figure 6.1 Spectral profiles of semi-urban landscape objects

In Section 2.3.1, it had been stated that shadow was a significant problem for classifying urban landscape from high spatial resolution imagery. Most of the cases, shadows are identified as an unclassified class therefore, a significant urban land cover information is lost (Zhou *et al.*, 2009). In this study, Obscured objects (Level I)

are identified as Shadows (Level II) then classified them into different land cover types (Level III).

The degree of interclass spectral variability for each of the semi-urban landscape objects has been assessed through analysing the spectral profiles (Figure 6.1). Scatter plots of the object signatures reveal that the spectral characteristics of the objects are complex and only a handful of objects are spectrally separable. In 4-band multispectral imagery, the mean values of the semi-urban objects indicates that some objects, such as different types of roofs, can be delineated, however other landscape objects cannot be recognised using only spectral signatures.

This research investigates the usefulness of fusion of multi-source data to overcome the difficulties described above. A hierarchical landscape classification system is developed to utilise an object's spectral and spatial properties, which can be recognised in multi-source data. For example, built-up areas may be isolated from vegetated areas in moderate spatial resolution optical imagery. However, without enough spectral and spatial information, an individual rooftop cannot be delineated. A high-resolution height data is needed to automatically separate concrete rooftops from parking lots. The landscape classification scheme must be compatible with the various levels of interpretation possible in the fusion model. A meaningful comparison of different fusion models is necessary to exploit the full potential of data collection and analysis techniques.

6.3 Methodology for Semi-urban Mapping

In a user-driven scenario, the fusions of aerial imagery with LiDAR data are illustrated in Figure 6.2.

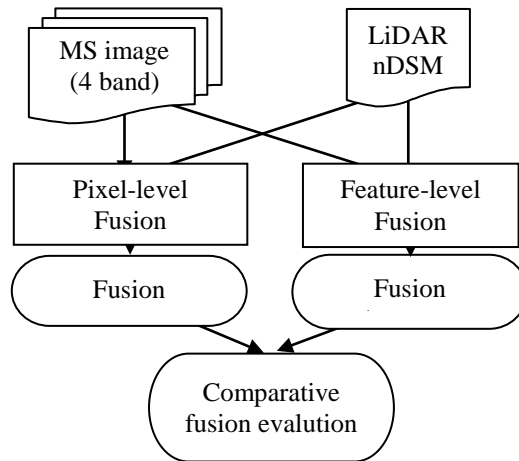


Figure 6.2 A conceptual model for comparing different user-driven fusion levels for semi-urban landscape mapping

The methodology uses 4-band multispectral imagery and LiDAR-derived nDSM data. A detailed description of the geometric corrections of these datasets was presented in Chapter 4. The pixel- and feature-level fusions are implemented and the accuracies are assessed. Detailed descriptions of these steps are given in the following sections.

6.4 Pixel-level Fusion for the Semi-urban Study Area

A wide range of pixel-level supervised classification algorithms was reviewed in Section 2.5.5.2 and they were applied in a variety of multi-source data fusion applications. In this research, a supervised classification algorithm is employed for the pixel-level fusion of aerial imagery with LiDAR data. In particular the parallelepiped classification algorithm is considered, since it depends only on the application of thresholding components of the datasets. Other algorithms such as maximum likelihood and minimum distance classification are commonly available; however the maximum likelihood classifier presents a number of difficulties in multi-source data fusion. These include incompatible statistics of disparate data types, with some data unable to be represented by normal class models (Richards and Jia, 2005).

The pixel-level supervised fusion starts with a stacked-vector dataset of 4-band multispectral imagery and LiDAR-derived nDSM data. A supervised parallelepiped classification procedure is then applied on it. The essential practical steps for the implementation of the supervised classification procedure were presented in Section 2.5.5.2. The flowchart in Figure 6.3 illustrates the implementation of the user-driven pixel-level fusion in this research.

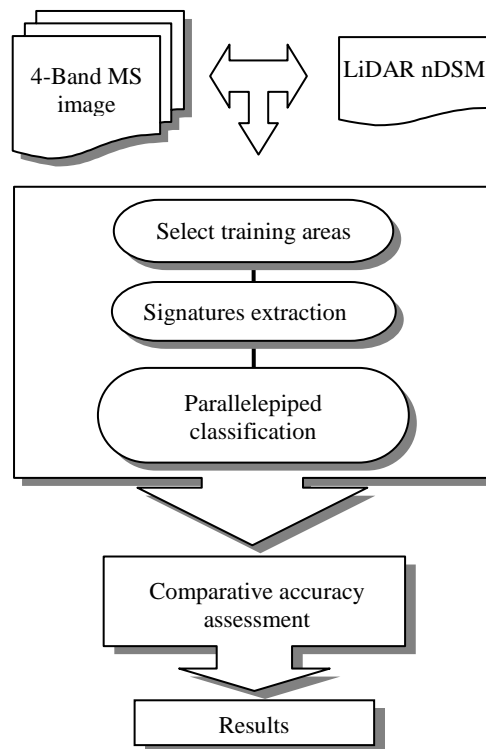


Figure 6.3 Flowchart for pixel-level fusion of aerial imagery with LiDAR-derived nDSM.

6.4.1 Selection of the training landscape objects

Much attention has been directed towards the collection of training samples from multi-source data for pixel-level supervised classification. The semi-urban landscape classification scheme is used to determine the landscape classes corresponding to training samples. Training samples are chosen by their reflectance characteristics and they are also used for spectral separability measure. Improved definition of training parameters for landscape classes leads to enhanced fusion performance. Critical parameters, including optimum training size, adequately measured class variances, and class separability, are selected to ensure that semi-urban landscape objects can be separated spectrally.

Training landscape objects are usually established according to a number of conflicting objectives. Firstly, they are selected to incorporate areas containing a uniform landscape of interest. This is rarely satisfied, as most landscapes comprise a mixture of components, therefore it is far more useful to select areas representative of the landscapes present. Secondly, spectral signatures should be unique, and to maximise separability between classes, class variance must be minimised which again suggests that training samples should be collected from areas of uniform landscape. However, if representative training data are to be obtained, sufficiently large areas must be sampled in order to assess the natural within-class variance. A measure of *sufficient size* of training samples is therefore required.



Figure 6.4 The training objects distribution on the subset of multispectral imagery of the semi-urban study area

The relevant semi-urban landscape objects have already been taxonomically identified using the semi-urban landscape classification system as present in Section 6.2. A description of individual semi-urban objects was presented in Section 3.6.

Figure 6.4 illustrates the spatial location of the training landscape objects used as the signatures of the semi-urban landscape objects. Total semi-urban study area covers 1.25km x 1km area or 2500 x 2000 pixels from their 51,302 pixels (or 10 percent) are selected as the training pixels. The breakdown of training pixels for each landscape classes is provided in Table 6.2. This number of the training pixels helps to accurately estimate the lower and upper threshold of the parallelepiped classifier and improve the overall accuracy of fusion results.

Spectral stratification of targets of this research has been applied to the derivation of training statistics for the *Vegetation* and *House* classes. In both cases, subclasses are identified. *Tree* and *Grass* are the subclass of *Vegetation*, and *Roof* is the subclass of *House*. Delineation of *Shadow* classes is the most challenging task in the pixel-level classification process. They cannot be separable in spectral domain as their formation is due to the suppression of obstacles. However, without any contextual information further classifications of *Shadow* are rarely achievable (Dare, 2005). Shadow training objects are carefully selected and later they are refined to produce results of an acceptable standard.

6.4.2 Signature extraction of semi-urban landscape objects

The spectral signatures of the semi-urban landscape objects are extracted from the training objects and listed in a library of signatures (Table 6.2). Later, in the fusion process, unknown pixels are compared with them and allocated to appropriate landscape object classes.

Table 6.2 Training statistics derived from 4-band multispectral image with nDSM data

Classes	Pixels	Band 1		Band 2		Band 3		Band 4		nDSM	
		Mean	Std. D.	Mean	Std. D.	Mean	Std. D.	Mean	Std. D.	Mean	Std. D.
Open-space	20342	2752.04	881.67	1246.64	348.01	814.21	168.07	2014.15	603.148	0.00	0.03
Roof	9556	4152.68	1480.46	2057.57	727.89	1335.24	457.12	2664.80	842.61	3.96	2.07
Roof shadow	2428	763.47	190.82	429.80	91.344	415.99	54.06	933.23	181.14	0.32	1.10
Grass	10456	1325.68	294.51	653.87	86.18	444.59	38.44	2335.29	276.35	0.07	0.78
Tree	3273	868.21	170.65	476.46	74.01	364.22	31.24	2314.90	322.05	5.36	3.27
Tree shadow	2559	478.23	140.97	286.70	60.97	312.06	32.87	806.84	191.04	0.15	0.86

Training signatures contain an adequate number of pixels and most of the major classes are separable in spectral and nDSM layers. This is important in a supervised classification to avoid misinterpretation of objects with similar spectral signatures. The most confusing landscape classes are the *shadows*. The mean of their spectral signatures for each band is very close (see Table 6.2). As a result, it is hard to distinguish them solely by their spectral signatures. *Shadow* signatures derived from the nDSM layer as not show any particular height attributes for each shadow class, making their classification in pixel-level fusion very difficult or impossible.

6.4.3 Fusion using parallelepiped classifier

Histograms of the individual spectral components of the available training data are inspected to train the parallelepiped classifier. The upper and lower significant boundaries on the histograms are identified and used to describe the range of the value for each layer of data for that landscape class. The range of the object's signature values in all layers described as a multi-dimensional box or parallelepiped. In the fusion process, pixels found to lie in such parallelepiped are labelled as belonging to that class. Spectral parallelepiped for each of the eight classes is derived from all the layers using the training statistics shown in Table 6.2

Fusion at Level II (Table 6.1) represented an aggregation of individual objects derived from Level III classes. Through post-fusion sorting the unique spectral characteristics of the objects at Level II are delineated. Aggregation of spectrally diverse signatures prior to fusion would have decreased the scope for discrimination between classes and reduced the accuracy of fusion results.

Small areas of training classes affect the ability to define reliable spectral signatures. Individual *shadow* classes occupied small areas so they were difficult in defining training sites. These classes particularly have mixed boundary pixels and have heterogeneous spectral pixels. Additionally, all *shadows* contain similar spectral signatures for each data layers therefore, pixel-based analysis gave unsatisfactory results (Dare, 2005).

6.5 Feature-level Fusion for the Semi-urban Study Area

Feature-level fusion starts with the identification of meaningful features over the area covered by multi-source data and then making a classification from these. The procedure thus consists of segmentation then classification. Feature-level fusion is implemented in this research using 4-band multispectral imagery with nDSM data. The flowchart in Figure 6.5 illustrates this implementation.

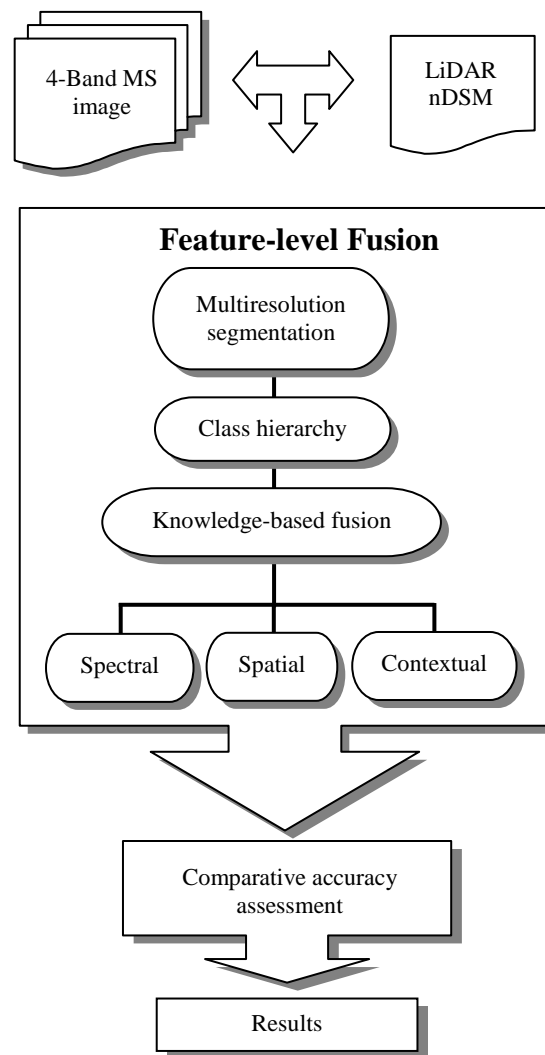


Figure 6.5 Flowchart for the user-driven feature-level fusion of multispectral imagery with LiDAR-derived nDSM for semi-urban landscape mapping

6.5.1 Multi-resolution segmentation

The basic processing units of feature-level fusion are segments or features, not single pixels. In the segmentation process, for each image feature a meaningful statistic is calculated in an uncorrelated feature space using shape, texture and topological features. This information improves the value of the final fusion and cannot be fulfilled by common, pixel-level approaches (Benz *et al.*, 2004). In this research, feature primitives are created through multi-resolution segmentation. These features are polygons of roughly equal size exhibiting internal homogeneity. The theoretical background of this segmentation process was presented in Section 2.6.1.2. In this segmentation process, the scale parameter determines the maximum allowed

heterogeneity for the resulting image objects. The size of the image objects can be varied by modifying the value of the scale parameter. For homogeneity, the relative weight applies to spectral versus shape criteria to reduce heterogeneity. Here shape, smoothness and compactness criterion are applied in a mixed form to define homogeneity for the image objects.

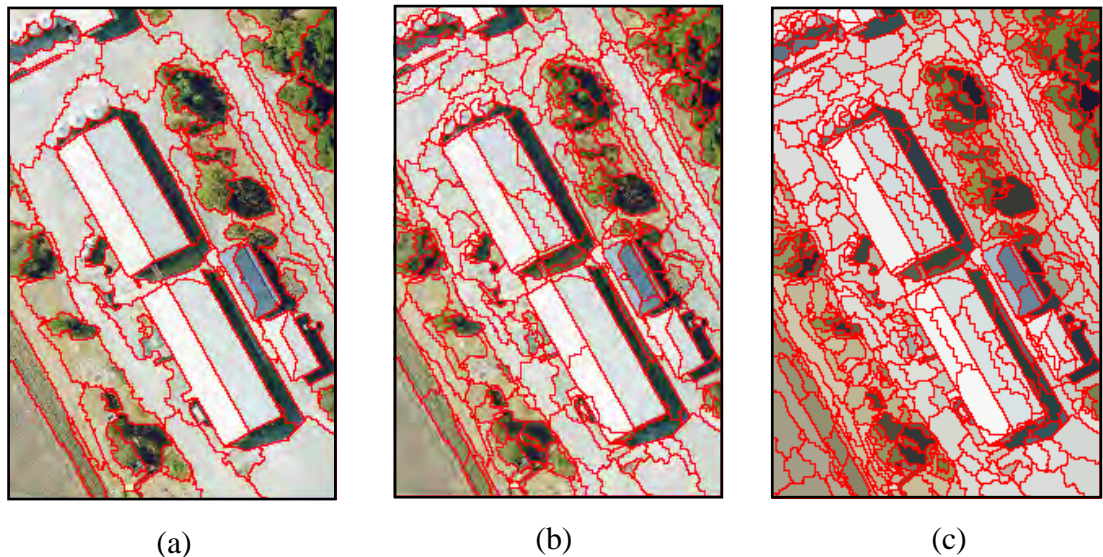


Figure 6.6 Scale-parameter for the multi-resolution segmentation: (a) coarse size features with large scale-parameter of 40, (b) medium size features with moderate scale-parameter of 25 and (c) small size features with small scale-parameter of 10

Figure 6.6 illustrates the effect of scale-parameters on the definition of feature size. The shape, smoothness, and compactness criteria are assigned the same values (0.275: 0.3: 0.7) for different scale parameters in order to compare effect of scale on feature definition. The large scale-parameter in Figure 6.7a has coarse segment size roughly equal to 40 pixels per feature. The large segment incorporates variability within a feature and dilutes separability among the features. Appropriate feature extraction is not achieved with this segmentation. The same problem arises for the small scale-parameters (segment size 10 pixels), as illustrated in Figure 6.7c. Small segments only include part of a feature and highlight noise within a feature. This segmentation behaves like a pixel-level approach and is not suitable for further fusion. However, the medium scale-parameter of 25 pixels segmented the features appropriately (see Figure 6.7b). By visually comparing different segmentation results, the scale parameter of 25 was chosen to create local homogeneity and to keep global heterogeneity.

For the fusion of multispectral imagery with nDSM data, equal weight is assigned to each of the multispectral bands. This emphasis is chosen because of the lack of colour homogeneity visually observed within the same features in the image. On the other hand, the LiDAR-derived nDSM has more homogeneity in grey levels; therefore more weight is given to this layer (see Table 6.3).

Table 6.3 Segmentation parameters for the aerial imagery with LiDAR-derived nDSM data fusion for level III landscape classification scheme

Parameters	MS & nDSM fusion
Weight	1 MS & 30 nDSM
Scale factor	25
Shape Factor	0.75
Compactness	0.3
Smoothness	0.7

A ratio of smoothness to compactness weight is specified as 3:7 for multispectral imagery with LiDAR-derived nDSM data (Table 6.3), emphasising the discrete, compact nature of *House Roofs*. A higher smoothness emphasis would be used to define objects observed to have greater variability between features (Baatz *et al.*, 2004). The compactness weight makes it possible to separate features that have quite different shapes but not necessarily a great deal of colour contrast, such as *House roofs* versus *Roads* within the semi-urban study area.

6.5.2 Class hierarchy

The class hierarchy is the framework of feature-level classification used to create the knowledge base for the fusion task. It contains all classes and is organised in a hierarchical structure (Baatz *et al.*, 2004). The class hierarchy passes down class descriptions from parent classes to their child classes. It reduces the redundancy and complexity in the class descriptions and creates a meaningful grouping of classes.

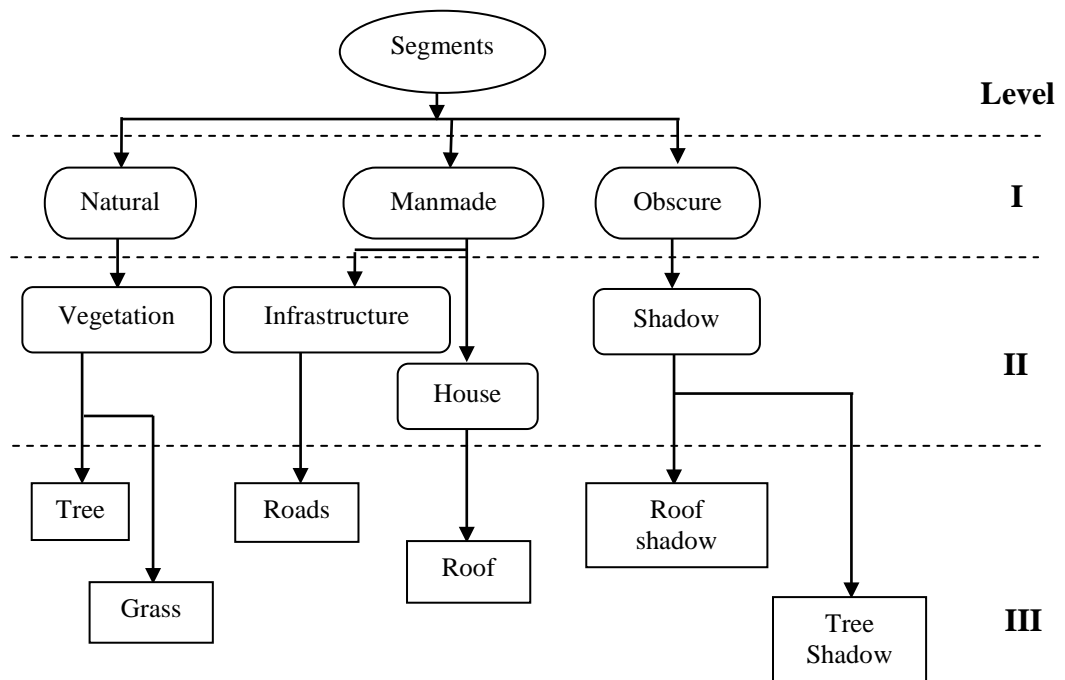


Figure 6.7 The class hierarchy for the feature-level fusion of semi-urban study area

In this research, the class hierarchy in feature-level fusion is developed through utilising a similar landscape classification scheme as was applied in pixel-level fusion (see Section 6.2). This class hierarchy is developed with the help of the urban classification scheme reviewed in Section 2.2.1. The landscape classification is defined as an inheritance hierarchy referring to the physical relations between the classes. In Level I, *Natural*, *Manmade* and *Obscure* features are the parent classes. In Level II, *Vegetation* is the child class of *Natural* feature, *Infrastructure* and *House* are the child classes of *Manmade* feature, and *Shadow* classes are the children of *Obscure* feature. In Level III, *Grass* and *Tree* classes became child of the *Vegetation* class (Figure 6.7).

Within this class hierarchy each class is described either by one or more fuzzy-membership functions, a nearest neighbour classifier, or by a combination of both. Fuzzy membership functions are determined by the semantic import (SI) model, which is based on expert knowledge of the features. A detailed review of the SI model and fuzzy membership functions was presented in Section 2.7.1. In a membership function the definition of a class requires more knowledge of the reality than in the nearest neighbour classifier. The membership process begins with a certain scale, then a search for the segments with similar statistical values. However, caution should be taken with the segments that represent extremes of a class

especially when high internal variance is indicated. A stepwise refinement of the class hierarchy is achieved using the inheritance mechanism. The membership functions utilise spectral, spatial, and contextual information to fuse segments for feature extraction. The following sections describe the details of the fusion procedures.

Fusion based on spectral properties

Since the generated segments hold more spectral information than the individual pixels, feature-level fusion offers a huge variety of derivative spectral features (Hofmann, 2001). Brightness and spectral ratios of the features are calculated using all image layers. Textural features are determined using standard deviations of layer values, spectral mean values, and average spectral differences. Contrast information is generated through spectral differences from neighbouring features. Context-related features are generated using mean spectral differences within a given class.

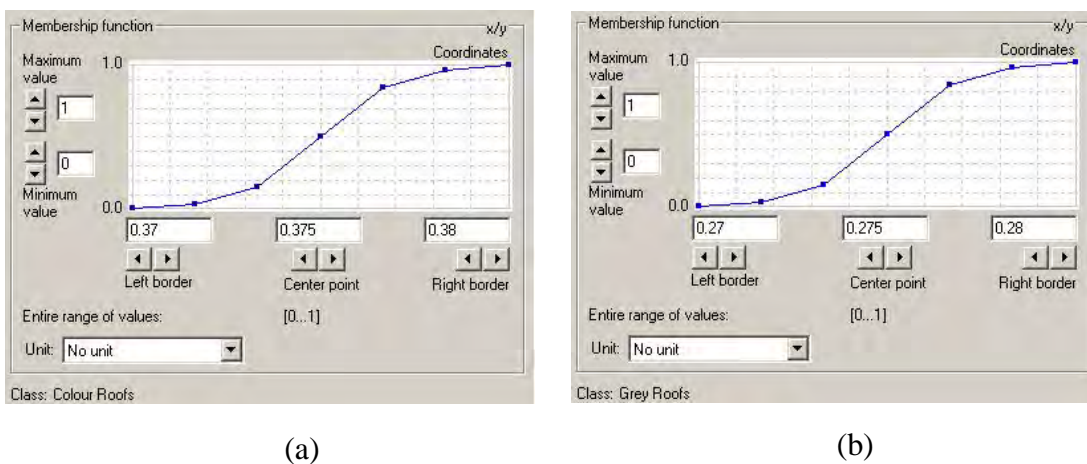
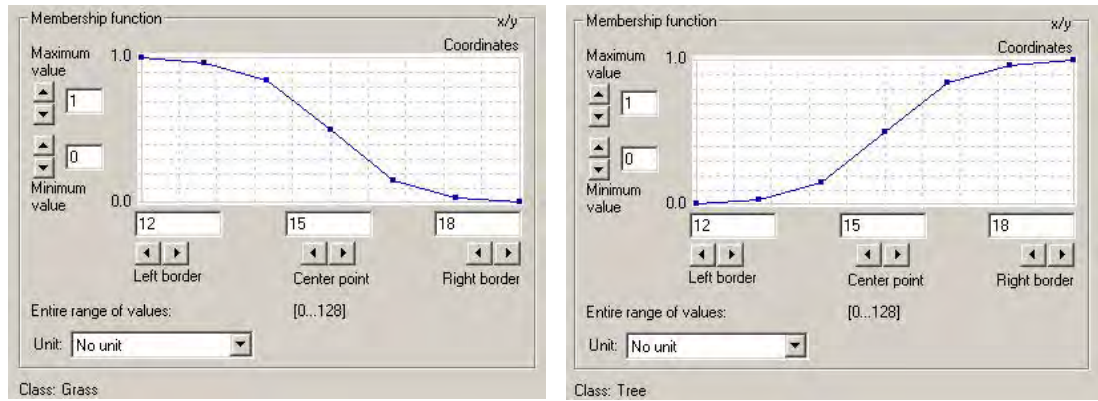


Figure 6.8 Fuzzy membership functions for the discrimination of *Manmade* and *Obscure/Shadow*

At Level I for fusion of 4-band multispectral imagery with nDSM data, the *Manmade* and *Obscure/Shadow* classes are discriminated using brightness values of the features. A feature is described as *Manmade* if its brightness value is more than or equal to 70, and a feature is *Obscure/Shadow* if it was less than 70.



(a)

(b)

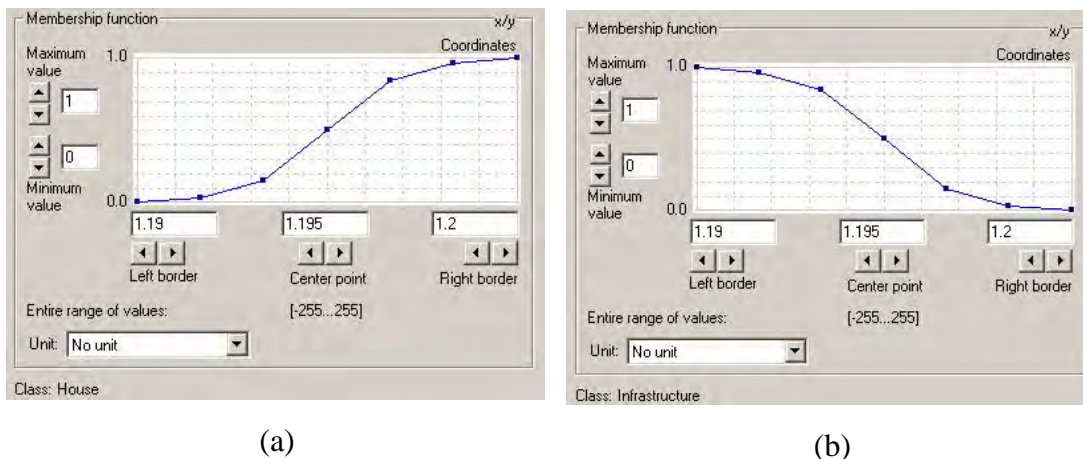
Figure 6.9 Fuzzy membership functions for the discriminating *Grass* and *Tree* classes in multispectral imagery

At Level II for fusion of multispectral imagery with nDSM data, the *Vegetation* class is separated from others by a fuzzy membership description of the mean and the ratio of the green spectral band. A segment is represented as a *Vegetation* feature if the mean of the green band is larger or equal to 30 and also the ratio is larger or equal to 0.37.

At Level III for fusion of multispectral imagery with nDSM data, *Vegetation* is further subdivided into *Tree* and *Grass* classes. Figure 6.10 illustrates the fuzzy membership functions for discriminating *Grass* and *Tree* classes. Using fuzzy membership function a *Vegetation* feature is represented as *Grass* if its standard deviation of the green band is smaller or equal to 18 (see Figure 6.9a). On the other hand, for a *Tree* the standard deviation is larger or equal to 18 (see Figure 6.9b).

Fusion based on LiDAR-derived nDSM properties

In feature-level fusion, the LiDAR-derived nDSM can be used to define feature class as the difference in elevation to neighbouring features. At Level II for fusion of multispectral imagery with nDSM data, *House* and *Infrastructure/Road* classes are discriminated through the mean height differences in the nDSM layer. A feature is represented as *Infrastructure/Road* class if the mean height difference in the nDSM is less than or equal to 1.2m (see Figure 6.10a). The *House* class has mean height greater than 1.2m (see Figure 6.10b).



(a)

(b)

Figure 6.10 Fuzzy membership function for the discrimination of *Roof* and *Open space* using nDSM values

At Level III for fusion of multispectral imagery with LiDAR-derived nDSM data, the *Tree* and *Grass* classes are discriminated on their height difference. *Vegetation* is defined as *Tree* if its height is greater than or equal to 1.8m below that it is *Grass*.

Fusion using contextual information

Normally elevated objects create shadows so most of the *Shadow* features can be detected and described by their source features. Additionally, *Shadow* areas can be classified according to their spectral properties. Pixel-based classification using these properties gave unsatisfactory results (Dare, 2005). Thus it is helpful to classify *Shadows* by describing contextual criteria in addition to any differences in spectral properties. Depending on the type, *Shadows* may inherit their spectral properties from an appropriate super-class and then be identified by their surroundings. The applied logic is: if a feature classified as *Shadow* is surrounded mainly by features classified as *Tree*, it should be classified as *Tree Shadow* (Zhou et al., 2009). At Level III classification, *Roof* and *Tree Shadows* are subclasses of *Shadow*. The different *Shadow* classes are discriminated by using the inherency and neighbour-object relationship. A *Shadow* is classified as a *Roof Shadow* if its border-to-neighbour relation for *Roof* is larger or equal to 0.025m. The same logic is applied for *Tree Shadow*. A *Shadow* object is classed as *Tree Shadow* if its border-to-neighbour for *Tree* is larger or equal to 0.02m.

6.6 Summary

This chapter has described the implementation of the pixel- and feature-level fusions employing user-driven algorithms for the delineation of a range of landscape features within the semi-urban study area. A hierarchical landscape classification system was developed for successful fusion and subsequent mapping of the landscape objects. It was necessary for the classification scheme to be hierarchical in structure such that information extracted from multi-source remotely sensed data could be incorporated into the fusion process. Traditional pixel-level fusion relies on all the facets of spectral characteristics only, whereas feature-level fusion is reliant on automated interpretation utilising spectral, spatial and contextual information.

4-band multispectral imagery with LiDAR-derived nDSM data was used for implementing pixel- and feature-level fusions. In pixel-level fusion, training data were obtained for the eight landscape classes that were subsequently used in a supervised classification. Analysis of training data, as part of the supervised classification process, indicated that most of the landscape classes were separable in spectral, height and textual space. The least amount of spectral separation among the *Shadow* classes is evident, with progressively greater separation in *Natural* and *Manmade* classes.

Feature-level fusion utilised multi-resolution segmentation for identification of meaningful features from multi-source data and subsequently fused the features according to their attributes. A class hierarchy similar to the classification scheme in pixel-level fusion was created for the feature-level fusion task. Spectral properties of the segmented features are used to identify *Vegetation* and *Roof* classes. The LiDAR-derived nDSM properties are used to distinguish *Roads* and *Houses* classes for their distinctive height difference. In addition, contextual attribute such as the border-to-neighbour relationship is utilised to discriminate different classes of *Shadows* (*Roof* and *Tree Shadows*) according to their source feature. A characteristic of this feature-level fusion is that all *Shadow* classes are well delineated due largely to utilising border-to-neighbour relationships.

CHAPTER 7

RESULTS AND DISCUSSION

In this chapter, the results of applying fusion models to landscape mapping are presented for two study areas. Thematic accuracy of the fusion results is assessed using field-surveyed reference data. Comparative thematic accuracies of pixel- and feature-level fusions are also presented. For each study area, the practical implications of pixel- and feature-level fusions are discussed for extracting a range of landscape details derived from remotely sensed data. Finally, research objectives are re-evaluated.

7.1 Data-driven Fusions for the Forest Study Area

The implementation of data-driven fusion models for the forest study area was presented in Chapter 5. Now the results and accuracy assessments of pixel- and feature-level fusion models are separately discussed.

For thematic accuracy assessment, an error matrix is generated by comparison of reference data derived from field-survey and aerial photo interpretation with corresponding samples of the fusion results. Overall thematic accuracy and the Kappa Coefficient are computed to measure the success of object interpretation using the fusion process. The User's and Producer's Accuracies, and elements of the error matrix are evaluated to assess thematic error patterns within each fusion result. Theoretical aspects of these evaluation processes were reviewed in Section 2.8.

Seven plots (marked in Figures 7.1 and 7.2) were selected for assessing the thematic accuracies of the fusion results. A total of 76 points were directly field surveyed within the 7 plots and the rest of the samples were collected from aerial photo interpretation. Descriptions of the sampling and field data collection procedures were given in Section 3.13.2. For preparing the reference dataset, from each plot (50m x 50m or 100 x 100 pixels) 43 points were selected using a stratified random sampling technique. This ensured that each of the four classes in the each plot contained at least ten validation points. In some cases, when all the classes were not present in a plot, additional samples were collected from other plots. Finally, a total of 300 points

(roughly 43 x 7) were identified for validation purposes, which equates with 60 samples per class on average. The thematic content of these points was determined by field survey as well as aerial photo interpretation and used as a reference dataset. Fusion results corresponding to the reference points were collected for accuracy assessment.

7.1.1 Pixel-level fusion results

The final pixel-level fusion map (Figure 7.1) shows that the tree classes have been well clustered and background materials are well separated. Nevertheless the tree cluster map displays minimum spatial homogeneity among tree species. The problem of mixed-pixel, or the ‘salt and pepper’ effect, is very prominent as the pixel-level fusion only uses the pixel spectral values of the multi-source data rather than the size and shape of the trees.

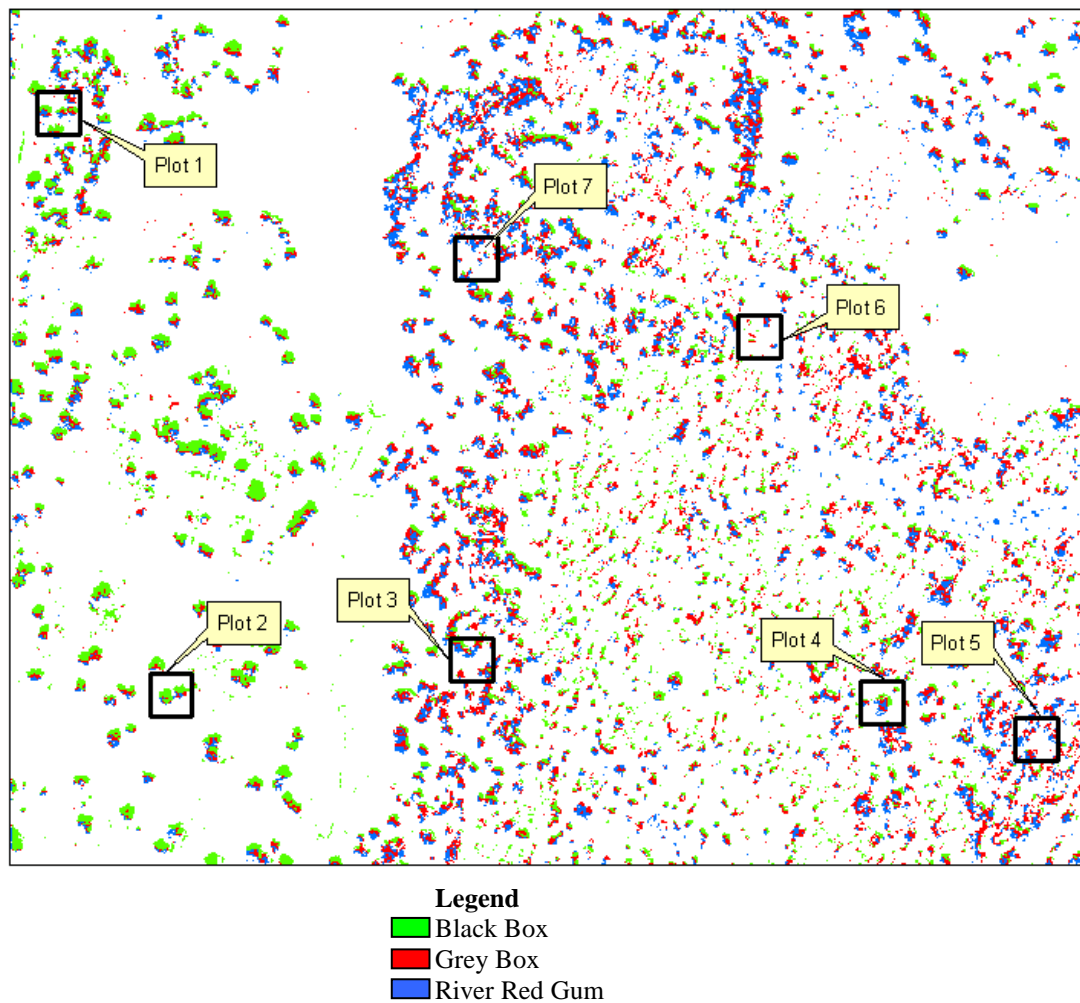


Figure 7.1 Tree species identification using pixel-level fusion of multispectral imagery with LiDAR-derived height and texture data

The main disadvantage of using this pixel-level fusion is that it does not clearly delineate individual trees when they are in a group. It delineates individual trees in an open area but suffers from the mixed-pixel problem as multiple tree species are shown within a singletree canopy. Most of the mixed-pixel problem is due to the presence of understorey and the variation of brightness in the tree canopy areas. The filtering procedures that are used to suppress shadow effects have very little influence on the fusion process as neighbouring pixels still have high spectral contrast.

Much of the success of the pixel-level fusion model depends on separability of tree clusters in the feature space. The statistical definition of separability is quite difficult with respect to the lack of homogeneity in spectral intensity within a tree crown. The heterogeneity is caused mainly by variation in crown structure (shadow effect), but also by other factors such as low crown density exposing background materials and different tree components (bark, stem). These are the major source of class variability even though most sunlit pixels are clustered according to tree species. These difficulties are expected from high spatial resolution datasets that allow analysis at the individual tree level.

7.1.2 Thematic accuracy of pixel-level fusion results

Table 7.1 shows the error matrix for tree species identification using pixel-level unsupervised fusion of multispectral imagery with LiDAR-derived nDSM data. The high User's (84.91 percent) and Producer's (75.47) accuracies for bare ground indicate that background materials are well separated from tree clusters. The User's Accuracy (68.57 percent) and Producer's Accuracy (74.90 percent) are relatively high for the river red gum (*Eucalyptus camaldulensis*) class indicating moderate agreement between fusion results and reference data thus expressing moderate confidence in identification of this species using the pixel-level fused data. Grey box (*Eucalyptus microcarpa*) had the lowest User's (40.98 percent) and Producer's (42.62 percent) accuracies. This result indicated that only 40 percent of the grey box (*Eucalyptus microcarpa*) had been correctly identified by the pixel-level fusion and 60 percent incorrectly identified. Black box (*Eucalyptus largiflorens*) had also not been reliably separated from other species (User's accuracy 49.32 percent and Producer's accuracy 57.14 percent).

Table 7.1 Error matrix for pixel-level fusion using multispectral imagery with LiDAR-derived nDSM data

Fused data	Reference data					User's Acc. %
	Black box	Grey box	River red gum	Bare ground	Total	
Black box	31	21	4	7	63	49.21
Grey box	14	25	13	9	61	40.98
River red gum	8	6	48	8	70	68.57
Bare ground	9	3	4	90	106	84.91
Total	63	61	70	106	300	
Prod Acc %	57.14	42.62	74.90	75.47		194
Overall Accuracy (Percent)						64.67
Kappa Coefficient (\hat{K}) (Percent)						51.69

The overall thematic accuracy is 64.67 percent and the Kappa Coefficient is 51.69 percent. The Kappa Coefficient indicates *moderate* agreement (see Table 2.4 for qualitative interpretation of the Kappa Coefficient) between the fusion results and the reference data. There is too much noise within the datasets in the pixel-level fusion to properly identify the tree species. However, the result indicates that forest and bare grounds can be separated using this method (high User's and Producer's accuracies).

Similar spectral properties and a high degree of within-target spectral variation of the tree species were the contributing factors for low accuracy. Pixel-level fusion of the spectral and height information created a mixed-pixel problem and substantially reduced the accuracy of tree species identification.

7.1.3 Feature-level fusion results

The results of the pixel-level fusion approach showed that the mixed-pixel problem complicated the fusion process and potentially hampered the recognition of tree species. Employing feature-level fusion overcame this problem by defining the tree canopy area first, then extracting the features from the segmentation and finally deriving the tree species map using unsupervised classification of the tree crown features.

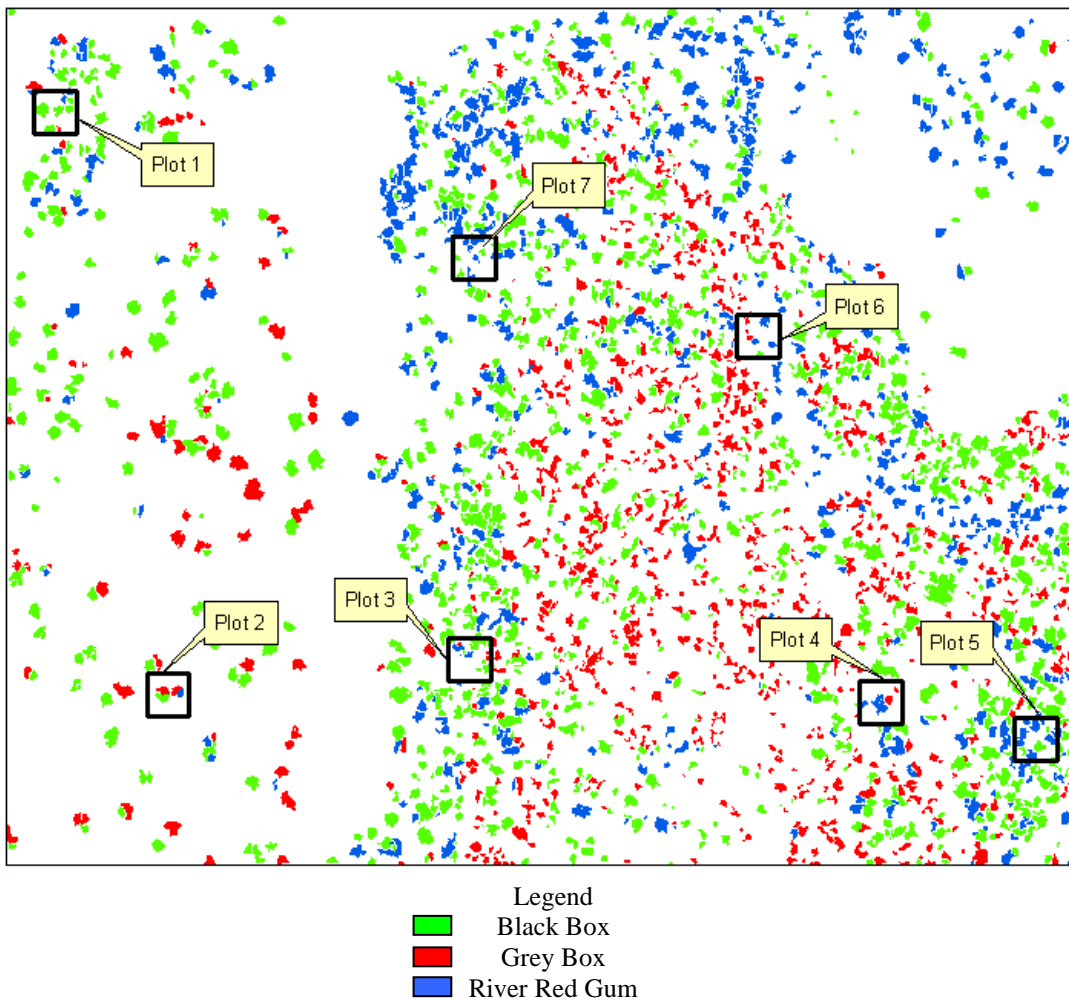


Figure 7.2 Tree species identification using feature-level fusion of multispectral imagery with LiDAR-derived height and texture data

Figure 7.2 illustrates that individual tree crowns were delineated properly and the background materials were well separated from the tree crown areas when feature-level fusion was employed. The tree cluster map displayed higher spatial homogeneity among tree species than did the pixel-level approach.

The advantage of the feature-level fusion was that it classified the whole tree crown area (tree feature) rather than the single pixels. As a result, it reduced the mixed-pixel problem and classified the tree more realistically than pixel-level fusion. The application of the unsupervised classification technique involved in the pixel-level fusion led to poor tree species mapping (accuracy only 64.67 percent) main because of the difficulty in defining the cluster centres and the critical distance parameter to enable individual tree crown recognition. Another reason for poor classification was the confusion within classes due to the noise effects such as shadows, background vegetation and lack of species spectral information.

7.1.4 Thematic accuracy of feature-level fusion results

Data-driven segmentation and subsequent classification were the decisive factors for improvement of the feature-level fusion result over pixel-level fusion. The enhancement can be seen in Figure 7.2 as well as in the error matrix (Table 7.2).

Table 7.2 Error matrix for feature-level fusion using multispectral imagery with LiDAR-derived nDSM data

Fused data	Reference data					User's Acc. %
	Black box	Grey box	River red gum	Bare ground	Total	
Black box	55	4	3	1	63	87.30
Grey box	4	53	1	3	61	86.89
River red gum	0	4	66	0	70	94.29
Bare ground	8	7	6	85	106	80.19
Total	63	61	70	106	300	
Prod Acc %	77.78	77.05	87.14	96.23		259
Overall Accuracy (Percent)						86.33
Kappa Coefficient (\hat{K}) (Percent)						81.61

Bare ground delineation has the highest Producer's accuracy (96.23 percent) and river red gum (*Eucalyptus camaldulensis*) has the highest User's accuracy (94.29 percent). Grey box (*Eucalyptus microcarpa*) and Black Box (*Eucalyptus largiflorens*) have lower Producer's and User's accuracies but identification of all three-tree species was considerably better using feature-level fusion than by using pixel-level fusion (Compare Table 7.2 with Table 7.1). The accuracy of feature-level fusion results outperforms the pixel-level fusion results by around 20 percent.

The overall thematic accuracy was 86.33 percent and the Kappa Coefficient was 81.61 percent. A Kappa Coefficient close to 1 indicates *substantial* agreement (see Table 2.4 for qualitative interpretation of The Kappa Coefficient) between fusion results and reference measurements, meaning that feature-level fusion produced a stable tree species map for the forest study area.

A semi-automated procedure similar to the feature-level fusion methodology was used by Meyer *et al.* (1996), where tree crowns were digitised from colour-infrared imagery, then identified tree species by classifying the pixels within crowns. The use of texture layers and a parallelepiped classifier achieved 80 percent accuracy and

outperformed on average 23 percent of maximum-likelihood classifier using original three-band imagery. Results of this study align with Meyer *et al.* (1996) results confirm that prior segmentation of tree crowns improves species classification. Overall accuracy of the current study (86.33 percent) is better than in Meyer *et al.*'s (1996) study (80 percent) where they used manual digitisation for delineating tree crowns.

Segmentation and subsequent tree feature extraction from all data layers are essential for achieving good results and for a more standardised feature-level fusion. Data-driven feature-level fusion can be used as a model for fusing high spatial resolution multispectral imagery with LiDAR data to assess forest attributes at an individual tree level. This fusion procedure has potential for minimising human involvement by automating interpretation of forest attributes. A similar study conducted by Koukoulas and Blackburn (2005) in broadleaved deciduous forest using airborne LiDAR and multispectral imagery achieved 91 percent accuracy. However, their method did not aim to delineate the whole tree crown but to extract top of the trees with information of their height, location and species type.

The following section presents a statistical comparison of the pixel- and feature-level fusion results.

7.1.5 Comparative accuracy of data-driven fusion results

The accuracy of tree species identification by different fusion models can be compared through visual interpretation. In some cases, map scatter plots from LiDAR combined with multispectral imagery and field data enabled identification of individual black box and grey box trees (*see* Figure 3.3). However, this is not always possible due to the variable datasets so identification is at the stand level only. Sites dominated by black box generally exhibits a lower proportion of single LiDAR returns compared to sites dominated by grey box. River red gum (*Eucalyptus camaldulensis*) can easily be separated for their unique habitat. This species is largely found on riverine wetland areas subject to periodic inundation.

Figure 7.3 shows the accuracy of tree species based on pixel- and feature-level fusions in the forest study area. Feature-level fusion delineates tree crowns more accurately than the pixel-level fusion.

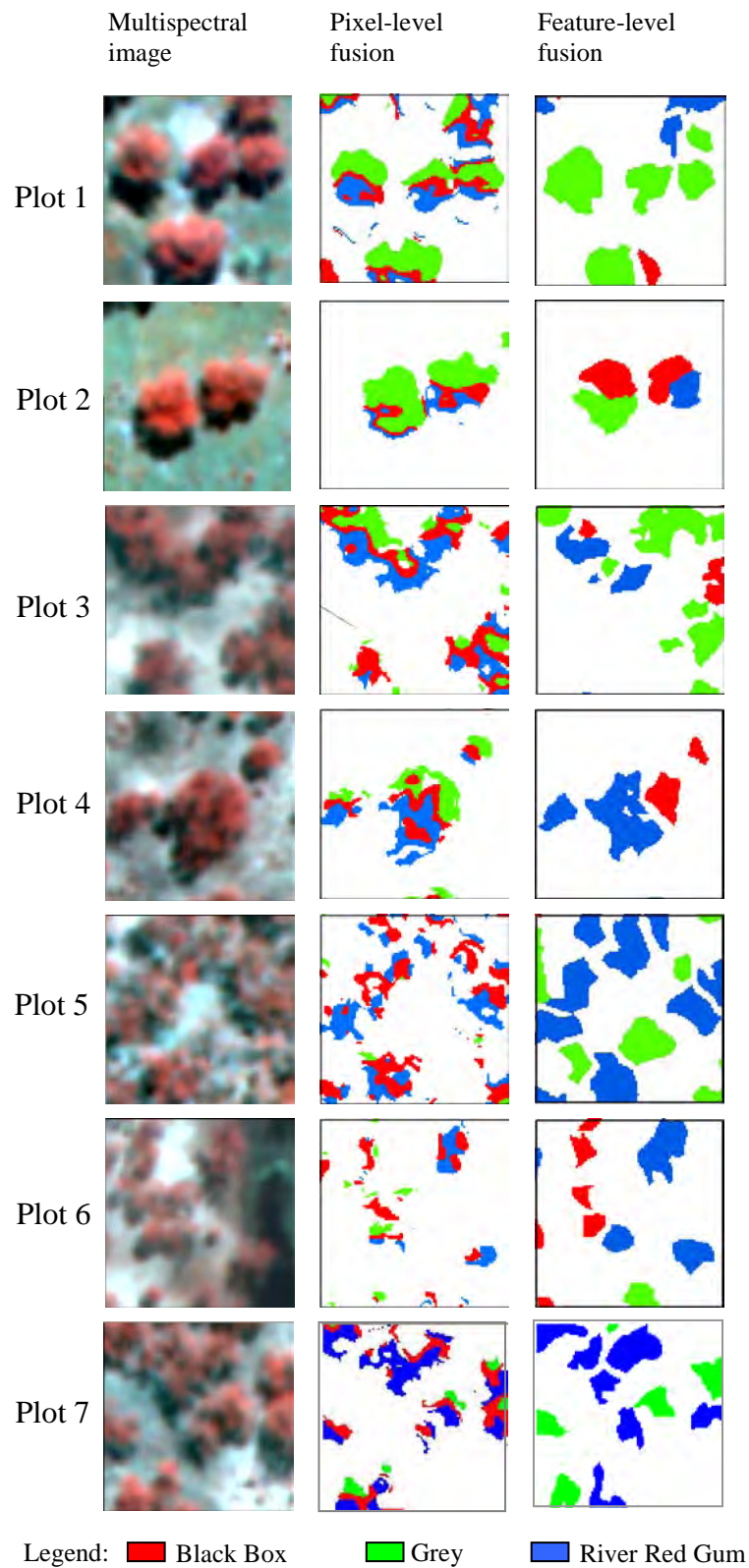


Figure 7.3 Visual comparisons of the pixel- and feature-level fusion results for tree species delineation

A detailed review of comparative accuracy assessment procedure of fusion results was presented in Section 2.9.2. Table 7.3 summarises thematic accuracy results and

provides a comparative Kappa Coefficient values for results derived from pixel- and feature-level fusions. The standard normal derivative of the Kappa Coefficient was used to evaluate whether a significant difference exists between the two results.

Table 7.3 Comparison of accuracy of tree species identification at the 95 percent confidence level using pixel- and feature-level fusions

Fusion	Pixel-level	Feature-level
Overall Accuracy (percent)	64.67	86.33
\hat{K} (percent)	51.69	81.61
$\hat{var}(K)$	0.001188	0.000700
z	15.00	30.84
$z_{K_1-K_2}$	6.68	
Significance*	S	

*NS = not significant, S = significant

Feature-level fusion greatly out-performed the pixel-level fusion in accuracy (86.33 vs 64.67). This was mainly due to the segmentation and subsequent tree feature definition. The Kappa Coefficients of the pixel- and feature-level fusions were 52.56 and 81.21 respectively, which indicated *moderate* and *substantial* agreement with reference measurements (see Table 2.4). Variance of Kappa and the Z test statistics point to thematic accuracies derived from the different fusions being significantly different from random results.

Tree species delineation results from pixel- and feature-level fusion (Table 7.3) indicated that, at the 95 percent confidence level, all computed values of the standard normal deviation of the Kappa Coefficient (Equation 2.25) were far greater than 1.96 (actual value 6.68) confirming that tree species delineation using feature-level fusion was significantly more accurate than pixel-level fusion result.

7.2 User-driven Fusions for the Semi-urban Study Area

The implementation of user-driven fusion models for identifying landscape objects of the semi-urban study area was presented in Chapter 6. Now the results and their thematic accuracy are discussed.

To evaluate thematic accuracy of the user-driven fusion results, a random sample of 200 points was generated across the semi-urban (1.25km x 1km) study areas that were not training sites. This random sample was used to determine the relative proportion of each landscape class within the image. Following Congalton and Green (1999), the random sample was supplemented by a stratified random sample of 256 point. This ensured that each class in the fusion results contained at least seventy validation points. The reference data for the 256 points were collected from the field and for the 200 points from aerial photos. Descriptions of the sampling techniques and the field data collection procedures are given in Section 3.13.2. Fusion results corresponding to the reference samples were assessed for thematic accuracy using an error matrix. Theoretical aspects of these evaluation processes were reviewed in Section 2.8.

7.2.1 Pixel-level fusion results

Pixel-level supervised fusion of multispectral imagery with LiDAR-derived nDSM data using the parallelepiped classifier is presented in Figure 7.4. The *Road/Open space* and *Grass* classes are relatively uniform and the boundaries compare favourably with the reference data shown in Figure 6.4. The *Roof* class is distributed all over *Road/Open space* classes, but their locations are not well matched with the reference data. The sealed *Road* appears to have been replaced by the *Roof* class. This effect is visible throughout the semi-urban study area indicating spectral confusion between these classes.



LEGEND







	Road/Open space		Grass
	Roof		Tree
	Roof shadow		Tree shadow

Figure 7.4 The results of the pixel-level fusion of multispectral imagery with LiDAR-derived nDSM data

However, the shadow and non-shadow areas were well delineated in pixel-level fusion and was consistent with the Zhou *et al.* (2009) and Dare (2005) studies. The most challenging part of shadow detection is to separate different types of shadows (such as *Roof* and *Tree Shadows*) as they almost have identical radiometric responses. In this study, *Roof shadow* class occupies most of the shadow areas and yet few *Tree shadow* areas are properly delineated. This study had proven that different types of shadows could not be distinguished solely from the radiometric responses.

7.2.2 Thematic accuracy of pixel-level fusion results

Thematic accuracy assessment with the Kappa Coefficient provides a measurement of the success of the application of different fusion models. A framework for thematic accuracy assessment was presented in Section 2.8. User's and Producer's Accuracies, and elements of the error matrix are a gauge of the accuracy of fusion results. Thematic accuracy of pixel-level fusion results for the semi-urban study area can be assessed from Table 7.4.

Table 7.4 Error matrix for pixel-level fusion results using multispectral imagery with LiDAR data

Fused data	Reference data							User's Acc. %
	Roof	Roof shadow	Tree	Tree shadow	Grass	Open space	Total	
Roof	58	5	4	2	1	8	78	74.36
Roof shadow	2	54	7	11		2	76	71.05
Tree	4		58	8	5	3	78	74.36
Tree shadow	5	9	4	49	2	3	72	68.06
Grass		4	7	1	56	6	74	75.68
Open space	5	2	3	2	7	59	78	75.64
Total	74	74	83	73	71	81	456	
Prod Acc %	78.38	72.97	69.88	67.12	78.87	72.84		334
Overall Accuracy (Percent)								73.25
Kappa Coefficient (\hat{K}) (Percent)								67.89

The overall thematic accuracy is 73.25 percent and the Kappa Coefficient is 67.89 percent. The Kappa Coefficient indicates *moderate* agreement (see Table 2.4 for qualitative interpretation of The Kappa Coefficient) between the fusion results and the reference data. In pixel-level fusion results, 'pepper and salt' effects are prominent due to the heterogeneous spectral response within the objects. Inclusion of nDSM data has very little influence in reducing the 'pepper and salt' effects. The *Roof* and *Open space* classes have higher User's and Producer's accuracies however in many cases, the *Roof* class is mis-identified as *Open space*. LiDAR height data is apparently not sufficiently well matched with multispectral imagery. This mis-identification is visible at the edges of *Roofs* in Figure 7.4. The identification of

Grass class has achieved the highest User's accuracy (75.68 percent) and Producer's accuracy (78.87 percent). The high User's accuracy (74.36 percent) of the *Tree* class indicates that spectral properties and LiDAR-derived height data play a significant role in separating *Tree* from *Grass*.

The *Shadow* classes are poorly separated as the low User's and Producer's accuracies show; this is also revealed visually (see Figure 7.4). *Roof Shadow* and *Tree Shadow* classes are particularly difficult to distinguish using only spectral signatures since they have very little spectral information and have a high degree of within-class variation in the multispectral image owing to high spatial resolution of multispectral imagery resulting in mixed-pixels (Zhou *et al.*, 2009; Dare, 2005).

7.2.3 Feature-level fusion results

Variation in spatial, spectral and radiometric resolution of the sensors is important because of the influence of these on discrimination of spectral, radiometric and spatial properties of the objects to be extracted. Different radiometric resolution such as 16-bit multispectral imagery and 8-bit LiDAR data provided rich datasets for fusion. Pixel-level fusion does not consider relative size, shape and distribution of landscape objects as a result; relation between objects cannot be used in this fusion process. Therefore, an alternative fusion technique needs to be considering for extracting landscape information. Figure 7.5 shows feature-level fusion results for delineating Level III landscape objects from multispectral imagery with LiDAR-derived nDSM data. *Roads/Open space* and *Grass* classes form the largest components of the scene; however, the *Tree* class also occupying significant areas. All main classes, especially *Roof* class, have smooth shapes compared to the roof shapes in the pixel-level fusion result.



LEGEND







	Road/Open space		Grass
	Roof		Tree
	Roof shadow		Tree shadow

Figure 7.5 Results of the feature-level fusion of multispectral imagery with LiDAR-derived nDSM data

An emerging characteristic of this feature-level fusion process is that all *Shadow* classes are well delineated due largely to each *Shadow* classes being explicitly defined through the radiometric responses and border-to-neighbour relationships. *Roof* and *Tree Shadow* objects are delineated using spectral signature from multispectral imagery and contextual information border-to-neighbour relationships derived from spatial analysis. Zhou (2009) used the same technique for delineating shadow sub-classes and his results were consistent with those of the current study.

7.2.4 Thematic accuracy of feature-level fusion results

The feature-level fusion of data for the semi-urban study area utilised user provided guidelines for segmentation and subsequent features classification employing

knowledge rules derived from colour, height and contextual properties. The visual depiction of the resulting classification indicated that feature-level fusion gave a superior result to pixel-level fusion (see Figure 7.5). Thematic accuracy of the feature-level fusion is assessed using an error matrix in Table 7.5.

Table 7.5 Error matrix for feature-level fusion using multispectral imagery with LiDAR data results

Fused data	Reference data							User's Acc. %
	Roof	Roof shadow	Tree	Tree shadow	Grass	Open space	Total	
Roof	73		2			3	78	93.59
Roof shadow		70	1	14		1	76	92.11
Tree	3	4	67	1	3		78	85.90
Tree shadow		10		60	2		72	83.33
Grass			5	2	64	3	74	86.49
Open space	5		1		3	69	78	88.46
Total	81	84	76	67	72	76	456	
Prod Acc %	90.12	83.33	88.16	89.55	88.89	90.79		403
Overall Accuracy (Percent)								88.38
Kappa Coefficient (\hat{K}) (Percent)								86.05

The overall thematic accuracy for the feature-level fusion is 88.38 percent and the Kappa Coefficient is 86.05 percent. The Kappa Coefficient indicates *excellent* agreement (see Table 2.4 for qualitative interpretation of The Kappa Coefficient) between the fusion results and the reference data.

Delineation of *Roof* class has the highest User's Accuracy (93.59 percent) and Producer's Accuracy (90.12 percent). Delineation of *Tree* and *Roof shadow* classes has improved dramatically using feature-level fusion as the contextual attribute border-to-neighbour relation is applied for defining a specific *Shadow* class Producer's Accuracy 88.16 percent and 85.90 percent respectively, and User's Accuracy 83.33 percent and 92.11 percent respectively. In pixel-level fusion, only spectral properties are used to separate *Shadow* classes and since shadows have the same or very similar spectral properties this approach fails to adequately separate *Tree Shadow* from *Roof Shadow* (Dare, 2005 and this study). On the other hand,

feature-level fusion uses contextual properties as well allowing accurate delineation of different *Shadows*. Zhou *et al.* (2009) used similar methodology for separating of shadows with overall accuracy of 81.5 percent, which is closed to the accuracy of current study average 88.40 percent. Interestingly, Zhou *et al.* (2009) achieved 83.1 percent User's accuracy for the *Tree Shadow* class; compared with this study's 83.33 percent. The *Roof Shadow* of Zhou *et al.*'s (2009) study had a low User's accuracy (66.7 percent) and a very high Producer's accuracy (95.2 percent), which indicated user's had difficulties to delineate *Roof Shadows*.

7.2.5 Comparative accuracy of results for user-driven fusions

Visual comparison between pixel- and feature-level fusion results (Figure 7.4 and 7.5) shows some substantial differences. Error matrices for thematic accuracy of the fusion results (Table 7.4 and 7.5) statistically confirm these differences. Overall accuracies of the pixel- and feature-level fusion results are 73.25 and 88.38 percent, respectively. The 15.13 percent difference between these values indicates that these two fusion processes produce very different results. A similar type of study by Platt and Rapoza (2008) also revealed that the accuracy of feature-level classification of multispectral imagery for urban area was considerably better than pixel-level classification (overall accuracy greater by 14 percent).

Feature-level fusion results landscape objects being better distinguished than by pixel-level fusion is well delineated due to contextual properties of the landscape features being taken into account. The pixel-level fusion results show considerable fragmentation of landscape objects.

The greatest improvement through using feature-level fusion is in the delineation of the *Shadow* classes. The User's and Producer's Accuracies are substantially better than pixel-level fusion results. In feature-level fusion, *Roof Shadow*'s User's and Producer's accuracies are 92.11 and 83.33 percent respectively. Compare to 71.05 and 72.97 percent in pixel-level fusion results. The *Tree Shadow* results are the same as *Roof Shadow* results for both fusion cases. The incorporation of contextual properties into the feature-level fusion is what leads to better results. Other studies done by Zhou *et al.* (2009), Yuan (2008) and Dare (2005) support this.

Kappa Coefficients derived from the same error matrices can be utilised to measure the level of agreements between pixel- and feature-level fusion results. Kappa Coefficients compares the degree of interclass confusion for better understanding of the thematic error patterns. This measurement helps to select appropriate fusion algorithm for achieving better thematic accuracy that is suitable for a particular study. A theoretical background and the process of statistical comparison were presented in Section 2.9.2.

Table 7.6 Comparison of accuracy of pixel- and feature-level fusion results at the 95 percent confidence level

Fusion	Multispectral imagery and LiDAR fusion	
	Pixel-level	Feature-level
Overall Accuracy (percent)	73.25	88.38
\hat{K} (Percent)	67.89	86.71
$\hat{var}(\hat{K})$	0.001269	0.001140
z	19.05	25.48
$z_{\hat{K}_1 - \hat{K}_2}$	3.70	
Significance*	S	

*NS = not significant, S = significant

Table 7.6 shows the z statistics for both fusion results, 19.05 for pixel-level fusion and 25.48 for feature-level fusion. At the 95 percent confidence level, the standard normal deviation of the Kappa Coefficient is 3.70, well above the 1.96 threshold showing that feature-level fusion of multispectral imagery with LiDAR data leads to significantly more accurate mapping of a semi-urban landscape than pixel-level fusion.

Both for forest and semi-urban study area the overall accuracy, User's and Producer's accuracies and pair-wise comparison of Kappa Coefficients revealed that feature-level fusion results were superior to the pixel-level fusion. The feature delineation using segmentation techniques and classification of features rather than pixels were the decisive factors for improvement over pixel-level fusions.

7.3 Discussion

The main goal of this research was to investigate different fusion models using aerial imagery with LiDAR-derived height data for a range of mapping applications particularly for forest and semi-urban landscape mapping. Pixel- and feature-level fusions were particularly investigated in order to evaluate their potential. Thematic accuracies for the fusion-driven objects were evaluated against ground truth reference data to identify optimum fusion models for two study areas. The following sections discuss findings in light of the original research objectives laid out in Section 1.4.

7.3.1 Methodologies for the fusion of aerial imagery with LiDAR data

The underlying objective of this research was to explore data- and user-driven models using pixel- and feature-level fusions for mapping forest and semi-urban areas. Data-driven models were applied to the forest study area to detect tree species. User-driven models were used for the semi-urban area because data only was not enough to separate the complex landscape classes. In the literature, pixel-level fusions using data- and user-driven models are well understood and have been tested widely (Richards and Jia, 2005; Rottensteiner *et al.*, 2004a; Stein, 2004; Teggi *et al.*, 2003; Laporterie and Flouzat, 2003; Pohl and van-Genderen, 1998; Schistad-Solberg *et al.*, 1994). In recent years, OBIA, or feature-level fusions, have been widely used (Blaschke, 2010; Johansen *et al.*, 2010; Hay and Castilla, 2008; Lang, 2008; Platt and Rapoza, 2008; Zhou and Troy, 2008; Yuan and Bauer, 2006); however their applications is still based on user-driven models. This research comprehensively reviewed and implemented the data- and user-driven models in pixel-and feature-level fusions for landscape mapping using aerial imagery and LiDAR data. A major contribution of this research was the introduction of data-driven fusion models in a forest study area. In a feature-level fusion scenario, this model used a watershed segmentation method to delineate tree crowns and then automatically identify tree species using an unsupervised classification algorithm. Another major contribution is the comprehensive review of different fusion techniques and their application for different landscape object extraction.

This study determined efficacy by comparing accuracy of the fusion results with field-surveyed reference data in error matrices. Comparative statistical analysis was

also conducted to find the best possible fusion model for identifying landscape objects for each study areas. The performance of feature-level fusions was, on average, 15 percent more accurate than pixel-level fusions for both forest and semi-urban study areas. A review of the literature (Geneletti and Gorte, 2003; Walter, 2004; Wang *et al.*, 2004; Kamagata *et al.*, 2005; Platt and Rapoza, 2008; Riggan and Weih, 2009; Zhou *et al.*, 2009) also support this assertion.

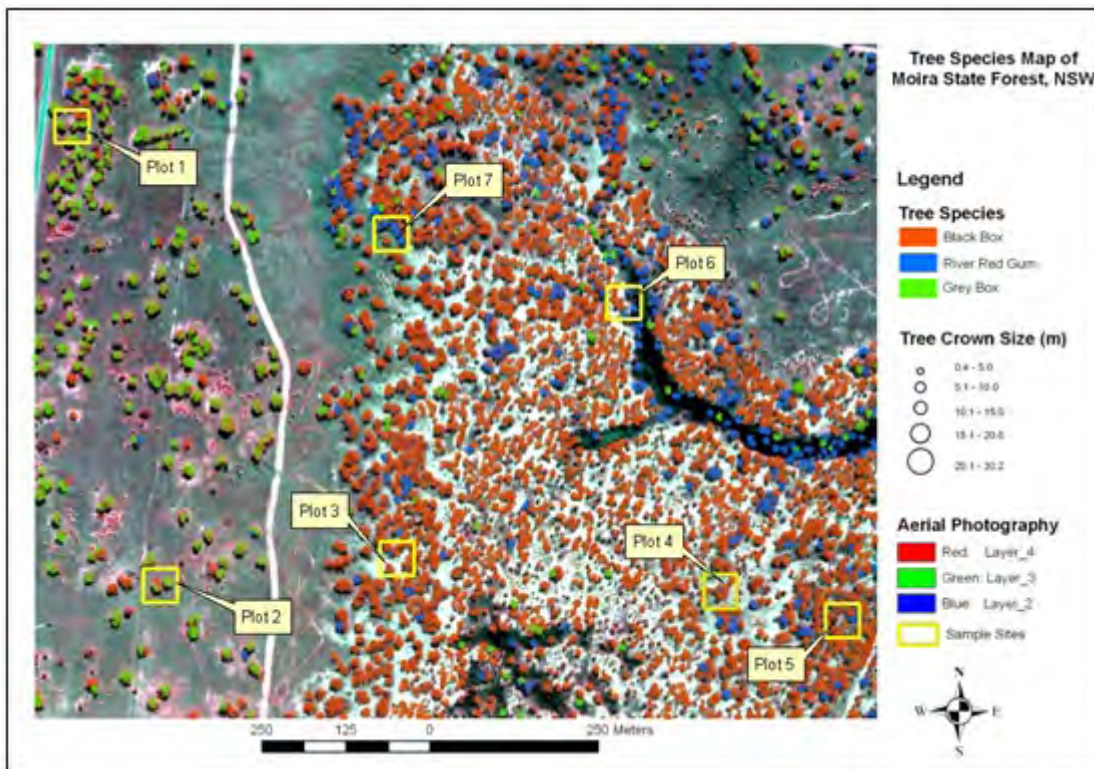


Figure 7.6 Final tree species map for the forest study area derived from feature-level fusion of multi spectral imagery and LiDAR data

Data-driven segmentation and subsequent feature-level fusion substantially improved tree species identification compared with pixel-level fusion. Figure 7.6 shows the tree species distribution within the forest study area using the data-driven feature-level fusion model. Feature-level fusion of multispectral imagery with LiDAR data not only delineated individual tree species but also extracted different tree attributes such as the individual tree crown diameters and tree heights without user interaction. Another advantage of this fusion process is that the derived tree attributes are readily available for use by GIS applications for forest management and forest ecological studies.

For the semi-urban study, user-driven fusion models were found more appropriate than data-derived models for number of reasons. Firstly, data-driven watershed segmentation uses only simple grey-scale imagery with limited measurement space on which to base segmentation. The process does not use all the data layers (4-band multispectral imagery with LiDAR-derived height and texture data) together in a segmentation process. Large number of semi-urban classes could not be delineated using this technique and produced highly unsuitable segmented features (Wealands *et al.*, 2005). User-driven segmentation such as multi-resolution segmentation on the other hand uses all the data layers and utilises scale, shape, texture and spectral properties to segment semi-urban features. Secondly, the number of semi-urban landscape classes was large (six) compared to the number of forest landscape classes (only three main classes). As a result, without *a priori* knowledge of the objects, which had complex spectral and spatial signatures, could not be extracted using data-driven fusion alone. Even user-derived pixel-level fusion did not produce satisfactory results so data-driven fusion model would not produce good results either. Contextual properties can only be utilised in fusion as knowledge rules, therefore without User's intervention the data-driven fusion model is required to allow feature-level fusion.

For the semi-urban study area, feature-level fusion utilised user-driven multi-resolution segmentation for identification of meaningful features from 4-band multispectral imagery with LiDAR data. A class hierarchy was created for the feature-level fusion using spectral, spatial and contextual properties to classify the features. Feature-level fusion overcomes the within-feature variation, whereas pixel-level fusion misclassifies them. For a given classification scheme, the finer the spatial resolution, the greater the chances of within-feature variation (Aplin *et al.*, 1999; Carleer *et al.*, 2005). This study reveals that the delineation of small semi-urban objects (such as *Roof* and *Trees*) is achieved with considerably greater accuracy using feature-level fusion than by pixel-level fusion.

The overall accuracy of feature-level fusion results was 88 percent, which was 15 percent higher than the results using pixel-level fusion. In both data- and user-driven models, feature-level fusions significantly improved results compared to pixel-level fusions owing to the use of segmentation and classification of features rather than just pixels.

7.3.2 The use of colour imagery as an alternative multispectral imagery in data fusions

4-band multispectral imagery was fused with LiDAR data for delineation of semi-urban objects. The recent availability of high spatial resolution colour imagery from satellite sensors such as IKONOS and QuickBird, as well as from digital aerial platforms, provides new opportunities for detailed urban land cover mapping at very fine scale (Zhou and Troy, 2008). On an experimental basis, 4-band multi spectral imagery was replaced by 3-band colour imagery fused with LiDAR-derived nDSM data. Subset data for the semi-urban study area were selected and similar pixel- and feature-level fusion models were applied. The thematic accuracies of these fusion results were compared with multispectral and LiDAR data fusion results. Detailed implementation and analysis are presented in Appendix 4.

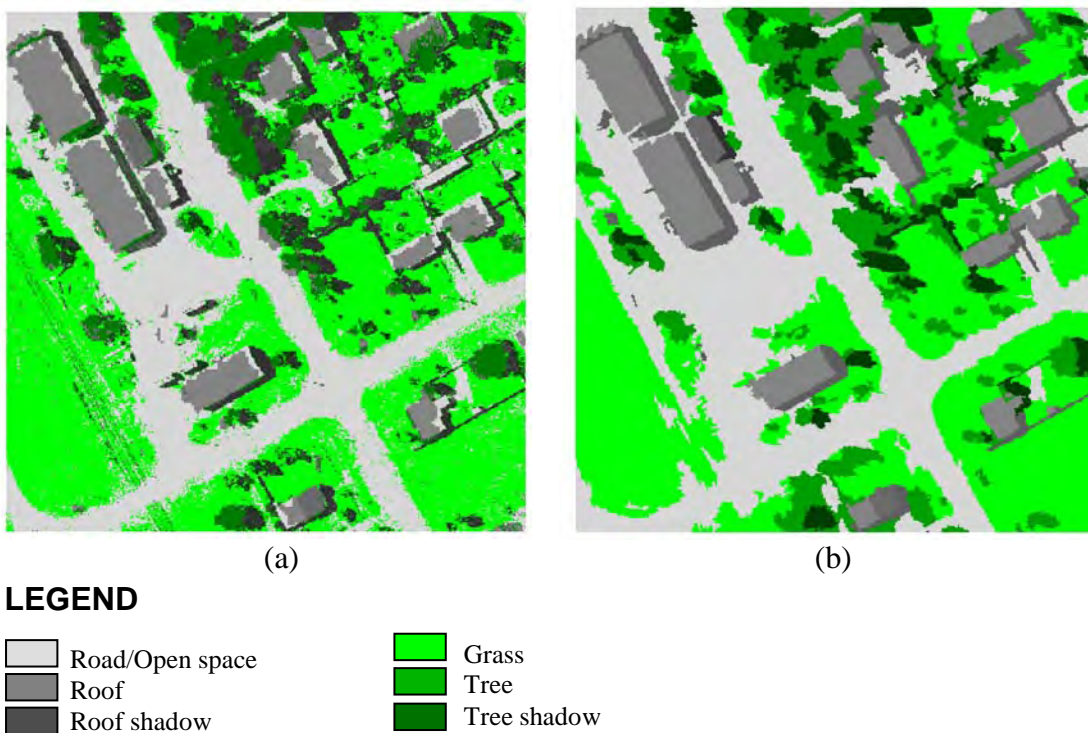


Figure 7.7 3-band colour imagery with nDSM data fusion results (a) pixel-level fusion and (b) feature-level fusions

The main contributing factor of colour imagery to the poor performance of the pixel-level fusion is the failure of the clustering process to recognize some classes. This problem was acute for the *Shadow* classes, as it could not distinguish different shadow classes on the basis of pixel's spectral content alone. The feature-level approach overcomes this problem by incorporating contextual information in the

fusion process. The replacement of multispectral imagery with colour imagery made the tiniest accuracy difference in classification by both pixel- and feature-level fusions. Multispectral imagery made *Shadow* extracting more accurate in pixel-level fusion with higher User's and Producer's accuracies than when using colour imagery (Table 7.4 and Table A4a). This was largely due to the variation of radiometric depth between the colour and multispectral imagery. The 16-bit multispectral imagery has more radiometric depth than 8-bit colour imagery: as a result, multispectral imagery with LiDAR data fusion had less misclassification than colour imagery with LiDAR data fusions. This findings agrees with Leberl and Gruber (2005) conclusions when they noticed that using imagery with high spectral depth (such as 16-bit) gave a significant improvement in classification compared to using imagery with low spectral depth (such as 8-bit).

Table 7.7 Comparison of pixel- and feature-level fusion results using colour imagery and LiDAR data (summary of Table A4 a & c)

Fusion	Colour imagery and LiDAR fusion	
	Pixel-level	Feature-level
Overall Accuracy (percent)	75.24	91.32
\hat{K} (Percent)	70.29	89.93
$\hat{var}(\hat{K})$	0.002594	0.001282
z	13.80	25.11
$z_{K_1 - K_2}$	3.15	
Significance*	S	

At the 95 percent confidence level *NS = difference not significant, S = difference significant

Shadow classification of this study can be compare to Zhou *et al.* (2009) study where object-based multisource data fusion achieved 88 percent overall accuracy. This study concluded that feature-level fusion provides an effective way to delineate shadow classes when multi-source data has radiometric difference and spatial misregistration. Application of the shadow detection and restoration methods helps to eliminate the shadow problem in land cover classification of high spatial resolution images in urban settings.

7.3.3 The effect of height data in fusion models

In this study, optical imagery and LiDAR-derived height data were fused in pixel- and feature-level models. The complementary nature of these two sources of data was the main reason for fusing them, however for argument's sake, the LiDAR-derived height data was excluded from fusion presented in Appendix 5 and accuracies of resulting classification compared with the classification obtained with colour imagery and LiDAR-derived height data fusion presented in Appendix 4.

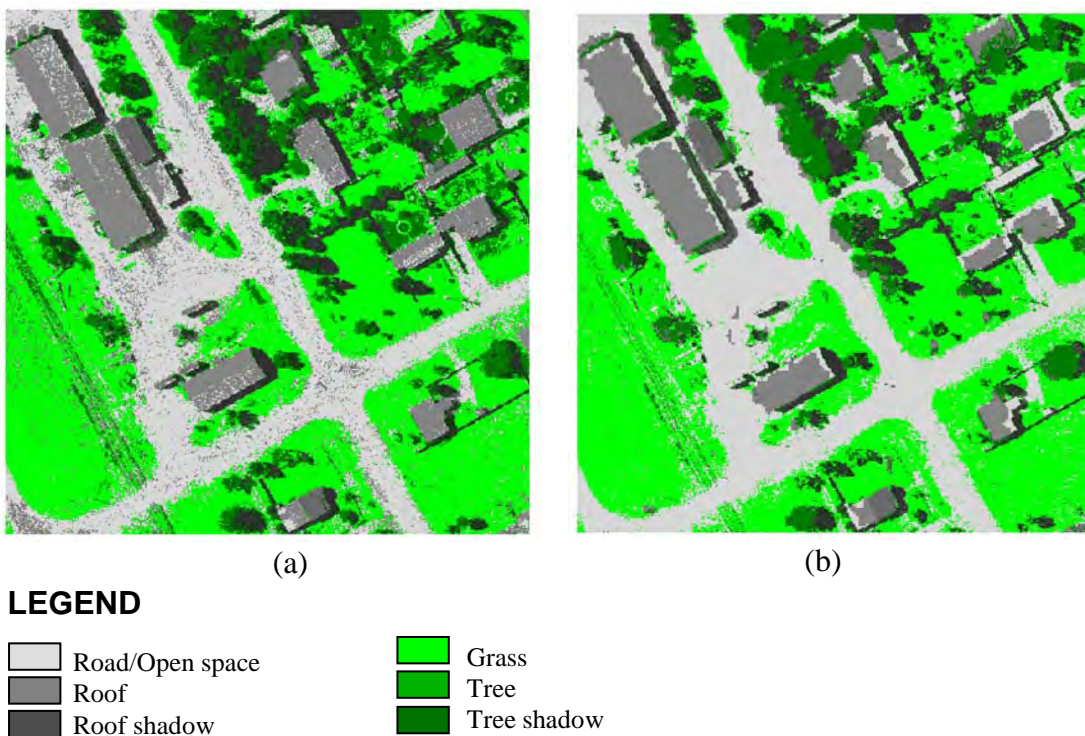
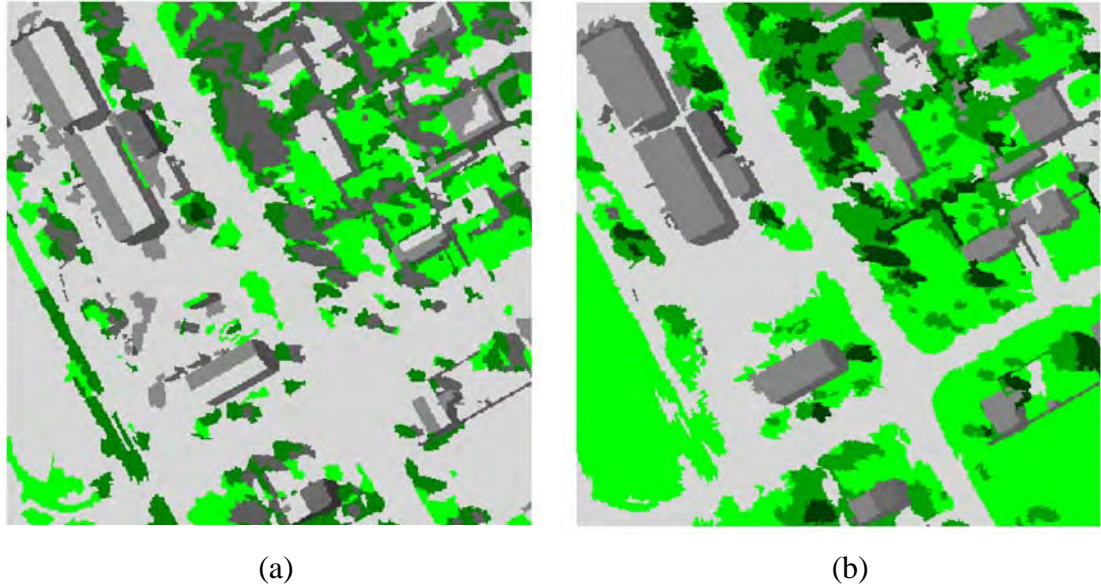


Figure 7.8 (a) Pixel-level classification of colour image, (b) Pixel-level fusion of colour image with LiDAR-derived nDSM data

Visually inspection of the results (Figure 7.8a and 7.9a) showed that object identification was severely disadvantaged by the exclusion of nDSM data. The inclusion of nDSM data improved the sharpness of object images (Figure 7.8b and 7.9b). This is particularly evident when observing *open space* and *roof* classes in pixel-level fusion of colour imagery with nDSM data. Exclusion of nDSM data gave a ‘pepper and salt’ effect due to the multiple misclassifications (*see* Figure 7.8a). Inclusion of LiDAR-derived nDSM results reduced the ‘pepper and salt’ effect and improved landscape mapping. However, misclassification of *shadow* classes still occurred (*see* Figure 7.8b) due to the limitation of the pixel-level fusion process. This

was addressed in feature-level fusion by incorporating contextual information. Obviously height data enhances object definition and so improves the capacity of both pixel- and feature-level fusions to correctly identify them.



LEGEND







	Road/Open space		Grass
	Roof		Tree
	Roof shadow		Tree shadow

Figure 7.9 (a) Feature-level classification of colour image, (b) Feature-level fusion of colour image with LiDAR-derived nDSM data

Object-based classification of colour imagery exhibited poor delineation results (see Figure 7.9a) as abrupt spectral changes occurred within the same classes due to the sun illumination angle. Again, inclusion of height data in the fusion process substantially improved object extraction (see Figure 7.9b). Comparative visual analysis of feature-level fusions showed that additional height data also improved the recognition of objects.

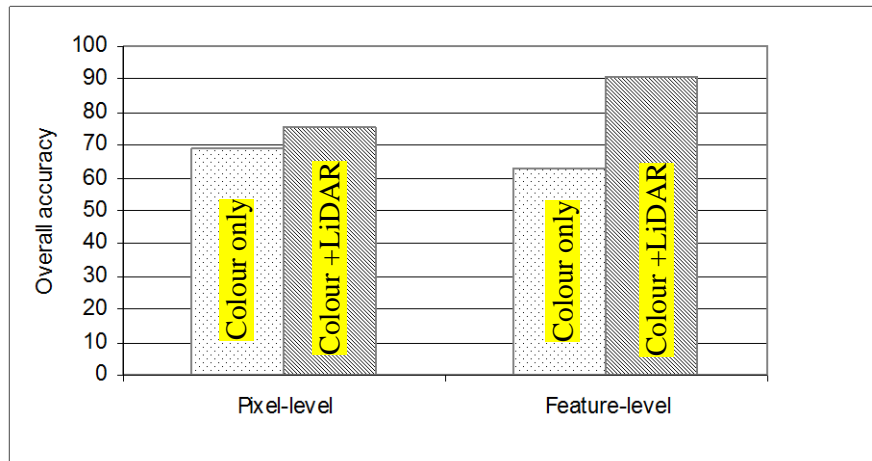


Figure 7.10 Summary of the overall object delineation accuracy of pixel- and feature-level fusions for the semi-urban study area

Classification accuracy when excluding or including nDSM data with colour imagery is presented in Figure 7.10. Implementation to classify a semi-urban area is presented in Appendix 4 and 5. Feature-level fusion of colour with nDSM data achieved the highest object delineation accuracy (91.32 percent), whereas using only colour imagery gave the worst accuracy result (66.56 percent). Abrupt spectral changes within the same classes, specifically in the *Roof* class, were the main reason for misclassification in feature-level fusion. In pixel-level approaches inclusion of nDSM data had less impact, whilst in feature-level approaches this had a significant effect. Both visual and statistical analyses (Appendix 4 and 5) show that inclusion of nDSM improves delineation of landscape objects.

7.3.4 Landscape object height derived from different DSM sources

Digital Terrain Model (DTM) not only represents terrain but also includes other derivatives such as elevation, slope, gradient, aspect, profile curvature, plane curvatures, and other geographical/environmental characteristics (Qiming *et al.*, 2008). These derivatives can be important for object classification (Dragut and Blaschke, 2008). For an example, we know that River red gum (*Eucalyptus camaldulensis*) only grows in riverine flood plains, therefore, if we can derive flood plains from DTM then we can use this attribute to assist with tree species grouping. Feature-level fusion using OBIA can use this sort of derivative as contextual information - i.e. as knowledge rules but such terrain derivatives can not be utilised by data-driven feature-level fusion due to the inability of this model to handle

knowledge rules. Pixel-level fusion cannot utilise terrain parameters either because only the values within pixels can be used in the classification.

Both aerial photography and LiDAR-derived DSMs have been used extensively for generating landscape object heights. Aerial photography has a long history of producing DSMs through analogue, analytical and digital methods (Mikhail *et al.*, 2001). The fundamental difference between LiDAR and photogrammetry is that LiDAR is based on a range measurement to a point from a single airborne position. Photogrammetry however, is based on stereo matching of images from two airborne positions. The stereo matching process requires the matching of a 'patch' of pixels covering a small area rather than a discrete point (footprint) as with LiDAR. In addition, often the algorithms used in the photogrammetric solution have been designed for smooth landscape modelling rather than the rapidly changing elevations of buildings in an urban environment (Smith *et al.*, 2004). According to Baltsavias (1999), LiDAR has some strengths compared to the photogrammetrically-derived DSM. The following sections discussed these issues in forest and semi-urban contexts.

DSM for forest study areas

For the conventional photogrammetric technique, creating automated Digital Surface Model (DSM) over a dense forest is prone to error because of the difficulty of matching pixels in the tree canopy between two stereo photos taken at different viewing angles. LiDAR is a good alternative tool to create better DSMs (Kato *et al.*, 2010). Leckie *et al.* (2003) compared LiDAR-derived Canopy Height Model (CHM) with the results from digital aerial photos. They found that digital aerial photos were better for delineating tree crowns, while the LiDAR-derived CHM was better for open canopy areas. The LiDAR-derived CHM was better suited in clearly distinct area among neighbouring pixels between trees and shrub (or ground). LiDAR was well suited to measure tree height and large tree crown delineation.

LiDAR provides ground and tree height information simultaneously. The penetration rate mainly depends on types of trees (deciduous or coniferous), season and the terrain roughness. Large-footprint LiDAR sensors can achieve results in dense tropical forests where 95 percent of the ground may be obstructed. The

photogrammetric method of image matching delivers very poor results due to shadow, less texture and the significant geometric and radiometric differences between images (Baltsavias, 1999). Automated photogrammetric matching process may provide high-density DSMs but matching results include locally significant errors that require manual editing. By contrast, LiDAR records first and last returns that normally represent the top and bottom of the objects. From these two returns high-density object height data can be derived. The height estimates derived from LiDAR are considered to be more reliable for the reasons that the highest point of the crown is located objectively. The apparently the highest point of a tree from the ground varies with observer's position lead into errors in field measurements. Note that wind effects on the crown might lead to minor errors in LiDAR-derived measurements too.

Results from previous studies had shown that separating northern hemisphere deciduous tree species using LiDAR data was difficult due to their complex structure (Chen *et al.*, 2006). However, use of the watershed segmentation algorithm with the LiDAR data achieved a satisfactory result for eucalypt trees. The success of the tree crown extraction algorithm was higher for old growth than for more juvenile trees where the crowns were more numerous and more scattered. It was also observed that large crowns were better delineated than small ones.

There were several reasons for measuring tree height and crown size from the LiDAR-derived DSMs rather than the stereo photogrammetrically-derived DSMs:

- (1) it is easier to identify individual trees from LiDAR-derived DSM due to the high pulse density of the LiDAR dataset,
- (2) additional texture information can be extracted from LiDAR intensity data, and
- (3) sampling in a LiDAR-derived canopy model can greatly reduce the workload and is not limited by the factors such as accessibility in the field.

Aerial photogrammetric techniques cannot reconstruct the objects accurately in three-dimensional spaces if they are not clearly visible on the photos (Helt *et al.*, 2006). On the other hand, LiDAR provides tree crown structure in a three-

dimensional space. Therefore, LiDAR-derived DSM can only be used in conjunction with optical imagery for tree delineation.

DSM for urban study areas

DSMs support the classification of urban structures beyond two-dimensional classifications. Wurmab *et al.* (2011) presented a hierarchical, object-based transferable framework to extract the urban structure at a high level of geometric detail for two test sites in Germany. Results showed accuracies above 90 percent for the land-use/land-cover classification for both test sites applying the same routines. DSMs from LiDAR and stereo-photo sources have been utilised for the extraction of individual building structures with accuracies of 90 percent and 80 percent, respectively. The LiDAR-derived methodology was suited to extract the urban structure at the level of individual buildings.

In urban environments, DSM-derived from LiDAR provide very dense and accurate 3D measurements of objects with sharp discontinuities, especially buildings. On the other hand, photogrammetric- derived DSMs are less detailed and smooth discontinuities. In addition photogrammetric image-matching include locally significant errors that require manual editing such as, low texture, shadows, multiple solutions, geometric and radiometric differences between the images, and poor approximate values, which are not applicable for LiDAR-derived DSM.

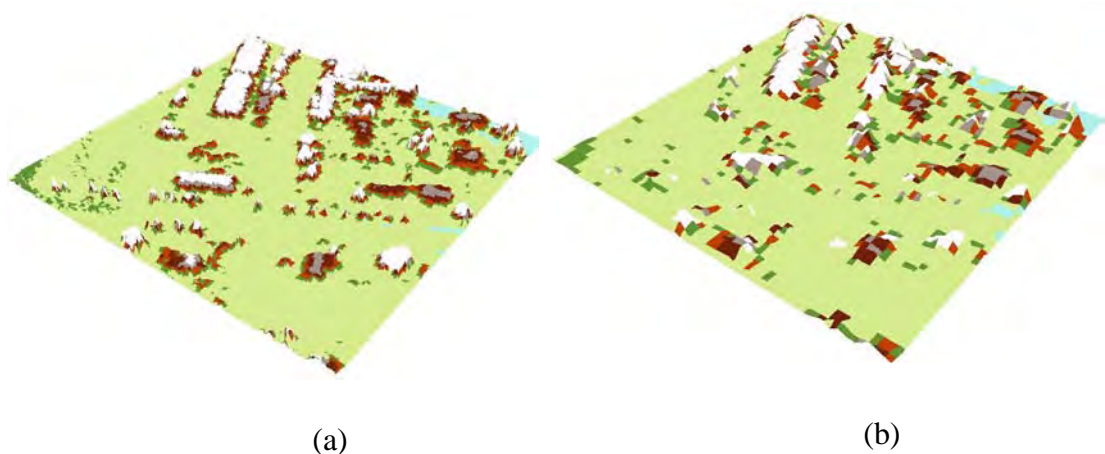


Figure 7.11 3D models of the same semi-urban area created from (a) a LiDAR-derived DSM, and (b) stereo photogrammetry

As an example, LiDAR-derived DSMs and stereo photogrammetry-derived DSMs were generated from the same semi-urban study area. Figure 7.11a represent the LiDAR-derived DSM and Figure 7.11b is the stereo photogrammetry-derived DSM. Without extensive manual editing, only a 4m spatial resolution DSM was generated from the photogrammetric technique, whereas from the LiDAR data a 1m spatial resolution DSM was generated. The latest LiDAR technology can produce a centimetre scale DSM using multi-return LiDAR data. Visual examination showed that the digital photogrammetric technique tends to smooth the surface models. As a result, landscape feature identification was much harder. LiDAR-derived DSMs generally produce superior three-dimensional models of landscape objects also.

7.3.5 Issues with comparative accuracy assessment using error matrices

Conventional accuracy assessments are based on a per-pixel approach and only consider the thematic or positional accuracy. However, feature-level fusion using OBIA produces results that have both thematic and geometric properties, therefore the traditional per-pixel error-matrix approach for accuracy assessment may be inadequate (Schiewe and Gahler, 2008). Concerns about pixel-based accuracy assessment were presented in Section 2.11. Review of comparative pixel- and object-based classification accuracy assessment (Riggan and Weih, 2009; Zhou *et al.*, 2009; Platt and Rapoza, 2008; Walter, 2004; Wang *et al.*, 2004; Geneletti and Gorte, 2003) reveal that error matrix-based cross validation was most popular as it summarizes the nature of the thematic accuracy and is also the basis for many further comparative statistical accuracy analyses. The overall accuracy derived from an error matrix only incorporates the major diagonal and excludes omission and commission errors. On the other hand, normalised accuracy includes the off-diagonal elements and represents the accuracy better (Congalton, 1991). However, normalisation can lead to a bias and have the effect of equalising the User's and Producer's accuracies where in fact they differ significantly (Stehman and Czaplewski, 1998). The user might benefit from normalization of the matrix (Smits *et al.*, 1999) but it would be impossible to work the other way and derive the original matrix from the normalized version (Foody, 2002). As a result, it is preferable to derive more than one measure of accuracy and provide an error matrix as a full description of classification accuracy (Arbia *et al.*, 1998; Muller *et al.*, 1998a; Stehman, 1997). This is important

as the use of different accuracy measures may result in different, possibly conflicting, interpretations and conclusions (Stehman, 1997).

KHAT accuracy is another error-matrix related measurement that indirectly incorporates the off-diagonal elements as products of the row and column margins. Depending on the amount of error included in the matrix, the overall, normalised and KHAT measures may not agree. In addition to these measures, Kappa is also a powerful technique for its ability to provide information about a single matrix as well as to statistically compare matrices. Kappa analysis tests the significance of each matrix alone and determines whether the results presented in the error matrix are significantly better than a random result.

The error matrix is therefore used to summarize the nature of the class allocations made by a fusion method and is the basis of many quantitative accuracy analyses. It is used to provide a site-specific assessment of the correspondence between the image fusion results and ground conditions. Given the range of viewpoints and problems surrounding the error matrix (see Section 2.11), it seems unlikely that a single universally acceptable standard for accuracy assessment and reporting can be specified. The error matrix lies at the core of much work on accuracy assessment and is frequently used without questioning its suitability (Foody, 2002).

Pixel- and feature-level fusion results were compared in order to evaluate the suitability of the two fusion techniques. The comparison was undertaken in a statistically rigorous way to provide an objective basis for comment and interpretation. Considering consistency, the same set of ground data was used for accuracy assessment for both fusion results (detail presented in Section 7.1 and 7.2). In the forest study area, feature-level fusion accuracy derived from error matrix was greater than the pixel-level fusion accuracy (86.33 percent vs 64.67 percent). The Kappa Coefficients of the pixel- and feature-level fusions (52.56 and 81.21 respectively) also indicated that feature-level fusion had *substantial* agreement compared to *moderate* agreement for pixel-level fusion. Variance of Kappa and the z test for each fusion result also indicated that thematic accuracies of the different fusion results were significantly different from random results and that feature-level fusion performed well above the pixel level fusion.

In the semi-urban study area, error-matrix based overall thematic accuracy results of pixel- and feature-level fusions were 73.25 percent and 88.38 percent respectively. The results from the fusion models are compared and the degree of interclass confusion is evaluated with a view to understanding the thematic error patterns. Pair-wise comparison of pixel- and feature-level fusion results indicated that, at the 95 percent confidence level, the standard normal deviation of the Kappa Coefficient is 3.15, well above the maximum 1.96 threshold; therefore, the feature-level fusion significantly improved the thematic accuracy as compared to the pixel-level fusion.

Cross validation techniques using unified samples indicate that error matrices play an important role in quantifying fusion accuracies and providing comparisons between different fusion models. Object-based accuracy assessment can be used with feature-level fusion but not with pixel-level fusion alone or for comparing results derived from pixel- and feature-level fusions.

7.4 Summary

This chapter had presented data- and user-derived fusion results for both of the study areas. Landscape object delineation from pixel- and feature-level fusions were presented and their comparative accuracies evaluated. This chapter also discussed different issues, which arose during the implementation phases.

For the forest study area, pixel-level fusion did not clearly delineate individual trees. The lack of homogeneity in the intensity distribution was the main reasons for the difficulty in spectrally separating the tree crowns. The advantage of feature-level fusion was that it classified the whole tree feature rather than the single pixel reducing the mixed-pixel problem and delineated trees more realistically. A thematic accuracy of 86.33 percent was achieved in feature-level fusion result outperforming the pixel-level fusion result by 22 percent. The Kappa Coefficients of the pixel- and feature-level fusions were 51.69 and 81.61 respectively, which indicated *moderate* compared to *substantial* agreement with field-surveyed data. Comparative deviation of the Kappa Coefficients indicated that significant difference was detected between pixel- and feature-level fusion results. Segmentation and subsequent feature classification substantially improved the tree species identification in feature-level fusion.

For the semi-urban study area, the results revealed that feature-level fusion exhibited higher thematic accuracy than pixel-level fusion. Pixel-level fusion of multispectral imagery with DSM data achieved 88.38 percent accuracy however, with kappa value 86.71 percent. The z statistics of the multispectral imagery with LiDAR data fusion at pixel- and feature-level were 19.95 and 3253 respectively. Comparative deviation of the Kappa Coefficient (3.12) revealed that feature-level fusion of multispectral imagery with LiDAR data was better than the pixel-level fusion results.

Pixel- and feature-level fusion models were carefully evaluated and their performance in landscape object delineation discussed. Comparative results for both the study areas revealed that feature-level fusion delineated objects better than pixel-level fusion. Data-driven feature-level fusion model was employed in the forest study area as datasets were rich with spectral, spatial, and textural information. Data-driven fusion models were found inappropriate for application of the semi-urban study area owing to the long list of landscape objects and the complex nature of the landscape classification scheme.

In the two different fusion processes, it was evident that inclusion of LiDAR-derived height data improved the correct delineation of landscape objects. In pixel-level fusion, this was particularly evident when observing different semi-urban landscape objects. Excluding DSM results gave a 'pepper and salt' effect due to serious misclassification of objects. Inclusion of LiDAR-derived DSM information reduced 'pepper and salt' effect and improved landscape delineation. Landscape object delineation using feature-level fusion of colour imagery with DSM data achieved an accuracy of 91 percent, whereas using only colour imagery achieved 66 percent accuracy.

Both aerial photography and LiDAR have been used extensively for generating DSMs of landscape objects; however, current stereo photogrammetric solutions have limitations in production high resolution DSMs without manual editing. Use of LiDAR, on the other hand, can produce very high-resolution 3D landscape models that can easily be incorporated into different fusion models.

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

The opening chapters of this thesis introduced the subject of data fusion in a *multi-source classification* context and explained how it consisted of two models: data- and user-driven. It was shown that the fusion of aerial imagery and LiDAR data is very important in landscape object extraction from remotely sensed data, deserving detailed research. This research set out to investigate the feasibility of using pixel- and feature-level fusions in a data- and user-driven context to identify a range of landscape objects. Data-driven fusion models were used for a forest study area for the presence of a handful of Eucalyptus species, and user-driven models for a semi-urban area for the presence of large number of classes with complex landscape structures.

This research reveals that a number of points need to be considered before one even begins to formulate a fusion project. The effects of scale and correct registration of multi-source data are overridingly important. The scale issue was addressed through the use of very high spatial resolution multi-source data in a hierarchical landscape classification schemes for specific sites. The complementary nature of optical imagery and LiDAR data was the main reason for deploying them in the fusion models for identifying a range of landscape objects. A parametric registration model was applied to bring optical imagery into the same geographic coordinate system as the LiDAR data. LiDAR data were processed to derive absolute height of the landscape objects. These heights were validated using field-surveyed data. Some temporal changes were expected in both study areas as the multi-source aerial data were collected over the 4-year period 2001-2004. The study areas were in a slow-growing region that had not seen many changes in that period, nevertheless some temporal effects were found due mainly to object movement - a common challenge with high spatial resolution data fusion.

8.1.1 Tree species identification in a forest area

Understanding the composition of tree species in a forest environment is an important task for managing our forest resources. Field survey is still considered the main source for collecting this information. However, the application of very high spatial resolution, remotely sensed data makes data collection a more timely and cheaper option. This study shows that data-driven fusion of aerial imagery and LiDAR data automatically identifies tree species without user intervention. Data-driven pixel-level fusion uses an unsupervised algorithm to fuse spectral, height and texture pixels for species identification. Data-driven feature-level fusion, on the other hand, employed watershed segmentation for delineating tree features first then used an unsupervised algorithm to classify trees into different species.

Pixel-level unsupervised fusion correctly separated tree and non-tree areas, however it could not delineate individual tree crowns. Misclassification of pixels was high within individual tree crown areas indicating that unsupervised pixel-level fusion is not efficient in separating tree species.

On the other hand, feature-level fusion could quite accurately identify tree species. The use of tree features (polygons) derived from watershed segmentation decisively improves fusion results through the assignment of the most frequent pixel of the layer to a particular tree polygon. Assigning a single digital number for each tree feature derived from multi-source data substantially reduces interclass confusion compared to using the original pixel values.

Thematic accuracies of data-driven fusion results as derived from error matrices achieved better overall accuracy (86 percent) for feature-level fusion compared to pixel-level fusion (65 percent). A Kappa Coefficient of 82 percent when employing feature-level fusion indicated *substantial* agreement with reference data whereas, for pixel-level fusion a value of 52 percent that indicated only *moderate* agreement with reference data. A comparison Kappa analysis statistically proved that feature-level fusion leads to a better result than pixel-level fusion. Thus feature-level fusion can be recommended over pixel-level fusion when attempting to identify the tree species.

8.1.2 Identifying objects in a semi-urban area

Using just multispectral imagery or LiDAR-derived data is not sufficient to recognise semi-urban landscape objects. The combine use of these two datasets improves object recognition. This study also reveals that data only is not sufficient to identify semi-urban objects for their complex nature and structure. The user plays a vital role in guiding the fusion process by providing inputs as knowledge-rules. Employing a supervised parallelepiped classifier for pixel-level fusion and multi-resolution segmentation for feature-level fusion greatly improves object identification.

It appears that the mixed-pixel problem, or ‘pepper and salt’ effect, is inevitable in pixel-level fusion, whereas objects appear sharper in feature-level fusion. Landscape features were well delineated in the segmentation stage of feature-level fusion and later these features were successfully classified using knowledge-driven spectral, spatial and contextual attributes.

Significant differences between pixel- and feature-level fusion results for the semi-urban study area were evident visually. Feature-level fusion significantly improved thematic quality as compared to pixel-level fusion, findings very similar to the results published by Kamagata *et al.* (2005).

Calculation of the Kappa Coefficients for results based on pixel-level fusion showed there was *substantial* agreement with reference data. On the other hand, the Kappa Coefficients for results from feature-level fusion was close to 1 indicating *excellent* agreement with reference data. Statistical comparison of pixel- and feature-level fusion results also revealed that object identification was better in feature-level fusion.

The greatest improvement in results using feature-level fusion over pixel-level fusion was in the delineation of *Shadow* classes. The User’s and Producer’s Accuracies are substantially better than for pixel-level fusion results. It is the incorporation of contextual properties into the feature-level fusion that leads to these better results.

It is interesting to note that replacement of multispectral imagery with colour imagery, did not making much difference to feature-level fusion results. However, in pixel-level fusion of multispectral imagery (16-bit) there was less misclassification

compared to fusion results derived from colour imagery (8-bit). This result reveals that radiometric depth of the imagery influences pixel-level fusion results.

Excluding LiDAR-derived height data from fusion models degrade results. In feature-level fusion, the degradation was massive and in pixel-level fusion the difference was substantial. So, it must be concluded that LiDAR-derived height data plays a huge role in differentiating objects when they have similar spectral properties.

Landscape object height can of course be derived by stereo matching aerial photographs but experiments during this study showed that height density data is not as accurate as LiDAR-derived data. Stereo matching aerial photographs cannot generate automatic high-density height data due to the shadow effect, less texture detail and the often significant geometric and radiometric differences between images. Thus, aerial photography-derived height data is not as suitable as LiDAR for use in the fusion process.

This research has provided a means for extracting landscape information for forest and semi-urban areas by fusing aerial imagery with LiDAR data. Feature-level fusion approach is useful for a wide variety of applications such as automated object extraction and for rapid mapping. Analysis of thematic accuracy detailed in this research provides an understanding of data quality issues associated with object recognition in different fusion models that should prove useful for designing future landscape mapping models.

8.2 Recommendations for Future Research

During this research some limitations of data fusion approaches were identified. The following recommendations are proposed for future work.

8.2.1 New data sources

Recent developments in satellite sensing systems have resulted in commercially available remotely sensed optical imagery with spatial resolution less than one metre. However, satellite-based high spatial resolution LiDAR has not yet been developed. LiDAR systems for landscape level mapping are still restricted to the aerial platform.

Currently the LiDAR system can generate and acquire multi-pulse returns so that it does not only record first and last returns but also intermediate returns such as second and third returns. A wide range of datasets are therefore available for fusion, each with their own enhanced spectral and spatial qualities.

Data from new sensors should be investigated to establish the information content and relevance to assessment of a wide range of landscape objects. These sensors are capable of expanding the number of spectral bands, level of detail (scale) and temporal information. Serious consideration should be given to the fusion of radar-derived high spatial resolution imagery with DEM data for automated landscape feature extraction. Constant evaluation of new data sources is needed to ensure continuity of landscape monitoring.

8.2.2 Automatic registration of multi-source data

The time lag between the collection of aerial imagery and LiDAR was not considered as a serious issue as the best available datasets were used in this research. Improved results may be achieved if the data sources do not have any time lag and phenological cycles of the landscape objects, such as tree flowering, are considered. This will presumably enable superior separation of landscape objects.

Geometric quality of the data needs to be investigated and assessments made of automatic image orthorectifications. A coherent well-rectified dataset can be produced when the optical and LiDAR sensors are on the same aerial platform and orthorectification is done automatically. Such a technique may be useful for minimising spatial errors in multi-source data and reducing the processing time for large data sets.

8.2.3 Virtual field data collection

The ability to accurately identify landscape objects from multi-source fused data is presently dependent on timely acquisition and accurate field data. However, field data collection is always time consuming and costly. Future research needs to be directed at minimising time spent collecting data in the field. An alternative is the creation of a virtual field scenario in a digital environment where the user can interact directly with the sample objects and derive measurements from them.

8.2.4 Automatic fusion models

Terrestrial photographs show the vertical structures of trees and sometimes tree species can be identified from their component facts. However, this sort of composition is unseen in remotely sensed data. For the forest study area, data-driven pixel-level unsupervised fusion results indicated that some tree species were not separable using the multi-source data even though they were identifiable in terrestrial photographs. Detailed examination of tree species' attributes needs to be undertaken prior to final design of any pixel-level fusion scheme. More sophisticated pixel-level approaches that include hierarchical clustering algorithms should be investigated in future. In a data-driven feature-level fusion approach, a high-end segmentation technique called watershed segmentation was successfully employed for feature definition. Further research should be performed using different segmentation algorithms and their accuracy for tree feature delineation compared.

For the semi-urban study area, the user-driven pixel-level fusion models gave only a moderate level of certainty in the allocation of pixels to classes when the parallelepiped decision rule was implemented. Many pixels exhibited multiple class membership due to conflict in class definition, or because of spatial complexity of landscape objects. More sophisticated pixel-level fusion approaches are needed that can handle multiple class memberships as well as use of more advanced and complex fusion schemes (such as a support vector machine) to take advantage of modern computing power.

The future of data fusion for landscape object extraction lies in the development of machine learning or artificial intelligence where a large numbers of scenes can be automatically processed and provide consistent results. This can be achieved by incorporating multiple complex fusion algorithms into a single machine process, which can deploy algorithms as necessary using artificial intelligence.

8.2.5 Machine-generated accuracy assessment

Accuracy assessment is essential for the evaluation of the quality and appropriateness of products. Consideration of the spatial and thematic components of analysis is required for a complete understanding of the accuracy of object delineation. This research evaluated the accuracy using point-based sampling and calculated an error

matrix with Kappa statistics as an indicator of accuracy of object delineation. Research should be directed in future to more sophisticated accuracy assessment, which not only assesses point-based thematic accuracy but also includes other mechanisms to assess accuracies such as positional accuracy and fuzziness of objects' boundaries.

In landscape mapping, an integrated machine-generated spatial and thematic accuracy approach is required to ensure fused data leads to proper interpretation of object classes. Accuracy issues that relate to the geo-referencing, boundary location, and thematic fusion need to be addressed in a comprehensive manner. It is clear that standard error matrices are inadequate for the task, and further research is required on more effective techniques.

REFERENCES

- Ackermann, F. (1999) Airborne laser scanning—present status and future expectations, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, p. 64–67.
- Aiazzi, B., Alparone, L., Argenti, F. and Baronti, S. (1999) Wavelet and pyramid techniques for multisensor data fusion: a performance comparison varying with scale ratios, *Proceedings of Europto, Image and signal processing for remote sensing V*, 22–24 September 1999, p. 251–262.
- Albrecht, F. (2008) "Assessing the spatial accuracy of object-based image classification", in: *Geospatial Crossroads @ GI_Forum '08*, Car, A., Griesebner, G. and Stoble, J. (eds.), Wichmann, Heidelberg, pp. 11-20.
- Amolins, K., Zhang, Y. and Dare, P. (2007) Applications of wavelet transforms in image fusion, *Proceedings of Urban Remote Sensing Joint Event*, IEEE, Paris, 11-13 April 2007, pp. 1-7.
- Anderson, J.R., Hardy, E.E. and Roach, J.T. (1976) A land use and land cover classification system for use with remote sensor data, US Government Printing Office, Washington, DCpp.
- Aplin, P., Atkinson, P.M. and Curran, P.J. (1999) "Per-field classification of land use using the forthcoming very fine spatial resolution satellite sensors: problems and potential solutions", in: *Advances in remote sensing and GIS Analysis*, Atkinson, P. M. and Tate, N. J. (eds.), John Wiley & Sons, Chichester, p. 219–239.
- Arbia, G., Griffith, D. and Haining, R. (1998) Error propagation modelling in raster GIS: overlay operations, *International Journal of Geographical Information Science*, Vol. 12, p. 145–167.
- Atkinson, P.M. and Aplin, P. (2004) Spatial variation in land cover and choice of spatial resolution for remote sensing, *International Journal of Remote Sensing*, Vol. 25, No. 18, p. 3687–3702.
- Atkinson, P.M. and Curran, P.J. (1997) Choosing an appropriate spatial resolution for remote sensing investigations, *Photogrammetric Engineering & Remote Sensing*, Vol. 63, No. 12, pp. 1345-1351.
- Atkinson, P.M., Foody, G.M., Curran, P.J. and Boyd, D.S. (2000) Assessing the ground data requirements for regional-scale remote sensing of tropical forest biophysical properties, *International Journal of Remote Sensing of Environment*, Vol. 21, p. 2571–2587.
- Axelsson, P. (1999) Processing of laser scanner data—algorithms and applications, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, p. 138–147.

- Baatz, M., Benz, U., Dehghani, S. and Heynen, M. (2004) *eCognition User Guide 4* [Online], Available: www.definiens-imaging.com (16 September, 2004).
- Baatz, M. and Schape, A. (2000) Multiresolution Segmentation-an optimization approach for high quality multi-scale image segmentation, *Proceedings of Angewandte Geogr. Informationsverarbeitung XII*, Strobl, J. and Blaschke, T. eds., Wichmann, Heidelberg, pp. 12-23.
- Baatz, M. and Schäpe, A. (1999) Object-Oriented and Multi-Scale Image Analysis in Semantic Networks, *Proceedings of 2nd International Symposium: Operationalization of Remote Sensing*, ITC, NL, 16-20 August.
- Ball, G.H. and Hall, D.J. (1965) A novel method of data analysis and pattern classification, Stanford Research Institute, Menlo Park, California, pp.
- Baltsavias, E.P. (1999) A comparison between photogrammetry and laser scanning, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, p. 83–94.
- Benson, B.J. and Mackenzie, M.D. (1995) Effects of Sensor Spatial Resolution on Landscape Structure Parameters, *Landscape Ecology*, Vol. 10, No. 2, pp. 113-120.
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M. (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 58, pp. 239-258.
- Bethune, S.D., Muller, F. and Donnay, J.P. (1998) Fusion of multispectral and pancromatic images by local mean and variance matching filtering techniques, *Proceedings of Fusion of Earth Data*, Sophia Antipolis, 28-30 January 1984, pp. 31-36.
- Beucher, S. (1991) The watershed transformation applied to image segmentation, *Proceedings of 10th Pfeifferkorn Conf. on Signal and Image Processing in Microscopy and Microanalysis*, Scanning Microscopy International, Cambridge, UK, 16-19 sept., pp. 299-314.
- Beucher, S. and Lantuejoul, C. (1979) Use of watersheds in contour detection, *Proceedings of International Workshop on Image Processing, Real-time Edge and Motion Detection/Estimation*, Rennes, France.
- Biging, G.S., Colby, D.R. and Congalton, R.G. (1999) "Sampling systems for change detection accuracy assessment", in: *Remote sensing change detection: environmental monitoring methods and applications*, Lunetta, R. S. and Elvidge, C. D. (eds.), Taylor and Francis, London, p. 281–308.
- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) *Discrete multivariate analysis*, MIT Press, Cambridge, MA. 557 pp.

- Blaschke, T. (2010) Object based image analysis for remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65, pp. 2-16.
- Blaschke, T., Conradi, M. and Lang, S. (2001) Multi-scale image analysis for ecological monitoring of heterogeneous, small structured landscapes, *Proceedings of SPIE*, Toulouse, pp. 35-44.
- Blaschke, T. and Strobl, J. (2001) What's wrong with pixels? Some recent developments interfacing remote sensing and GIS, *Proceedings of GIS – Zeitschrift für Geoinformationssysteme 6/2001*, Hüthig GmbH & Co. KG, Heidelberg, pp. 12-17.
- Bo, Y.C., Wang, J.F. and Li, X. (2005) Exploring the Scale Effect in Land Cover Mapping from Remotely Sensed Data: the Statistical Separability-based Method, *IGARSS*, pp. 3884-3887.
- BOMa (2006) *Bureau of Meteorology, Climate Averages for Tocumwal Airport* [Online], Available: http://www.bom.gov.au/climate/averages/tables/cw_074106.shtml (9th November).
- BOMb (2006) *Bureau of Meteorology, Climate Averages for Deniliquin (Wilkinson St)* [Online], Available: http://www.bom.gov.au/climate/averages/tables/cw_074128.shtml (9th November).
- BOMc (2006) *Bureau of Meteorology, Climate Averages for Echuca Aerodrome* [Online] (9th November).
- Bork, E.W. and Su, J.G. (2007) Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis, *Remote Sensing of Environment*, Vol. 111, p. 11–24.
- Botkin, D.B., Estes, J.E., MacDonald, R.M. and Wilson, M.V. (1984) Studying the Earth's Vegetation from Space, *Bioscience*, Vol. 34, pp. 508-514.
- Brandtberg, T., Warner, T.A., Landenberger, R.E. and McGraw, J.B. (2003) Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America, *Remote Sensing of Environment*, Vol. 85, p. 290–303.
- Bremaud, P. (1999) *Markov chains Gibbs field, Monte Carlo simulation, and queues*, Springer-Verlag, New York. 186 pp.
- Brown, L.G. (1992) A survey of image registration techniques, *ACM Computing Surveys*, Vol. 24, p. 326–376.
- Bruzzone, L. and Prieto, D.F. (2000) Automatic analysis of the difference image for unsupervised change detection, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 3, p. 1171– 1182.

- Buiten, H.J. and van Putten, B. (1997) Quality Assessment of Remote Sensing Image Registration - Analysis and Testing of Control Point Residuals, *Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 57-73.
- Burrough, P.A. (1986) *Principles of Geographical Information Systems for Land Resources Assessment*, Clarendon Press, Oxford. 194 pp.
- Burrough, P.A. and McDonnel, R.A. (1998) *Principles of Geographical Information Systems*, Oxford University Press, New York, USA. 333 pp.
- Canny, J.F. (1986) A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 337-356.
- Canter, F. (1997) Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification, *Photogrammetric Engineering and Remote Sensing*, Vol. 63, p. 403-414.
- Cao, C. and Lam, N.S. (1997) "Understanding the Scale and Resolution Effects in Remote sensing and GIS", in: *Scale in Remote Sensing and GIS*, Quattrochi, D. A. and Goodchild, M. F. (eds.), Lewis Publishers, Boca Raton, Fla., pp. 57-72.
- Carleer, A.P., Debeir, O. and Wolff, E. (2005) Assessment of very high spatial resolution satellite image segmentations, *Photogrammetric Engineering and Remote Sensing*, Vol. 71, pp. 1285-1294.
- Carper, W.J., Lillesand, T.M. and Kieffer, R.W. (1990) The use of Intensity-Hue-Saturation transformations for merging SPOT Panchromatic and multispectral image data, *Photogrammetric Engineering & Remote Sensing*, Vol. 56, pp. 459-467.
- Chang, Y.L. and Li, X. (1994) Adaptive image region-growing, *IEEE Image Process*, Vol. 3, pp. 868-872.
- Chavez, P.S. and Bowell, A.J. (1988) Comparison of the spectral information content of Landsat TM and SPOT for three different sites in the Phoenix, Arizona Region, *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 12, pp. 1699-1708.
- Chavez, P.S. and Kwarteng, A.Y. (1989) Extracting spectral contrast in Landsat Thematic Mapper image data using selective Principal Component Analysis, *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 3, pp. 339-348.
- Chavez, P.S., Sides, J.S.C. and Anderson, J.A. (1991) Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT Panchromatic, *Photogrammetric Engineering and Remote Sensing*, Vol. 57, No. 3, pp. 295-303.

- Chen, Q., Balddocchi, D., Gong, P. and Maggi, K. (2006) Isolating individual trees in a Savanna Woodland using small footprint lidar data, *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No. 8, pp. 923-932.
- Chen, Y., Fung, T., Lin, W. and Wang, J. (2005) An Image Fusion Method Based on Object-Oriented Image Classification, *Proceedings of IGARSS 2005 Symposium*, Seoul, Korea, July 25-29, 2005.
- Cherchali, S., Amram, O. and Flouzat, G. (2000) Retrieval of temporal profiles of reflectance from simulated and real NOAA AVHRR data over heterogeneous landscapes, *International Journal of Remote Sensing*, Vol. 21, No. 4, pp. 753-775.
- Chong, J. (2003) Analysis of management of unseasonal surplus flow in the Barmah-Millewa Forest, Cooperative research centre for Catchment Hydrology, 15 pp.
- Chow, C.K. and Kaneko, T. (1972) Automatic boundary detection of the left-ventricle from cineangiograms, *Comput. Biomed. Res.*, Vol. 5, pp. 388-410.
- Clark, D.B., Castro, C.S., Alvarado, L.D.A. and Read, J.M. (2004) Quantifying mortality of tropical rain forest trees using high-spatial-resolution satellite data, *Ecological Letters*, Vol. 7, pp. 52-59.
- Cohen, J. (1960) A coefficient of agreement for nominal scales, *Edu. Psychol. Meas.*, Vol. 20, pp. 37-46.
- Congalton, R.G. (1991) A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, Vol. 37, No. 1, pp. 35-46.
- Congalton, R.G. and Green, J. (1999) *Assessing The Accuracy of Remotely Sensed Data: Principles and Practice*, CRC press, USApp.
- Congalton, R.G. and Green, K. (1993) A practical look at the sources of confusion in error matrix generation, *Photogrammetric Engineering and Remote Sensing*, Vol. 59, p. 641– 644.
- Congalton, R.G., Oderwald, R.G. and Mead, R.A. (1983) Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques, *Photogrammetric Engineering & Remote Sensing*, Vol. 49, No. 12, pp. 1671-1678.
- Cowen, D.J., Jensen, J.R. and Bresnahan, P.J. (1995) The design and implementation of an integrated geographic information system for environmental applications, *Photogrammetric Engineering and Remote Sensing*, Vol. 61, pp. 1393-1404.

- Crosetto, M. and Tarantola, S. (2001) Uncertainty and sensitivity analysis: tools for GIS-based model implementation, *International Journal of Geographical Information Science*, Vol. 15, No. 5, pp. 415-437.
- Dai, X. and Khorram, S. (1998) The effects of image misregistration on the accuracy of remotely sensed change detection, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 36, No. 5, pp. 1566-1577.
- Dare, P. (2005) Shadow analysis in high-resolution satellite imagery of urban areas, *Photogrammetric Engineering & Remote Sensing*, Vol. 71, No. 2, pp. 169-177.
- Darwish, A., Leukert, K. and Reinhardt, W. (2003) Image segmentation for the purpose of object-based classification, *IEEE International, Geoscience and Remote Sensing Symposium, IGARSS*, Vol. 3, pp. 2039-2041.
- Davis, F.W. and Simonett, D.S. (1991) "GIS and Remote Sensing", in: *Geographical Information Systems: Principles and Applications*, Vol. 14, Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (eds.), Longman Scientific and Technical, Harlow, UK, pp. 191-213.
- Davis, L.S. (1975) A survey of edge detection techniques, *Comput. Graphics Image Process.*, Vol. 4, pp. 248-270.
- DCE (1992) Barmah Management Plan for Barmah State Park and Barmah State Forest, Department of Conservation and Environment (DCE), East Melbournepp.
- deKok, R., Schneider, T. and Ammer, U. (1999) Object based classification, applications in the alpine forest environment, *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, No. 7-4-3 W6, 3-4 June, 1999.
- de-Kok, R., Schneider, T. and Ammer, U. (1999) Object based classification, applications in the alpine forest environment, *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, No. 7-4-3 W6, 3-4 June, 1999.
- Deravi, F. and Pal, S.K. (1983) Gray level thresholding using second-order statistics, *Pattern Recognition Letter*, Vol. 1, No. 417-422.
- Dikshit, O. and Roy, D.P. (1996) An empirical evaluation of image resampling effects upon the spectral and textural supervised classification of a high spatial resolution multispectral image, *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 9, pp. 1085-1092.
- Dowman, I. and Dare, P. (1999) Automated procedures for multi sensor registration and orthorectification of satellite images, *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, No. 7-4-3 W6, pp. 37-44.

- Dragut, D.L. and Blaschke, T. (2008) "Terrain Segmentation and Classification using SRTM data", in: *Advances in Digital Terrain Analysis*, Qiming, Z., Lees, G. B. and Tang, G. (eds.), Springer, Berlin, pp. 141-158.
- Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F. and Prince, S.D. (2002) Estimation of tropical forest structural characteristics using large-footprint lidar, *Remote Sensing of Environment*, Vol. 79, pp. 305-319.
- Ehlers, M. (1991) Multisensor image fusion techniques in remote sensing, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 46, pp. 19-30.
- Ehlers, M. (1997) "Rectification and Registration", in: *Integration of Geographic Information Systems and Remote Sensing*, Star, J. E., Estes, J. E. and McGwire, K. C. (eds.), Cambridge University Press, Cambridge, UK, pp. 13-36.
- ERDAS (2002) *ERDAS Field Guid*, Leica Geosystems, GIS & Mapping Division, Atlanta, USA. 226 pp.
- Ewart, A.J. (1931) *Flora of Victoria*, Melbourne Uni. Press, Melbournepp.
- Flanders, D., Hall-Beyer, M. and Pereverzoff, J. (2003) Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction, *Canadian Journal of Remote Sensing*, Vol. 29, No. 4, pp. 441-452.
- Flouzat, G., Amram, O., Laporterie, F. and Cherchali, S. (2001) Multiresolution analysis and reconstruction by a morphological pyramid in the remote sensing of terrestrial surfaces, *Signal Processing*, Vol. 81, No. 10, p. 2171–2185.
- Fonseca, L.M.G. and Manjunath, B.S. (1996) Registration Techniques for Multisensor Remotely Sensed Imagery, *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 9, pp. 1049-1056.
- Foody, G.M. (1996) Approaches for the production and evaluation of fuzzy land cover classification from remotely -sensed data, *International Journal of Remote Sensing*, Vol. 17, No. 7, pp. 1317-340.
- Foody, G.M. (2002) Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, Vol. 80, pp. 185-201.
- Foody, G.M., Campbell, N.A., Trood, N. and Wood, T.F. (1992) Derivation and applications of probabilistic measures of class membership for the maximum-likelihood classification, *Photogrammetric Engineering & Remote Sensing*, Vol. 59, No. 9, pp. 1335-1341.

- Forbes, A.D. (1995) Classification-algorithm evaluation: five performance measures based on confusion matrices, *Journal of Clinical Monitoring*, Vol. 11, p. 189–206.
- Forester, B.C. and Trinder, J.C. (1984) An examination of the effects of resampling on classification accuracy, *Proceedings of Landsat 84 - Third Australasian Conference on Remote Sensing*, 106-115, Brisbane, Old.
- Franklin, J., S. R. Phinn, Woodcock, C.E. and Rogan, J. (2003) "Rationale and conceptual framework for classification approaches to assess forest resources and properties", in: *Methods and applications for remote sensing of forests: Concepts and case studies*, Wulder, M. and Franklin, S. E. (eds.), Kluwer, New York, p. 279–300.
- Franklin, S.E. and Blodgett, C.F. (1993) An example of satellite multisensor data fusion, *Computers and Geoscience*, Vol. 19, pp. 577-583.
- Fu, K.S. and Mui, J.K. (1981) A survey on image segmentation, *Pattern Recognition*, Vol. 13, pp. 3-16.
- Fukada, Y. (1980) Spatial clustering procedures for region analysis, *Pattern Recognition*, Vol. 12, pp. 395-403.
- Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, p. 721–741.
- Geneletti, D. and Gorte, B.G.H. (2003) A method for object-oriented land cover classification combining Landsat TM data and aerial photographs, *International Journal of Remote Sensing*, Vol. 24, No. 6, pp. 1273-1286.
- Gillespie, A.R., Kahle, A.B. and Walker, R.E. (1986) Colour enhancement of highly correlated images: I. Decorrelation and HSI contrast stretches, *Remote Sensing of Environment*, Vol. 29, No. 20, pp. 209-235.
- Gong, P., Biging, G.S., Lee, S.M., Mei, X., Sheng, Y., Pu, R., Xu, B., Schwarz, K. and Mostafa, M. (1999) Photo ecometrics for forest inventory, *Geographic Information Science*, Vol. 5, pp. 9-14.
- Gong, P. and Howarth, P.J. (1990) An Assessment of Some Factors Influencing Multispectral Land-cover Classification, *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 5, pp. 597-603.
- Gonzalez, R.C. and Woods, R.E. (2002) *Digital Image Processing, 2nd edition*, Prentice Hall, NJpp.
- Gonzalez, R.C., Woods, R.E. and Eddins, S.L. (2004) *Digital image processing using MATLAB* (2nd Edition), Prentice Hall, Upper Saddle River, N. Jpp.

- Haala, N. and Brenner, C. (1999) Extraction of buildings and trees in urban environments, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, p. 130–137.
- Haddon, J.F. (1988) Generalized threshold selection for edge detection, *Pattern Recognition*, Vol. 21, pp. 195-203.
- Hall, D.L. and McMullen, S.A.H. (1997) *Mathematical techniques in multi-sensor data fusion* (2nd Edition), Artech House, Boston, pp.
- Hall, E.L. (1979) *Computer Image Processing and Recognition*, Academic Press, New York pp.
- Han, P., Gong, J.Y., Li, Z.L. and Cheng, L. (2008) The study on the choice of optimal scale in image classification, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII, No. B2, pp. 385-390.
- Haralick, R.M. (1983) Decision making in context, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 4, pp. 417-428.
- Haralick, R.M. and Shapiro, L.G. (1985) Survey image segmentation techniques, *Computer Vision Graphics and Image Processing*, Vol. 29, pp. 100-132.
- Harris, J.R. and Murray, R. (1989) IHS transform for the integration of radar imagery with geophysical data, *Proceedings of The 12th Canadian Symposium on Remote Sensing*, CSRS, Vancouver, Canada, pp. 923-926.
- Harris, J.R., Murray, R. and Hirose, T. (1990) IHS transform for the integration of radar imagery with other remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, Vol. 56, pp. 1631-1641.
- Harrison, B.A. and Jupp, D.L.B. (1990) *Introduction to image processing: MicroBRIAN Resource Manual, Part 2*, CSIRO Publications, Melbourne, 156 pp.
- Hay, G.J. and Castilla, G. (2008) "Geospatial object-based image analysis (GEOBIA): A new name for a new discipline", in: *Object Based Image Analysis (GEOBIA)*, Blaschke, T., Lang, S. and Hay, G. J. (eds.), Springer-Verlag, Berlin, pp. 75-90.
- Helt, M., Olsena, R.C. and Puetzb, A.M. (2006) Analysis of LIDAR and spectral imagery for scene classification in forested areas, *Proceedings of SPIE*, Shen, S. S. and Lewis, P. E. eds., pp. 62331P1-9.
- Herold, M., Schiefer, S., Hostert, P. and Roberts, D.A. (2006) "Applying image spectrometry in urban areas", in: *Urban Remote Sensing*, Qihao, W. and Quattrochi, D. A. (eds.), CRC Press.

- Hofmann, P. (2001) Detecting buildings and roads from IKONOS data using additional elevation information, *Proceedings of GIS – Zeitschrift für Geoinformationssysteme 6/2001*, Hüthig GmbH & Co. KG, Heidelberg, pp. 28-33.
- Holliday, I. (1969) *A Field Guide to Australian Trees*, Rigby Publishingpp.
- Holliday, I. and Hill, R.A. (1974) *Field Guide to Australian Trees*, Frederick Muller Ltd.pp.
- Hord, R.M. (1982) *Digital Image Processing of Remotely Sensed Data*, Academic Press, New Yorkpp.
- Hsiaofei, C., Yanglin, W., Zhengguo, L. and Ichen, H. (2006) Zoning by functions of small-scale forest ecosystems: a case study of Hui-Sun Forest Station in Taiwan Province, China, *Frontiers of Forestry in China*, Vol. 1, No. 1, pp. 21-27.
- Hu, Y.H., Lee, H.B. and Scarpace, F.L. (2001) Optimal Linear spectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 7, pp. 1533-1536.
- Husak, G.J., Hadley, B.C. and McGwire, K.C. (1999) Landsat Thematic Mapper registration accuracy and its effects on the IGBP validation, *Photogrammetric Engineering and Remote Sensing*, Vol. 65, p. 1033–1039.
- Hyypä, J., Hyypä, H., Inkinen, M., Engdahl, M., Linko, S. and Zhu, Y. (2000) Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes, *Forest Ecology and Management*, Vol. 128, No. 1-2, pp. 109-120.
- Hyypä, J., Kelle, O., Lehtikoinen, M. and Inkinen, M. (2001) A Segmentation-Based Method to Retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 5, pp. 969-975.
- ITT (2011)[Online], Available: <http://www.ittvis.com/language/en-us/productsservices/envi/envifeatureextractionmodule.aspx>.
- Jäger, G. and Benz, U. (2000) Measures of Classification Accuracy Based on Fuzzy Similarity, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 3, pp. 1462-1467.
- Janssen, L.L.F. and Molenaar, M. (1995) Terrain objects, their dynamics and their monitoring by the integration of GIS and remote sensing, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 33, pp. 749-758.
- Janssen, L.L.F. and van-der Wel, F.J.M. (1994) Accuracy assessment of satellite derived land-cover data: a review, *Photogrammetric Engineering & Remote Sensing*, Vol. 60, No. 4, pp. 419-426.

- Jensen, J.R. (1996) *Introductory Digital image Processing A Remote Sensing Perspective* (2nd Edition), Prentice Hall. 205-277 pp.
- Johansen, K., Bartolo, R. and Phinn, S.R. (2010) Special feature- Geographic Object-Based Image Analysis, *Journal of Spatial Science*, Vol. 55, No. 1, pp. 3-7.
- Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J. and Baret, F. (2000) Developments in the 'validation' of satellite sensor products for the study of the land surface, *International Journal of Remote Sensing*, Vol. 21, p. 3383–3390.
- Kamagata, N., Akamatsu, Y., Mori, M., Li, Y.Q., Hoshino, Y. and Hara, K. (2005) Comparison of pixel-based and object-based classifications of high resolution satellite data in urban fringe areas, *Proceedings of the 26th Asian Conference on Remote Sensing*, Hanoi, Vietnam, 7 - 11 November.
- Kardoulas, N.G., Bird, A.C. and Lawan, A.I. (1996) Geometric Correction of SPOT and Landsat Imagery: A Comparison of Map and GPS-Derived Control Points, *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 10, pp. 1173-1177.
- Kartikeyan, B., Sarkar, A. and Majumder, K.L. (1998) A segmentation approach to classification of remote sensing imagery, *International Journal of Remote Sensing*, Vol. 19, No. 9, pp. 1695-1709.
- Kasetkasema, T., Arorab, M.K. and Varshney, P.K. (2005) Super-resolution land cover mapping using a Markov random field based approach, *Remote Sensing of Environment*, Vol. 96, pp. 302-314.
- Kato, A., Moskalb, L.M., Schiessb, P., Calhounc, D. and Swansond, M.E. (2010) True orthophoto creation through fusion of LiDAR-derived digital surface model and aerial photos, *Proceedings of ISPRS TC VII Symposium – 100 Years ISPRS*, W., W. and Székely, B. eds., IAPRS, Vienna, Austria, July 5–7, 2010, pp. 88-93.
- Kelly, M., Shaari, D., Guo, Q.H. and Liu, D.S. (2004) A comparison of standard and hybrid classifier methods for mapping hardwood mortality in areas affected by "sudden oak death", *Photogrammetric Engineering & Remote Sensing*, Vol. 70, pp. 122-1239.
- Kennedy, M. (2002) *The Global Positioning System and GIS*, Taylor and Francis Inc, New York. 336 pp.
- Kettig, R. and Landgrebe, D. (1976) Classification of multispectral image data by extraction and classification of homogenous objects, *IEEE Transactions on Geoscience Electronics*, Vol. GE-14, No. 14, pp. 29-44.
- Kettig, R.L. and Landgrebe, D.A. (1975) Computer classification of multispectral image data by extraction and classification of homogeneous objects, *LARS*

Information Note 050975, The Laboratory for Application of Remote Sensing, Purdue University, West Lafayette, Indiana.

- Keys, L.D., Schmidt, N.J. and Phillips, B.E. (1990) A prototype example of sensor fusion used for a siting analysis, *Technical Papers 1990, ACSM-ASPRS Annual Convention, Image Processing and Remote Sensing*, Vol. 4, pp. 238-249.
- Klinger, A. (1973) Data structures and pattern recognition, *Proceedings of First International Joint Conference on Pattern Recognition*, Washington D. C., October, pp. 497-498.
- Koukoulas, S. and Blackburn, G.A. (2005) Mapping individual tree location, height and species in broadleaved deciduous forest using airborne Lidar and multispectral remotely sensed data, *International Journal of Remote Sensing*, Vol. 26, No. 3, pp. 431-455.
- Kundu, M.K. and Pal, S.K. (1986) Thresholding of edge detection using human psychovisual phenomena, *Pattern Recognition Letter*, Vol. 4, pp. 433-441.
- Labovits, M.L. and Marvin, J.W. (1986) Precision in geodetic correction of TM data as a function of the number, spatial distribution and success in matching control points: A simulation, *Remote Sensing of Environment*, Vol. 20, pp. 237-252.
- Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data, *Biometrics*, Vol. 33, pp. 159-174.
- Lang, S. (2008) "Object-based image analysis for remote sensing applications: Modeling reality - Dealing with complexity", in: *Object Based Image Analysis*, Blaschke, T., Lang, S. and Hay, G. J. (eds.), Springer, Heidelberg, Berlin, New York, pp. 1-25.
- Lang, S., Albrecht, F., Kienberger, S. and Tiede, D. (2010) Object validity for operational tasks in a policy context, *Journal of Spatial Science*, Vol. 55, No. 1, pp. 9-22.
- Laporterie, F. and Flouzat, G. (2003) The morphological pyramid concept as a tool for multi-resolution data fusion in remote sensing, *Integrated Computer-Aided Engineering*, Vol. 10, No. 1, p. 63-79.
- Leberl, F. and Gruber, M. (2005) Ultracam-d: understanding some noteworthy capabilities, *Proceedings of Photogrammetric Week 2005*, Stuttgart, Germany.
- Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L. and Shreenan, R. (2003) Combined high density LiDAR and multispectral imagery for individual tree crown analysis, *Canadian Journal of Remote Sensing*, Vol. 29, No. 5, p. 633-649.

- Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G. and Shugart, H.H. (1999) Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA, *Remote Sensing of Environment*, Vol. 67, pp. 83-98.
- Li, H., Manjunath, B.S. and Mitra, S.K. (1995) Multisensor Image Fusion Using the Wavelet Transform, *Graphical Models and Image Processing*, Vol. 57, No. 3, pp. 235-245.
- Li, S.Z. (2009) *Markov Random Field Modeling in Image Analysis*, Springer. 21 pp.
- Lillesand, T.M. and Kiefer, R.W. (1994) *Remote sensing and image interpretation* (3rd Edition), John Wiley & Sons, Inc., New York, USA. 679-685 pp.
- Lindenberger, J. (1973) Guide to multispectral data analysis using LARSYS, Purdue University, West Lafayette, Indiana, 88 pp.
- Liow, Y. and Pavlidis, T. (1990) Use of shadows for extracting buildings in aerial images, *Computer Vision, Graphics, and Image Process*, Vol. 49, pp. 242-277.
- Lo, C.P. and Watson, L.J. (1998) The influence of geographic sampling methods on vegetation map accuracy evaluation in a swampy environment, *Photogrammetric Engineering & Remote Sensing*, Vol. 64, No. 12, pp. 1189-1200.
- Lu, Y.H. and Trinder, J. (2003) Data fusion applied to automatic building extraction in 3D reconstruction, *Proceedings of Annual ASPRS Conference*, Anchorage, Alaska, May 2003, p. 114–122.
- Lucieer, A. and Stein, A. (2002) Existential uncertainty of spatial objects segmented from satellite sensor imagery, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 11, pp. 2518-2521.
- Maas, H.-G. (1999) Fast determination of parametric house models from dense airborne laser scanner data, *IAPRS XXXII*, Vol. 2W1, p. 1–6.
- Maling, D.H. (1988) "The concept of the accurate map", in: *Measurements from Maps: Principles and Methods of Cartometry*, Pergamon Press, Oxford, UK.
- Marceau, D.J., Howarth, P.J. and Gratton, D.J. (1994) Remote sensing and the measurement of geographical entities in a forest environment. 1. The scale and spatial aggregation problem, *Remote Sensing of Environment*, Vol. 49, pp. 93-104.
- Marpu, P.R., Neubert, M., Herold, H. and Niemeyer, I. (2010) Enhanced evaluation of image segmentation results, *Journal of Spatial Science*, Vol. 55, No. 1, pp. 55-68.

- Martin, L.R.G. (1989) Accuracy Assessment of Landsat-based Visual Change Detection Methods Applied to the Rural-urban Fringe, *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 2, pp. 209-215.
- McCoy, M.R. (2004) *Field methods in remote sensing*, The Guilford Press pp.
- McGwire, K.C. and Goodchild, M.F. (1997) "Accuracy", in: *Integration of Geographic Information Systems and Remote Sensing*, Strar, J. L., Estes, J. E. and McGwire, K. C. (eds.), Cambridge University Press, Cambridge, UK, pp. 110-133.
- McKeown, D. (1988) Building knowledge-based systems for detecting man-made structure from remotely sensed imagery, *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, Vol. 324, No. 1579, p. 423–35.
- MDBC (2006) *Murray-Darling Basin Initiative, Fact Sheet 6: The Barmah Choke* [Online], Available: http://www.mdbc.gov.au/rmw/river_murray_system/dartmouth_reservoir/hume_and_dartmouth_dams_operations_review/backgrounders/6_the_barmah_choke (9th November).
- Meyer, F. and Beucher, S. (1990) Morphological segmentation, *Journal of Visual Communication and Image Representation*, Vol. 1, No. 1, pp. 21-46.
- Meyer, P., Staenz, K. and Itten, K.I. (1996) Semi-automated procedures for tree species identification in high spatial resolution data from digitized colour infrared-aerial photography, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 51, pp. 5-16.
- Mezneda, N., Abdeljaoueda, S. and Boussemab, M.R. (2010) A comparative study for unmixing based Landsat ETM+ and ASTER image fusion, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 12, No. 1, pp. 131-137.
- Mikhail, E.M., Bethel, J.S. and McGlone, C.J. (2001) *Introduction to Modern Photogrammetry*, John Wiley & Sons. 118 pp.
- Minghell-Roman, A., Mangolini, M., Petit, M. and Polidori, L. (2001) Spatial resolution improvement of MeRIS images by fusion with TM images, *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 39, No. 7, pp. 1533-1536.
- Moller, M., Lymburner, L. and Volk, M. (2007) The comparison index: A tool for assessing the accuracy of image segmentation, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 9, p. 311–321.
- Muell, F. (2007) *Eucalyptus largiflorens-Black box* [Online], Available: http://gardenbed.com/source/27/2653_cul.asp (8th January 2007).

- Muhammad, S., Wachowicz, M. and de-Carvalho, L.M.T. (2002) Evaluation of wavelet transform algorithms for multiresolution image fusion, *Proceedings of The 5th International conference on information fusion*, International Society of Information Fusion, pp. 1573-1780.
- Muller, S.V., Walker, D.A., Nelson, F.E., Auerach, N.A., Bockheim, J.G., Guyer, S. and Sherba, D. (1998a) Accuracy assessment of a land-cover map of the Kuparuk river basin, Alaska: considerations for remote regions, *Photogrammetric Engineering and Remote Sensing*, Vol. 64, p. 619– 628.
- Muller, S.V., Walker, D.A., Nelson, F.E., Auerbach, N.A., Bockheim, J.G., Guyer, S. and Sherba, D. (1998b) Accuracy assessment of a land cover map of the Kuparuk River Basin, Alaska: considerations of remote regions, *Photogrammetric Engineering & Remote Sensing*, Vol. 64, No. 6, pp. 619-628.
- Mutlu, M., Popescu, S.C., Stripling, C. and Spencer, T. (2008) Mapping surface fuel models using lidar and multispectral data fusion for fire behavior, *Remote Sensing of Environment*, Vol. 112, p. 274–285.
- Næsset, E. (1997) Determination of mean tree height of forest stands using airborne laser scanner data, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 49-56.
- Nakagawa, Y. and Rosenfeld, A. (1979) Some experiments on variable thresholding, *Pattern Recognition*, Vol. 11, pp. 191-204.
- Nandhakumar, N. (1990) A phenomenological approach to multisource data integration: analyzing infrared and visible data, *Proceedings of Multisource Data Integration in Remote Sensing*, NASA Conference Publication 3099, Maryland, USA, 14-15 June 1990, pp. 61-73.
- Nishii, R. (2003) A Markov random field-based approach to decision-level fusion for remote sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 10, p. 2316– 2319.
- Pajares, G. and Cruz, J.M. (2004) A wavelet-based image fusion tutorial, *Pattern Recognition*, Vol. 37, p. 1855 –1872.
- Pal, N.R. and Pal, S.K. (1993) A review on image segmentation techniques, *Pattern Recognition*, Vol. 26, No. 9, pp. 1277-1294.
- Pala, V. and Pons, X. (1995) Incorporation of relief in polynomial-based geometric corrections, *Photogrammetric Engineering & Remote Sensing*, Vol. 61, No. 7, pp. 935-944.
- Paradella, W.R., Bignelli, P.A., Veneziani, P., Pietsch, R.W. and Tourin, T. (1997) Airborne and spaceborne synthetic aperture radar SAR integration with Landsat TM and gamma ray spectrometry for geological mapping in a

tropical rainforest environment, the Carajas Mineral Province, Brazil, *International Journal of Remote Sensing*, Vol. 18, No. 7, pp. 1483-1501.

Pavel, M. and Sharma, R.K. (1996) Fusion of Radar images rectification without the flat Earth assumption, *Proceedings of SPIE*, May 1996, p. 108–118.

Phen-Lan, L. and Po-Ying, H. (2008) Fusion methods based on dynamic-segmented morphological wavelet or cut and paste for multifocus images, *Signal Processing*, Vol. 88, No. 6, pp. 1511-1527.

Plantnet (2009a) *Eucalyptus camaldulensis* Dehnh [Online], Available: <http://plantnet.rbg Syd.nsw.gov.au/cgi-bin/NSWfl.pl?page=nswfl&lvl=sp&name=Eucalyptus~camaldulensis> (31 July 2009).

Plantnet (2009b) *Eucalyptus microcarpa* Maiden [Online], Available: <http://plantnet.rbg Syd.nsw.gov.au/cgi-bin/NSWfl.pl?page=nswfl&lvl=sp&name=Eucalyptus~microcarpa> (31 July 2009).

Platt, R.V. and Rapoza, L. (2008) An evaluation of an object-oriented paradigm for land use/land cover classification, *The Professional Geographer*, Vol. 60, No. 1, pp. 87-100.

Pohl, C. and van-Genderen, J.L. (1998) Multisensor image fusion in remote sensing: concepts, methods and applications, *International Journal of Remote Sensing*, Vol. 19, No. 5, pp. 823-854.

Popescu, S.C., Wynne, R.H. and Nelson, R.F. (2002) Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size, *Computers and Electronics in Agriculture*, Vol. 37, No. 1-3, pp. 71-95.

Price, J. (1987) Combining panchromatic data and multispectral imagery from dual resolution satellite instruments, *Remote sensing of environment*, Vol. 21, p. 119–128.

Qiming, Z., Lees, G.B. and Tang, G. (2008) "Advances in Digital Terrain Analysis: The TADTM Initiative", in: *Advances in Digital Terrain Analysis*, Qiming, Z., Lees, G. B. and Tang, G. (eds.), Springer, Berlin, pp. 3-10.

Quattrochi, D.A. and Goodchild, M.F. (1997) *Scaling in Remote sensing and GIS*, CRC/Lewis Publisher, Inc., Boca Raton, FL. 102-114 pp.

Quegan, S., A. Rye, Hendry, A., Skingley, J. and Oddy, C. (1988) Automatic interpretation strategies for synthetic aperture radar images, *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, Vol. 324, No. 1579, pp. 409-20.

- Raffy, M. (1994) Change of scale theory: a capital challenge for space observation of Earth, *International Journal of Remote Sensing*, Vol. 15, No. 2, pp. 2353-2357.
- Ranchin, T. and Wald, L. (2000) Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation, *Photogrammetric Engineering and Remote Sensing*, Vol. 66, No. 1, p. 49–61.
- Rego, F. and Koch, B. (2003) Automatic classification of land cover with high resolution data of the Rio de Janeiro city Brazil comparison between pixel and object classification. In: *Proceedings of The International archives of the photogrammetry, remote sensing and spatial information sciences*, Carstens, J. ed., Regensburg, Germany, 27-29 June.
- Richards, J.A. (1993) *Remote Sensing Digital Image Analysis - An Introduction* (2nd Edition), Springer-Verlag, Berlin,. 340 pp.
- Richards, J.A. and Jia, X. (2005) *Remote Sensing Digital Image Analysis - An Introduction* (4th Edition), Springer-Verlag, Berlin,. 340 pp.
- Riggan, N.D. and Weih, R.C. (2009) A Comparison of Pixel-based versus Object-based Land Use/Land Cover Classification Methodologies, *Journal of the Arkansas Academy of Science*, Vol. 63, pp. 145-152.
- Robertson, T.V. (1973) Extraction and classification of objects in multispectral images, *IEEE 73 CHO 837-2GE*, Purdue University, West Lafayette, Indiana, 3B-27-3B-34 pp.
- Robinson, V.B. (1988) Some implications of Fuzzy Set theory applied to Geographic Database, *Computers, Environment and Urban System*, Vol. 12, No. 1, pp. 89-97.
- Rogers, R.H. and Wood, L. (1990) The history and status of merging multiple sensor data: an overview, *Proceedings of ACSM-ASPRS Annual Convention, Image processing and remote sensing*, pp. 352-360.
- Rosenfeld, A. and Kak, A.C. (1982) *Digital Picture Processing*, Academic Press, New Yorkpp.
- Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K. (2004a) Fusing Airborne Laser Scanner Data and Aerial Imagery for The Automatic Extraction of Buildings in Densely Built-up Areas, *Proceedings of XXth ISPRS Congress, Commission 3*, Istanbul, Turkey, p. 512 ff.
- Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K. (2004b) Using the Dempster–Shafer method for the fusion of Lidar data and multispectral images for building detection, *Information Fusion*, Vol. 6, No. 4, pp. 283-300.
- Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K. (2007) Building detection by fusion of airborne laser scanner data and multispectral images: Performance

evaluation and sensitivity analysis, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. (in Press).

- Salvador, E., Cavallaro, A. and Ebrahimi, T. (2001) Shadow identification and classification using invariant color models, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1545-1548.
- Schenk, T. and Csatho, B. (2002) Fusion of Lidar data and aerial imagery for a more complete surface description, *IAPGIS*, Vol. XXXIV, No. 3A, p. 310–317.
- Schetselaar, E. (1998) Fusion by its transform: Should we use cylindrical or spherical coordinates?, *International Journal of Remote Sensing*, Vol. 19, No. 4, pp. 759-765.
- Schiewe, J. and Gahler, M. (2008) "Modelling uncertainty in high resolution remotely sensed scenes using a fuzzy logic approach", in: *Object-Based Image Analysis*, Blaschke, T., Lang, S. and Hay, G. J. (eds.), Springer, Berlin, pp. 755-768.
- Schiewe, J., Tufte, L. and Ehlers, M. (2001) Potential and problems of multi-scale segmentation methods in remote sensing, *Proceedings of GIS – Zeitschrift für Geoinformationssysteme 6/2001*, pp. 34-39.
- Schistad-Solberg, A.H., Jain, A.K. and Taxt, T. (1994) Multisource classification of remotely sensed data: fusion of Landsat TM and SAR image, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, pp. 768-778.
- Schowengert, R. (1980) Reconstruction of multispatial, multispectral image data using spatial frequency content, *Photogrammetric Engineering and Remote Sensing*, Vol. 46, No. 10, p. 1325–1334.
- Shackelford, A.K. and Davis, C.H. (2003) A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 10, pp. 2354-2363.
- Shettigara, V.K. (1992) A generalised component substitution technique for spatial enhancement of multispectral images using a higher resolution data set, *Photogrammetric Engineering & Remote Sensing of Environment*, Vol. 58, No. 5, pp. 561-567.
- Shettigara, V.K. and Sumerling, G.M. (1998) Height determination of extended objects using shadows in SPOT images, *Photogrammetric Engineering & Remote Sensing*, Vol. 64, No. 1, pp. 35-44.
- Singh, A. (1989) Digital change detection techniques using remotely-sensed data, *International Journal of Remote Sensing*, Vol. 10, No. 6, pp. 989-1003.

- Skole, D. and Tucker, C. (1993) Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988, *Science*, Vol. 260, p. 1905–1910.
- Smith, J.H., Stehman, S.V., Wickham, J.D. and Yang, L. (2003) Effects of landscape characteristics on land-cover class accuracy, *Remote Sensing of Environment*, Vol. 84, p. 342–349.
- Smith, M.J., Asal, F.F.F. and Priestnall, G. (2004) The Use of Photogrammetry and Lidar for Landscape Roughness Estimation in Hydrodynamic Studies, *Proceedings of ISPRS 2004 International Society for Photogrammetry and Remote Sensing*, Istanbul, Turkey, 12-23 July 2004, p. 714.
- Smits, P.C., Dellepiane, S.G. and Schowengerdt, R.A. (1999) Quality assessment of image classification algorithms for land-cover mapping: a review and proposal for a cost-based approach, *International Journal of Remote Sensing of Environment*, Vol. 20, p. 1461–1486.
- Sohn, G. and Dowman, I. (2007) Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 62, p. 43–63.
- Soille, P. (2003) *Morphological image analysis: principles and applications* (2nd Edition), Springer, Berlin ; New York. 316 pp.
- Soille, P. and Vincent, L. (1990) Determining watersheds in digital pictures via flooding simulations, *Proceedings of Visual Communications and Image*, Kunt, M. ed., Society of Photo Instrumentation Engineers, Bellingham, pp. 240-250.
- Solberg, A.H.S., Taxt, T. and Jain, A.K. (1996) A Markov random field model for classification of multisource satellite imagery, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 34, p. 100– 113.
- Stehman, S.V. (1997) Selecting and interpreting measures of thematic classification accuracy, *Remote Sensing of Environment*, Vol. 62, p. 77– 89.
- Stehman, S.V. (1999) Basic probability sampling designs for thematic map accuracy assessment, *International Journal of Remote Sensing*, Vol. 20, No. 12, pp. 2423-2441.
- Stehman, S.V. and Czaplewski, R.L. (1998) Design and analysis for thematic map accuracy assessment: fundamental principles, *Remote Sensing of Environment*, Vol. 64, p. 331– 344.
- Stein, A. (2004) Use of single- and multi-source image fusion for statistical decision-making, *International Journal of Applied Earth Observations and Geoinformation*, Vol. 6, No. 3-4 [SPECIAL ISSUE], pp. 229-239.

- Sua'raza, J.C., Ontiveros, C., Smith, S. and Snape, S. (2005) Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry, *Computers & Geosciences* (2005), Vol. 31, No. 2, p. 253–262.
- Tao, G. and Yasuoka, Y. (2002) Combining High Resolution Satellite Imagery and Airborne Laser Scanning Data for Generating bare land DEM in Urban Areas, *Proceedings of International Workshop on Visualization and Animation of Landscape*, International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, Kunming, China, 26 - 28 February 2002.
- Taxt, T., Flynn, P.J. and Jain, A.K. (1989) Segmentation of document images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 12, pp. 1322-1229.
- Teggi, S., Cecchi, R. and Serafini, F. (2003) TM and IRS-1C-PAN data fusion using multiresolution decomposition methods based on the 'a trous' algorithm, *International Journal of Remote Sensing*, Vol. 24, No. 6, pp. 1287-1301.
- Tickle, P.K., Lee, A., Lucas, R.M., Austin, J. and Witte, C. (2006) Quantifying Australian forest floristics and structure using small footprint LiDAR and large scale aerial photography, *Forest Ecology and Management*, Vol. 223, p. 379–394.
- Townshend, J.R.G., Justice, C.O., Gurney, C. and McManus, J. (1992) The Impact of Misregistration on Change Detection, *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 30, No. 5, pp. 1054-1060.
- Tso, B. and Mather, P.M. (2001) "Multisource classification", in: *Classification Methods for Remotely Sensed Data*, Taylor & Francis, New York, pp. 271-298.
- Tso, B.C. and Mather, P.M. (1999) Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 3, p. 1255–1260.
- Van-Der-Meer, F. (1999) Iterative spectral unmixing, *International Journal of Remote Sensing*, Vol. 20, No. 17, pp. 3431-3436.
- van-der-Sande, C.J., de-Jong, S.M. and de-Rooc, A.P.J. (2003) A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 4, No. 3, p. 217–229.
- van-der-Wal, D. and Herman, P.M.J. (2007) Regression-based synergy of optical, shortwave infrared and microwave remote sensing for monitoring the grain-size of intertidal sediments, *Remote Sensing of Environment*, Vol. 111, No. 1, pp. 89-106.

- van-der-Werf, H.M.A. and van-der-Meer, F.D. (2008) Shape-based classification of spectrally identical objects, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 63, No. 2, pp. 251-258.
- Vijayaraj, V. (2004) A Quantitative Analysis Of Pansharpened Images, *Master of Science*, Department of Electrical & Computer Engineering, Mississippi State University, Mississippi, Mississippi Statepp.
- VLS (2011) *Feature Analyst* [Online], Available: http://www.featureanalyst.com/feature_analyst.htm (15 July, 2011).
- Wal, D.v.-d. and Herman, P.M.J. (2007) Regression-based synergy of optical, shortwave infrared and microwave remote sensing for monitoring the grain-size of intertidal sediments, *Remote Sensing of Environment*, Vol. 111, pp. 89-106.
- Wald, L. (1998) A European proposal for terms of reference in data fusion, *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, No. 7, pp. 651-654.
- Wald, L. (1999) Some terms of reference in data fusion, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 3, pp. 1190-1193.
- Walter, V. (2004) Object-based classification of remote sensing data for change detection, *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 58, p. 225– 238.
- Walter, V. (2005) Object-based classification of integrated multispectral and Lidar data for change detection and quality control in urban areas, *Proceedings of ISPRS WG VII/1 "Human Settlements and Impact Analysis" 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas (URBAN 2005) and 5th International Symposium Remote Sensing of Urban Areas (URS 2005)*, Tempe, AZ, USA, March 14 - 16.
- Waltz, E. (2001) "The principles and practice of image and spatial data fusion", in: *Handbook of multisensor data fusion*, Hall, D. L. and Llinas, J. (eds.), CRC Press, Washington.
- Wang, L., Sousa, W.P. and Gong, P. (2004) Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery, *International Journal of Remote Sensing*, Vol. 25, No. 24, pp. 5655 - 5668.
- Wang, L., Sousa, W.P., Gong, P. and Biging, G.S. (2004b) Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama, *Remote Sensing of Environment*, Vol. 91, p. 432–440.
- Wang, Z. (1990) *Principles of Photogrammetry (with Remote Sensing)*, Press of Wuhan Technical University of Surveying and Mapping, Beijing, China. 257 pp.

- Wealands, S.R., Grayson, R.B. and Walker, J.P. (2005) Quantitative comparison of spatial fields of hydrological model assessment - some promising approaches, *Advances in Water Resources*, Vol. 28, pp. 15-32.
- Wehch, R. and Ehlers, M. (1988) Cartographic feature extraction from integrated SIR-B and Landsat TM images, *International Journal of Remote Sensing*, Vol. 9, pp. 873-889.
- Wehr, A. and Lohr, U. (1999) Airborne laser scanning—an introduction and overview, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, No. 68–82.
- Welch, R. (1982) Spatial resolution requirements for urban studies, *International Journal of Remote Sensing*, Vol. 3, No. 2, pp. 139-146.
- Welch, R. and Ehlers, M. (1987) Merging multiresolution SPOT HRV and Landsat TM data, *Photogrammetric Engineering & Remote Sensing*, Vol. 53, pp. 301-303.
- Welch, R., Jordan, T.R. and Ehlers, M. (1985) Comparative Evaluations of the Geodetic Accuracy and Cartographic Potential of Landsat-4 and Landsat-5 Thematic Mapper Image Data, *Photogrammetric Engineering & Remote Sensing*, Vol. 51, No. 9, pp. 1249-1262.
- Weszka, J.S. (1978) A survey of threshold selection techniques, *Computer Graphics and Image Processing*, Vol. 7, pp. 259-265.
- Weszka, J.S. and Rosenfeld, A. (1978) Threshold evaluation techniques, *IEEE Trans. Syst. ManCybern.*, Vol. 8, pp. 622-629.
- Wilson, N. (1995) *The Flooded Gum Trees: Land Use and Management of River Red Gums in New South Wales*, Nature Conservation Council of NSW, Sydneypp.
- Winkler, G. (1995) *Image analysis random fields and dynamic Monte Carlo methods*, Springer-Verlag, New Yorkpp.
- Wolter, P.T., Mladenoff, D.J., Host, G.E. and Crow, T.R. (1995) Improved Forest Classification in the Northern Lake States Using Multi-Temporal Landsat Imagery, *Photogrammetric Engineering & Remote Sensing*, Vol. 61, No. 9, pp. 1129-1143.
- Woodcock, C.E. and Strahler, A.H. (1987) The factor of scale in remote sensing, *Remote Sensing of Environment*, Vol. 21, pp. 311-332.
- Woodcock, C.E., Strahler, A.H. and Jupp, D.L.B. (1988) The use of variograms in remote sensing I: Scene models and simulated images, *Remote Sensing of Environment*, Vol. 25, p. 323–348.
- Wurmab, M., Taubenböcka, H., Schardtcd, M., Escha, T. and Dechab, S. (2011) Object-based image information fusion using multisensor earth observation

data over urban areas, *International Journal of Image and Data Fusion*, Vol. 2, No. 2, pp. 121-147.

- Yan, G., Mas, J.-F., Maathuis, B.H.P., Xiangmin, Z. and Dijk, P.M.V. (2006) Comparison of pixel-based and object-oriented image classification approaches - a case study in a coal fire area, Wuda, Inner Mongolia, China, *International Journal of Remote Sensing*, Vol. 27, No. 18, p. 4039–4055.
- Yang, X. (1997) Georeferencing CAMS Data: Polynomial Rectification and Beyond, *Doctoral Dissertation*, College of arts and sciences, University of South Carolinapp.
- Yanowitz, S.D. and Bruckstein, A.M. (1989) A new method for image segmentation, *Computer Vision Graphics and Image Processing*, Vol. 46, pp. 82-95.
- Yao, J. and Zhang, Z. (2006) Hierarchical shadow detection for color aerial images, *Computer Vision and Image Understanding*, Vol. 102, pp. 60-69.
- Yastikli, N. and Jacobsen, K. (2005) Direct sensor orientation fro large scale mapping - potential, problems, solutions, *The Photogrammetric Record*, Vol. 20, No. 111, pp. 274-284.
- Yuan, F. (2008) Land-cover change and environmental impact analysis in the Greater Mankato area of Minnesota using remote sensing and GIS modeling, *International Journal of Remote Sensing*, Vol. 29, No. 4, pp. 1169-1184.
- Yuan, F. and Bauer, M.E. (2006) Mapping impervious surface area using high resolution imagery: a comparison of object-oriented classification to per-pixel classification, *Proceedings of American Society of Photogrammetry and Remote Sensing Annual Conference*, Reno, NV, May 1–5, 2006, pp. CD-ROM.
- Zadeh, L.A. (1965) Fuzzy set, *Information and Control*, Vol. 8, pp. 338-353.
- Zhang, J. and Goodchild, M.F. (2002) *Uncertainty in Geographical Inforamtion*, Taylor & Francis, New York. 266 pp.
- Zhang, Z. and Blum, R.S. (1999) A categorization of multiscale decomposition-based image fusion schemes with a performance study for a digital camera application, *Proceedings of IEEE*, p. 1315–1326.
- Zhou, W., Huang, G., Troy, A. and Cadenasso, M.L. (2009) Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study, *Remote Sensing of Environment*, Vol. 113, p. 1769–1777.
- Zhou, W. and Troy, A. (2008) An object-oriented approach for analysing and characterizing urban landscape at the parcel level, *International Journal of Remote Sensing of Environment*, Vol. 29, No. 11, pp. 3119-3135.

Zhu, Z., Yang, L., Stehman, S.V. and Czaplewski, R.L. (2000) Accuracy assessment for the U.S. Geological Survey regional land-cover mapping programme: New York and New Jersey region, *Photogrammetric Engineering and Remote Sensing*, Vol. 66, p. 1425– 1435.

Zitova, B. and Jan, F. (2003) Image registration methods: a survey, *Image and Vision Computing*, Vol. 21, pp. 977-1000.

APPENDIX

Appendix 1. The forest study area field data collection form

Trip:	Site:	Date:
--------------	--------------	--------------

Photo reference/s: _____

Category	Attribute	Tree 1	Tree 2	Tree 3	Tree 4
Location	Coordinates (GDA94)	Easting			
		Northing			
Foliage & Canopy	Overall Height (m)				
	Canopy Area (sq. m)				
	Foliage Density (%)				
DBH	Bole Diameter (m)				
Tree Species	RRG, GB or BB				
Vertical Strata	Distance from Transect origin (m)				
	Transect Bearing (degree)				
	Transect point of origin				
	Bare Ground (%)				
	Canopy cover (%)				
Other	Ground vegetation cover (%)				

Appendix 2. The semi-urban study area field data collection form

Trip:	Site:	Date:
--------------	--------------	--------------

Photo reference/s: _____

Category	Attribute	Object 1	Object 2	Object 3	Object 4
Location	Coordinates (GDA94)	Easting			
		Northing			
Class	Roofs/Road/Trees/..etc				
	Overall Height (m)				
	Area cover (sq. m)				
Vertical Strata	Distance from Transect origin (m)				
	Transect Bearing (degree)				
	Transect point of origin				
	Mixed class (%)				
Other					

Appendix 3. Summary statistics of the initial 10 clusters of data-driven pixel-level fusion

Cluster	Layers	DN Mean	DN St. Dev.
1	Blue layer of original image	608.76	119.45
	Green layer of original image	329.51	48.48
	Red layer of original image	342.85	28.78
	Infrared layer of original image	887.97	164.05
	Height layer: LiDAR-derived nDSM	3.45	4.44
	Texture layer: LiDAR 1st return intensity	134.84	49.26
2	Blue layer of original image	777.92	77.81
	Green layer of original image	400.65	31.91
	Red layer of original image	373.97	21.82
	Infrared layer of original image	1353.78	109.88
	Height layer: LiDAR-derived nDSM	8.058	4.162
	Texture layer: LiDAR 1st return intensity	81.08	43.13
3	Blue layer of original image	899.85	68.26
	Green layer of original image	447.70	29.74
	Red layer of original image	406.20	18.73
	Infrared layer of original image	1174.03	85.19
	Height layer: LiDAR-derived nDSM	1.51	3.139
	Texture layer: LiDAR 1st return intensity	153.32	41.86
4	Blue layer of original image	1066.50	56.74
	Green layer of original image	517.44	24.04
	Red layer of original image	445.11	19.08
	Infrared layer of original image	1274.63	53.29
	Height layer: LiDAR-derived nDSM	0.29	1.42
	Texture layer: LiDAR 1st return intensity	172.64	29.24
5	Blue layer of original image	1072.53	59.79
	Green layer of original image	522.82	25.95
	Red layer of original image	443.79	17.95
	Infrared layer of original image	1420.30	53.15
	Height layer: LiDAR-derived nDSM	1.15	2.83
	Texture layer: LiDAR 1st return intensity	166.04	43.68
6	Blue layer of original image	1224.71	62.77
	Green layer of original image	574.76	24.29
	Red layer of original image	474.62	18.65
	Infrared layer of original image	1482.48	73.53
	Height layer: LiDAR-derived nDSM	0.71	2.14
	Texture layer: LiDAR 1st return intensity	164.96	37.01
7	Blue layer of original image	987.20	106.56
	Green layer of original image	483.75	42.25
	Red layer of original image	415.55	29.97
	Infrared layer of original image	1684.53	128.26
	Height layer: LiDAR-derived nDSM	6.73	4.34
	Texture layer: LiDAR 1st return intensity	94.49	47.85
8	Blue layer of original image	1365.51	94.34
	Green layer of original image	621.62	36.46
	Red layer of original image	498.14	26.71
	Infrared layer of original image	1688.86	110.51
	Height layer: LiDAR-derived nDSM	1.88	3.16
	Texture layer: LiDAR 1st return intensity	142.64	46.51
9	Blue layer of original image	1677.09	116.80
	Green layer of original image	734.52	48.13
	Red layer of original image	563.25	34.83
	Infrared layer of original image	1827.54	111.74
	Height layer: LiDAR-derived nDSM	1.13	2.34
	Texture layer: LiDAR 1st return intensity	147.448	38.71
10	Blue layer of original image	2160.23	241.90
	Green layer of original image	903.15	80.88
	Red layer of original image	651.26	50.79
	Infrared layer of original image	2102.31	157.64
	Height layer: LiDAR-derived nDSM	0.63	1.66
	Texture layer: LiDAR 1st return intensity	156.65	34.78

Appendix 4. 3-band colour imagery and LiDAR data fusion

The analysis of this Appendix was presented in section 7.3.2. Implementation of this experiment was described in this Appendix. A subset of semi-urban study area data was used for the experiment. The area covered 0.5km x 0.5km of semi-urban Mathoura. A total of 311 sample points were collected with a stratified random sampling technique to insure a balanced representation from all semi-urban landscape classes.

Pixel-level fusion

3-band colour imagery and LiDAR-derived nDSM data was used for the pixel-level fusion. An independent, supervised parallelepiped classification procedure was applied to them. The essential practical steps for the implementation of the supervised classification procedure were presented in Section 2.5.5.2. Landscape training signatures were also generated from 3-band colour imagery with LiDAR-derived nDSM data. The distance between signatures was assessed through both tests on their statistical separability and spectral profiling. The pixel-level fusion results were presented in the Section 7.5: accuracy assessment is presented in Table A4a.

Table A4a Error matrix using pixel-level fusion of colour imagery with LiDAR data in a semi-urban area

Fused data	Reference data							User's Acc. %
	Roof	Roof shadow	Tree	Tree shadow	Grass	Open space	Total	
Roof	38	5		1	3	6	53	71.70
Roof shadow	4	33		13		1	51	64.71
Tree			44	3	4		51	86.27
Tree shadow	2	8	7	34			51	66.67
Grass	1	1	4	2	43	1	52	82.69
Open space	6	2			3	42	53	79.25
Total	51	49	55	53	53	50	311	
Prod Acc %	74.51	67.35	80.00	64.15	81.13	84.00		234
Overall Accuracy (Percent)								75.24
Kappa Coefficient (\hat{K}) (Percent)								70.29

Visual inspection of the pixel- and object-level fusions of colour imagery with LiDAR data reveals the sharpness of the classes is dramatically changed (Compare Figure 7.7a and 7.7b). This is particularly evident when *Grass* and *Roof* results are examined. In pixel-level fusion, the ‘pepper and salt’ effect is prominent resulting in misclassification. Visual comparison by Kamagata *et al.* (2005) also revealed that pixel-level fusion failed to recognise shape variation within forest canopy and mixed vegetation. Feature-level fusion, on the other hand, did extract the boundaries between forest types. Similarly, this study concludes that feature-level fusion has a great potential for analysing landscape patterns even in highly heterogeneous and rapidly-changing urban areas.

The pixel-level error matrix (see Table A4a) reveals that *Roof shadow* class is poorly delineated, with User’s Accuracy 64.71 percent and Producer’s Accuracy 67.35 percent. The delineation of the *Tree Shadow* class is also relatively poor, with User’s Accuracy 66.67 percent and Producer’s Accuracy 64.15 percent. Feature-level fusion improves the *Roof Shadow* classification (User’s Accuracy 90.20 percent and Producer’s Accuracy 88.46 percent) by incorporating contextual information. Improvement is also noticed in classifying *Tree Shadow* by feature-level fusion (User’s accuracy 86.27 percent and Producer’s accuracy 91.67 percent compared to pixel-level fusions).

Feature-level fusion

For colour imagery with nDSM fusion, equal weight was given to both colour imagery and LiDAR data sources due to their equal radiometric depth (8-bit). The given scale factor was 25, which delineated feature with sufficient sizes. Smoothness and compactness weights were specified as 0.3 and 0.7 respectively (Table A4b), emphasising the discrete, compact nature of *House Roofs*.

Table A4b Segmentation parameters for the colour imagery with LiDAR-derived nDSM data fusion for level III landscape classification scheme

Parameters	Colour & nDSM fusion
Weight	1 Colour & 1 nDSM
Scale factor	25
Shape Factor	0.15
Compactness	0.3
Smoothness	0.7

The methodology for multispectral imagery and LiDAR data fusion was given in section 6.5. At Level I for fusion of colour imagery with nDSM data, the *Manmade* and *Obscure/Shadow* classes are discriminated using brightness values of the features. At Level II for fusion of colour imagery with nDSM data, the *Vegetation* class is separated from others by a fuzzy membership description of the mean and the ratio of the green spectral band. At Level III for fusion of colour imagery with nDSM data, *Vegetation* is further subdivided into *Tree* and *Grass* classes. The accuracy assessment of the feature-level fusion is presented Table A4c.

Table A4c Error matrix for feature-level fusion results for a semi-urban area using colour imagery with LiDAR data

Fused data	Reference data							User's Acc. %
	Roof	Roof shadow	Tree	Tree shadow	Grass	Open space	Total	
Roof	50	1				2	53	94.34
Roof shadow	2	46		1	1	1	51	90.20
Tree	1		47	1	2		51	92.16
Tree shadow		4	1	44	2		51	86.27
Grass	1		2	1	48		52	92.31
Open space		1	1	1	1	49	53	92.45
Total	54	52	51	48	54	52	311	
Prod Acc %	92.59	88.46	92.16	91.67	88.89	94.23		284
Overall Accuracy (Percent)								91.32
Kappa Coefficient (\hat{K}) (Percent)								89.93

Feature-level results compared with those of pixel-level fusion confirm a significant improvement in identification of classes throughout the scene (error matrices presented in Appendix 4). The summary of fusion results accuracy in Table 7.7 shows the Kappa Coefficient calculated for the summary accuracy is 3.15 confirming the feature-level fusion result is better than the pixel-level fusion result irrespective of whether colour or multispectral data is used in the fusion.

Appendix 5. Accuracy assessment of pixel- and feature-level classification using colour imagery only

The analysis of this Appendix was presented in section 7.3.3. Here the implementation part was presented. Semi-urban landscape object delineation resulting from the exclusion or inclusion of nDSM data with colour imagery is evaluated through statistical analysis. Pixel- and feature-level classification accuracy assessments are performed here to compare results obtained from excluding height of objects derived from LiDAR data. The steps of pixel- and feature-level classification were given in section 6.4 and 6.5 respectively. To ensure consistency, the same sampling technique was adopted for each of the fusion processes and compared with results in Appendix 4. A total of 311 sample points were collected with a stratified random sampling technique, which insured a balanced representation from all the classes. The object delineation accuracies were assessed independently to reduce systematic bias and to counter the temporal effect on overall accuracy.

Table A5a Error matrix for pixel-level classification using colour imagery only for a semi-urban area

Fused data	Reference data							User's Acc. %
	Roof	Roof shadow	Tree	Tree shadow	Grass	Open space	Total	
Roof	36	4	1	3	2	7	53	67.92
Roof shadow	2	37		8	1	3	51	72.55
Tree			41	4	3	3	51	80.39
Tree shadow	1	10	5	33	2		51	64.71
Grass	1	2	7	4	34	4	52	65.38
Open space	12	3			4	34	53	64.15
Total	52	56	54	52	46	51	311	
Prod Acc %	69.23	66.07	75.93	63.46	73.91	66.67		215
Overall Accuracy (Percent)								69.13
Kappa Coefficient (\hat{K}) (Percent)								62.96

Table A5b Error matrix for feature-level classification using colour imagery only

Fused data	Reference data							User's Acc. %
	Roof	Roof shadow	Tree	Tree shadow	Grass	Open space	Total	
Roof	35	7	1		1	9	53	66.04
Roof shadow	6	35		8	2		51	68.63
Tree			34	9	5	1	51	66.67
Tree shadow		8		35	4	1	51	68.63
Grass	2	1	7	1	39	2	52	75.00
Open space	15	3	1	1	4	29	53	54.72
Total	58	56	46	54	55	42	311	
Prod Acc %	60.34	62.50	73.91	64.81	70.91	69.05		207
Overall Accuracy (Percent)								66.56
Kappa Coefficient (\hat{K}) (Percent)								60.95