Enhancing the Spatial Resolution of Landsat Data for Mapping Urban Areas

by

RAHMAN MOMENI, BSc, MSc

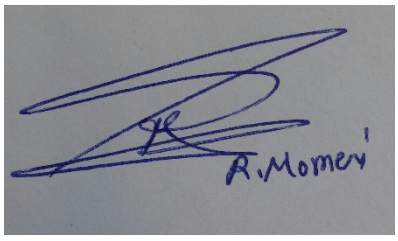


Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

June 2017

This thesis is the result of work done wholly, while I was in registered postgraduate candidature at the University of Nottingham.

Rahman Momeni 31/05/2017



Abstract

Detailed land cover information is crucial for mapping and managing complex urban environments across local and regional scales (Zhou and Qiu, 2015). This thesis is based on the proposition that spatial resolution is the most influential factor when mapping complex urban environments, compared to imagery's spectral properties and type of classifier. As such, the modern "very high resolution" sensors (i.e., WorldView-2) offer a significant advantage for mapping, however using such imagery is a costly and resource-hungry approach. The coarser resolution of Landsat data (30m) is the key limitation for using these data, yet they are free and now have a temporal legacy. This doctoral research assesses the potential of using an approach that enhances the spatial resolution of Landsat data for urban land cover mapping, namely sparse representation. Focusing on the land cover mapping of the urban area of Nottingham, UK, and after establishing the superior role of spatial resolution on the accuracy of that mapping, this research demonstrates the potential of this approach. Moreover, some parameters around its use are established, in particular, the transferability of this method over space and time. It should be noted the potential of sparse representation can be even more significant by using finer spatial resolution products (i.e., Sentinel-2 and SPOT with 10m). This reaffirmed the importance of the spatial resolution for urban land cover mapping. Then it presents the sparse representation as a successful method to enhance the spatial resolution of Landsat data for urban land cover mapping.

“There is only one corner of the universe you can be certain of improving, and that is your own self”.

Aldous Huxley

Publications

Momeni, R., Boyd, D., Foody, G., and Aplin, P., Using sparse representation to enhance the spatial capacity of Landsat data for urban land cover mapping [in the process of writing],

Momeni, R., Aplin, P., and Boyd, D., (2016), Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach, *Remote Sensing Journal* (with **impact factor 3.2**), Vol. 8 (2), 88.

Momeni, R. and Aplin, P., (2014), Assessing the Performance of OBIA and Per-Pixel Based Classification Approaches to Map Detailed Urban Area, [Poster], 5th Annual GEOBIA Conference, 21–23 May, Thessaloniki, Greece.

Momeni, R., Aplin, P., and Boyd, D., (2013), A Simple Experiment to Test the Potential of WorldView-2 Imagery for Urban Land Cover Classification: a Comparison of Spatial, Spectral and Classifier Characteristics, [Poster], RSPSoc 2013 Annual Conference, Earth Observation for Problem Solving, 4–6 September, Glasgow, UK.

Recent oral presentations

Momeni, R., Boyd, D., Foody, G., and Aplin, P., Urban mapping at high resolutions: exploiting the long-term EO record, Earth Observation for Knowledge Discovery, RSPSoc Annual Conference 2016 at The University of Nottingham.

Momeni, R., Aplin, P., and Boyd, D., A rigorous comparison of spatial, spectral and classifier properties when mapping complex urban land cover from spaceborne imagery, *Earth Observation in the Sentinel Era, RSPSoc, NCEO and CEOI-ST Joint Annual Conference 2015,* The University of Southampton.

Momeni, R., Aplin, P., and Boyd, D., First year PhD student Presentation to providing an overview on my research Ordnance, 2012 October, University of Nottingham

Momeni, R., Classification of Land-Use Patterns: Based on Extracting the Impervious Surface in the City of Sheffield, Ordnance Survey Annual MSc GIS/RS Seminar 2010 in The Southampton, UK.

Lecturing experience at the Kingston University

- Remote sensing module (Spring 2016-17)
- GIS field trip (May 2017)

Demonstration experience at the University of Nottingham

- Earth Observation (Autumn and Spring 2012-13; 2013-14; 2014-15)
- Introduction to GIS (Autumn 2012-13; 2013-14; 2014-15; 2015-16)
- Digital Explorer (Autumn and Spring 2013-14; 2014-15; 2015-16)
- Mobile field GIS (Spring 2012-13; 2013-14; 2014-15)
- Geospatial Information Services (Autumn 2012-13; 2013-14)

- Advances in Remote Sensing (Autumn 2014-15)

Acknowledgements

First of all, I would like to express my sincere gratitude and thanks to all my three supervisors, Dr. Doreen Boyd, Professor Giles Foody, and Professor Paul Aplin. They all did a great effort on guiding my research and improving my writing skill.

I am grateful to Professor Aplin for accepting me as his student and providing guidance, helps and continuous support throughout the first three years of this study. Also, special thanks to Dr Boyd for guiding the research in this study and writing up the process. My thanks also go for my co-supervisor Professor Foody for his consultant during this study.

I would like to convey my thanks to my colleagues Dr Tracey Mooney, Dr Xiao Feng, and Dr Hugo Gomes da Costa for their kind help throughout my study.

Thanks also to all academic and admin staff of the School of Geography, the University of Nottingham for their co-operation and support especially to Andrea Payne and Ian Conway.

I would like to thank my parents for dedicating their life to providing their children with the support needed to continually push themselves to succeed. I also like to thank them for being a constant source of inspiration.

Last but not the least; I would like to thank my wife who supports me during this study.

List of Figures

Figure 1.1. The structure of thesis……12

Figure 2.1. The relationship between objects and pixels depend on variation of spatial resolution……35

Figure 3.1. Nottingham, UK study area location and WorldView-2 image……45

Figure 3.2. WV-2 spectral wavebands (top line) and four and six spectral bands data sets used for comparative analysis……46

Figure 3.3. The 36 image data set/classifier combinations (4 spatial resolutions × 3 spectral band sets × 3 classifiers) used for comparative classification analysis……48

Figure 3.4. The multi-stage OBIA\NN procedure (this chart especially demonstrates the OBIA\NN of WV-2 imagery with 2m spatial resolution)……60

Figure 3.5. Land cover maps (detail) for the 36 data set/classifier combinations, plus the WorldView-2 image (© DigitalGlobe, Inc. All Rights Reserved) and MasterMap vector data……67

Figure 3.6. The overall classification accuracies for the 36 data set/classifier combinations (ML, SVM, and OBIA/NN)……68

Figure 3.7. Matrix of McNemar test results showing the statistical significance of differences between all classification pairs (ML = maximum likelihood, SVM = support vector machine, OBIA/NN = object-based image analysis/ nearest neighbour; 30m, 10m, 4m and 2m refer to spatial resolution; four, six and eight bands refer to the number of spectral bands)……74

Figure 3.8. The overall accuracy of 36 land cover maps by highlighting The influence of spatial resolution……76

Figure 3.9. The McNamar's statistical test of a 30m spatial resolution against 10m, 4m, and 2m spatial resolution……77

Figure 3.10. The influence of using various spectral bands on overall classification accuracy of 36 land cover maps……79

Figure 3.11. The relationship between spectral and spatial resolution……80

Figure 3.12. Overall classification accuracies: the relationship between spectral resolution and classification approach……82

Figure 4.1. Image observation model which demonstrates factors impacting the quality of the image during acquisition (Yang and Huang, 2010). It highlights the key components of signal distortion including discrete sampling, motion, blur, and down-sampling. It is possible to retrieve the high-quality image by inverting this process……90

Figure 4.2. Image aliasing: (a) is the condition with aliasing, (b) is the condition without aliasing……91

Figure 4.3. A trained dictionary with total 512 atoms which trained with each atom of size9……104

Figure 4.4. The diagram summarises the sparse representation as an inverse problem……109

Figure 4.5. a) A fine resolution image ($\boldsymbol{y_h}$) degraded by zoom factor 3 creating. B) A course resolution image ($\boldsymbol{x_l}$). C referred as performing bicubic interpolation to up sampled it according the to by $\boldsymbol{y_h}$ scale……110

Figure 4.6. The workflow of the training phase……114

Figure 4.7. The true RGB colour of training image with 2m spatial resolution……115

Figure 4.8. The true RGB colour of degraded 2m into 30m……116

Figure 4.9. The true RGB colour of bilinear interpolation 30m to 2m……116

Figure 4.10. The true RGB colour of subtracting $\boldsymbol{Y_h - Y_l}$……117

Figure 4.11. The true RGB colour of high pass filtering the $\boldsymbol{y_l}$……118

Figure 4.12. The workflow of the reconstruction phase……121

Figure 4.13. The true RGB colour of Landsat OLI with 30m ($\boldsymbol{x_l}$)……123

Figure 4.14. The true RGB colour of up-sampled Landsat 30m to 2m ($\boldsymbol{y_l}$)……123

Figure 4.15. The true RGB colour of Landsat OLI high pass filtered with 2m resolution ($\boldsymbol{y_l'}$)……123

Figure 4.16. The true RGB colour of PCA on filtered Landsat OLI with 2m resolution ($\boldsymbol{y_l''}$)……124

Figure 5.1. The Landsat OLI imagery of Nottingham as the study area……129

Figure 5.2. Spectral sets visual demonstration for Landsat OLI and WV-2……131

Figure 5.3. Visual comparison of pan-sharpening of a. HCS, b. IHS, c. Brovery, d. Gram-Schmidt, e. Wavelet merge resolution, f.WV-2 and g. Landsat OLI.....141
Figure 5.4. a) zoom factor role for reconstruction phase by increasing the spatial resolution from 2.5m to 10m, b) zoom factor role for training phase by degrading spatial resolution from 10m to 2.5m.....145
Figure 5.5. OBIA/SVM two-stage segmentation and classification diagram.....154
Figure 5.6. 15 small subsets of five different images (RGB) of Nottingham: i) Landsat OLI 30m, ii) pan-sharpened Landsat OLI 15m, iii) spatially enhanced Landsat OLI 2m using sparse representation, iv) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation, and v) WV-2 2m) according to three scales (1:20,000, 1:10,000, and 1:5000).....160
Figure 5.7. 10 small subsets (750 × 750m) of land cover maps which were produced by performing SVM and OBIA/SVM over five data sets: i) Landsat OLI 30m, ii) pan-sharpened Landsat OLI 15m, iii) spatially enhanced Landsat OLI 2m using sparse representation, iv) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation, and v) WV-2 2m.....161
Figure 5.8. The overall classification accuracies for the ten land cover maps.....163
Figure 5.9. The user accuracy of each class for five classified land cover maps (by averaging two classifiers).....176
Figure 5.10. The producer accuracy of each class for five classified land cover maps (by averaging two classifiers).....177
Figure 6.1. Landsat OLI image of Derby.....182
Figure 6.2. Landsat ETM+ image captured Nottingham on 16th of May 2000.....183
Figure 6.3. Comparing the distribution of spectral properties of Landsat ETM+ (without thermal band) and WV-2 images.....185
Figure 6.4. 9 small subsets of three different images of Derby (i) Landsat OLI 30m, ii) spatially enhanced Landsat OLI 2m using sparse representation, iii) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation) according to three scales (1:20,000, 1:10,000, and 1:5000).....193
Figure 6.5. 9 small subsets of three different images of historical Nottingham (i) Landsat ETM+ 30m, ii) spatially enhanced Landsat ETM+ 2m using sparse

representation, iii) spatially enhanced pan-sharpened Landsat ETM+ 2m using sparse representation) according to three scales (1:20,000, 1:10,000, and 1:5000)……..194
Figure 6.6. 6 small subsets of classified Land cover maps (a. Landsat OLI of Derby, b. spatially enhanced Landsat OLI of Derby, c. spatially enhanced pan-sharpened Landsat OLI of Derby, d. Landsat ETM+ of Nottingham, e. spatially enhanced Landsat ETM+ of Nottingham, f. spatially enhanced pan-sharpened Landsat ETM+ of Nottingham……196
Figure 6.7. The overall accuracies of performing the sparse representation on original and pan-sharpened Landsat dada (i.e., OLI for Derby and ETM+ for Nottingham)……197
Figure 6.8. The McNemar statistical test on the capability of sparse representation method with regard to spatial transferability factor, to spatially enhance Landsat OLI for land cover mapping……201
Figure 6.9. The McNemar statistical test on the capability of sparse representation method with regard to temporal transferability factor, to spatially enhance Landsat ETM+ for land cover mapping……201
Figure 6.10. The producer accuracies of seven classes for Derby……203
Figure 6.11. The user accuracies of seven classes for Derby……203
Figure 6.12. The producer accuracies of seven classes of Nottingham……204
Figure 6.13. The user accuracies of seven classes of Nottingham……204
Figure 7.1. User and producer accuracies of 11 land cover classes highlighting the influence of spectral resolution (averaging the spatial resolution and classifier types)……217
Figure 7.2. The comparison between the pixel-based and object-based accuracies on mapping vegetated areas by summarising the influence of spectral resolution……222
Figure7.3. The overall accuracies of Landsat data and its spatially enhanced versions (i.e., using SVM classier), representing through three scenarios……225
Figure 7.4. 3 small subsets with 750m × 750m size of land cover maps representing when there is spatial and temporal coincident between WV-2 and Landsat OLI (a); Landsat OLI with 30m, (b); enhanced pan-sharpened Landsat OLI using sparse representation method with 2m resolution, (c); actual WV-2 imagery with 2m resolution……228

Figure7.5. The producer accuracies of Landsat OLI of Nottingham and Theirs's spatially enhanced versions when there are spatial and temporal coincident between Landsat OLI and WV-2 (the first scenario)……240
Figure7.6. The producer accuracies of Landsat OLI of Derby and their spatially enhanced versions when there is a spatial separation between Landsat OLI and WV-2 (the second scenario)……240
Figure7.7. The producer accuracies of Landsat ETM+ of Nottingham and their spatially enhanced versions when there is a temporal separation between Landsat ETM+ and WV-2 (the third scenario)……241

List of Tables

Table 2.1. The specifications of IKONOS, QuickBird, GeoEye-1, WV2, and WV-3……..22

Table 3.1. Classification scheme of Nottingham urban land cover with eleven members……49

Table 3.2. Segmentation parameter settings for multi-stage object-based classification……61

Table 3.3. Error matrix for performing OBIA/NN on WV-2 with 2m spatial resolution and eight spectral bands……70

Table 3.4. Error matrix for performing OBIA/NN on degraded WV-2 with 4m spatial resolution and eight spectral bands……70

Table 3.5. Error matrix for performing SVM on WV-2 with 2m spatial resolution and six spectral bands……71

Table 3.6. Error matrix for performing SVM on degraded WV-2 with 4m spatial resolution and six spectral bands……71

Table 3.7. Error matrix for performing ML on degraded WV-2 with 10m spatial resolution and four spectral bands……72

Table 3.8. Error matrix for performing ML on degraded WV-2 with 30m spatial resolution and four spectral bands……72

Table 4.1. A general comparison of frequency and spatial domain……97

Table 5.1. Spectral properties of Landsat OLI with their corresponding spatial resolution……130

Table 5.2. The result of five metrics, highlighted values represent the most accurate……138

Table 5.3. Normalised quantitative and metric pan-sharpening quality assessment……139

Table 5.4. The parameters were allocated for both training and reconstruction phase……148

Table 5.5. The land cover classification scheme used for Nottingham……150

Table 5.6. The details on values were allocated for scale, shape, and compactness……155

Table 5.7. The error matrix for the OBIA/SVM classification on spatially increased Landsat OLI with 2m spatial resolution using sparse representation……164

Table 5.8. The error matrix for the SVM classification on spatially increased Landsat OLI with 2m spatial resolution using sparse representation……164

Table 5.9. The error matrix for the OBIA/SVM classification on spatially increased pan-sharpened Landsat OLI with 2m spatial resolution using sparse representation……165

Table 5.10. The error matrix for the SVM classification on spatially increased pan-sharpened Landsat OLI with 2m spatial resolution using sparse representation……165

Table 5.11. The error matrix for the SVM classification on WV-2 imagery with 2m spatial resolution……166

Table 5.12. The error matrix for the SVM classification on Landsat OLI imagery with 30m spatial resolution……166

Table 5.13. The result of McNemar statistical test (statistically significant at 99% level of confidence shows bolded and underlined)……167

Table 6.1. The Landsat ETM+ spectral properties and their corresponding spatial characteristics……184

Table 6.2. The land cover classification scheme which for land cover mapping of Derby……185

Table 6.3. The error matrix of performing SVM on Landsat ETM+ of Nottingham with 30m spatial resolution……198

Table 6.4. The error matrix of performing SVM on Landsat OLI of Derby with 30m spatial resolution ……199

Table 6.5. The error matrix of performing SVM on spatially increased Landsat OLI of Derby with 2m spatial resolution, using sparse representation ……199

Table 6.6. The error matrix of performing SVM on spatially increased pan-sharpened Landsat OLI of Derby with 2m spatial resolution, using sparse representation……200

Abbreviations

AFI: Area Fit Index

CC: Correlation Coefficient

DIV: Difference in variance

ERGAS: Erreure Relative Globale Adimensionnelle de Synthese

HCS: Hyperspectral Colour Space

HIS: Hue-intensity-saturation

K-SVD: k means clustering method-singular value decomposition

L\H-resolution: Low\High resolution

Landsat ETM+: Landsat Enhanced Thematic Mapper plus

Landsat MSS: Multispectral Scanner System

Landsat OLI: Landsat Operational Land Imager

Landsat TM: Landsat Thematic Mapper

ML: Maximum Likelihood

NDBRI: Normalized Difference Brick Roof Index

NDBSI: Normalized Difference Bare Soil Index

NDVI: Normalized Difference Vegetation Index

NIR: Near Infrared

NN: Nearest Neighbour

OBIA: Object-Based Image Analysis

OLI: Operational Land Imager

OSGB: Ordnance Survey Great Britain

PCA: Principal component analysis

QR: Quality Rate

RGB: red, green, and blue

RMSE: Root Mean Square Error

Sentinel-2 MSI: Sentinel-2 Multi Spectral Instrument

SPOT: Pour l'Observation de la Terre

SRM: Super-resolution Mapping

SRR: Super-resolution Restitution

SVM: Support Vector Machine

TD: Transformed Divergence

VHR: Very High Resolution

WV-2: WorldView-2

Table of contents

Abstract ..... **i**
Publications ..... **iii**
Recent oral presentations ..... **iii**
Lecturing experience at the Kingston University ..... **iv**
Demonstration experience at the University of Nottingham ..... **iv**
Acknowledgements ..... **v**
List of Figures ..... **vi**
List of Tables ..... **xi**
Abbreviations ..... **xiii**
Table of contents ..... **xiv**
Chapter 1: Introduction **1**
1.1 Research rationale 1
1.2 Very High Spatial Resolution Sensors 3
1.3 Exploiting VHR Data 5
1.4 Research objectives 8
1.5 Outline 9
Chapter 2: Remote Sensing and Urban Environment **13**
2.1 Overview 13
2.2 Urbanisation and its environmental effects 14
2.3 Urban remote sensing 15
2.3.1 Urban remote sensing challenges 17
2.4 Image spatial resolution 18
2.4.1 Medium spatial resolution multispectral imagery 18
2.4.2 Fine spatial resolution multispectral imagery 20
2. 5 Image spectral characteristics 23
2.6 Land cover classification 24
2.6.1 Pixel-based classification 25
2.6.2 Object-based image analysis 33
Chapter 3: Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set, and Classification Approach **41**
3.1 Overview 41
3.2 Research materials 44

3.2.1. Study area 44
3.2.2 Data and classification scheme 46
3.3 Research methods 49
3.3.1 Pre-processing 49
3.3.2 Spectral indices 50
3.3.3 Reference data 51
3.4 Land cover classification 51
3.4.1 Pixel-based training 52
3.4.2 Pixel-based classification 55
3.4.3 Object-Based Image Analysis (OBIA/NN) 57
3.5 Accuracy assessment and statistical testing 64
3.5.1 Error matrix 64
3.5.2 McNemar statistical Test 65
3.6 Results 66
3.6.1 General overview 66
3.6.2 Spatial enhancement 75
3.6.3 Spectral enhancement 77
3.6.4 Classification approach 80
3.7 Summary 83
Chapter 4: Super Resolution Image Analysis 85
4.1 Overview 85
4.2 Super-resolution mapping (SRM) 86
4.2.1 Types of SRM methods 87
4.2.3 Pros and cons of SRM 88
4.3 Super-resolution restitution 89
4.4 Image observation model 90
4.5 Super-resolution and the reconstruction phase 91
4.5.1 Frequency domain 92
4.5.2 Spatial domain 93
4.6 Super-resolution and the number of training samples 98
4.6.1 Multi image and example-based 98
4.6.2 Example-based super-resolution 99
4.7 Dictionary 102
4.7.1 Types of dictionaries 104
4.7.2 Joint dictionary 105

4.8 Sparse representation 106
4.8.1 Sparse representation principle 106
4.8.2 Compressed sensing 108
4.8.3 Assumptions 108
4.8.4 Sparse representation algorithm 112
Chapter 5: Using Sparse Representation to Enhance Landsat OLI Imagery Using A Single WorldView-2 Image 126
5.1 Overview 126
5.2 Research Materials 128
5.2.1 Study area 128
5.2.2 Data 129
5.3 Landsat OLI spatial enhancement strategies 132
5.4 Pan-sharpening Landsat OLI 133
5.4.1 Assessing the pan-sharpening quality 133
5.5 Sparse representation to enhance Landsat OLI imagery 142
5.5.1 Introduction 142
5.5.2 Pre-processing 143
5.5.3 Training phase 144
5.5.4 Reconstruction phase 147
5.5.5 Post-processing 148
5.6 Classification 149
5.6.1 Classification scheme 149
5.6.2 Classification training 150
5.6.3 Classification methods 152
5.7 Accuracy assessment and statistical test 156
5.8 Results 157
5.8.1 Class level details 171
5.9 Summary 177
Chapter Six: Assessing the Potential of Sparse Representation Method for Transferability Over Space and Time 180
6.1 Introduction 180
6.2 Research Materials 181
6.2.1 Study area 181
6.2.2 Data and classification scheme 182
6.3 Pan-sharpening of the Landsat OLI 185
6.4 Sparse representation analysis 186

6.4.1 Introduction 186
6.4.2 Pre-processing 187
6.4.3 Training phase 187
6.4.4 Reconstruction phase 188
6.4.5 Post-processing 188
6.5 Classification 189
6.5.1 Training procedure 189
6.5.2 Classification approaches 190
6.6 Accuracy assessment and statistical testing 191
6.7 Results 191
6.7.1 Class level details 205
6.8 Summary 209
Chapter 7: Discussions and Conclusions 211
7.1 Overview 211
7.2 The influence of spatial resolution 213
7.3 The influence of spectral resolution 215
7.4 The influence of the choice of classifier 218
7.5 Project requirements versus project resources 223
7.6 Sparse representation and Landsat data 224
7.7 Sparse representation and pan-sharpened Landsat data 226
7.8 Classification scheme and the overall mapping accuracy 228
7.9 Classifier choice 230
7.10 Training image characteristics 231
7.11 Zoom factor 234
7.12 Sparse representation algorithm 235
7.13 Applications 236
7.13.1 Urban mapping 236
7.13.2 Historical applications 237
7.13.3 Disaster management 238
7.14 Summary 238
7.15 Conclusions 241
7.16 Future work 242
7.14.1 Applications 242
7.14.2 Sparse representation algorithm 243
Bibliography 246

8. Appendix 269

8.1 OBIA procedure for data sets with 30, 10, and 4m spatial resolution from Chapter 3.....269

8.2 The remaining 30 error matrices from Chapter 3 270

8.3 The remaining four error matrices from Chapter 5 285

8.4 List of all satellite images.....287

8.4.1 Images were used in the chapter 3 287

8.4.2 Images were used in the chapter 5 288

8.4.3 Images were used in the chapter 6 289

8.5 MATLAB scripts.....289

Chapter 1: Introduction

1.1 Research rationale

Although urban areas cover less than 3% of the global terrestrial surface, they have a significant effect on environmental processes including hydrology, biodiversity, biogeochemistry and climate at a range of scales (i.e., local, regional and global (Grimm *et al.,* 2008). According to the United Nations Population Division, currently more than half of the world's population lives in urban areas, and this number is expected to increase to 67.2% by 2050 (United Nations, 2012). As such, there is a pressing need for urban analysis to help the understanding of the status of urban areas and the quality of life of urban dwellers (Nichol, 2009). Site-specific information on the morphology and composition of urban areas is required in order to take appropriate action to mitigate undesirable impacts and plan for the future (Sanchez, 2004). However, this is required for ever increasing temporal resolutions to track the rate of change, particularly in the larger and most complex of urban areas, cities.

The synoptic view afforded by remote sensing offers great potential for data collection about cities. However, the data desired often tends to be related to population numbers or distributions and/or land use, and this is not directly observable by remote sensing (Nichol, 2009). Often the directly observable land cover information is the first step to the collection of the desired data on cities, and this can be done so at a range of scales (Lu and Weng, 2007; Zhou *et al*., 2011). Land cover data for cities provides crucial input data for a number of

applications including green space analysis (Lang *et al*., 2008), urban hydrological modelling (Han and Burian, 2009), urban land-use mapping (Herold *et al*., 2003b) and other such urban planning and management activities (Rogan and Chen, 2004; Müller *et al*., 2010). Thus to avoid error propagation downstream, there is a need for highly accurate land cover data of cities.

Although remote sensing is a common source of land cover data and often the only practical and cost-effective means of generating such information over large areas (Zhou and Qiu, 2015), the mapping of land cover of cities poses a significant challenge for remote sensing due to the high spatial frequency of surface features (Herold *et al*., 2003a; Lu *et al*., 2010; Moreira and Galvao, 2010). Urban land is highly heterogeneous comprising a mosaic of both human-made materials (such as asphalt, concrete, roof tiles and other impervious surfaces) and semi-natural surfaces (i.e., grass, trees, bare soil, and water). Early work using the relatively coarse spatial resolution sensors (i.e., Landsat Thematic Mapper (TM) imagery with its 30m multispectral spatial resolution) was compromised by the spatial resolution of the remotely sensed data (Guindon *et al*., 2004; Small and Lu, 2006). This level of spatial detail inevitably leads to mixed pixels, whereby each pixel exhibits some spectral average representing multiple surface features (Lu *et al*., 2011; Ling *et al*., 2012; Zhang *et al*., 2014). Commonly there was the situation as described by Strahler *et al*.'s (1986) L-resolution model, and attempts to overcome this included spectral unmixing (i.e., a single L-resolution pixel of the dense residential area may be

recognisable as the signatures of several different ground materials such as vegetation, buildings, and roads).

1.2 Very High Spatial Resolution Sensors

A technological advancement in satellite remote sensing came around the turn of the millennium with the advent of so-called “very high resolution” (VHR) satellite sensors, led by the 4m spatial resolution (multispectral) IKONOS mission in 1999. It was then followed by a series of other instruments including OrbView-3 (with 4m spatial resolution), QuickBird (2.4m) and GeoEye-1 (1.6m) (Aplin, 2003). Thus, a different paradigm in the land cover mapping of urban areas ensued. Here Strahler *et al*.’s (1986) H-resolution model prevailed (i.e., the scene elements are typically larger than image pixel). Many studies demonstrated how the fine spatial resolution VHR remotely sensed data enabled relatively accurate identification of small urban features (Inglada and Michel, 2009, Weng, 2012, Qin *et al*., 2015). This potential was further cemented when a generation of VHR satellite sensors emerged with enhanced spectral, as well as advanced spatial, properties. It is expected that consideration of more wavebands and higher spectral resolution may prove especially valuable for urban mapping, enabling subtly different spectral classes to be identified (Poursanidis *et al*., 2015). This advantage is likely to be most pertinent where a detailed classification scheme is involved, for instance when identifying many different land cover classes in complex urban environments. For instance, a planning agency official conducting a land use

inventory may wish to map several different types of roofing materials rather than a single “buildings” class (Fiumi, 2012).

That said, in contrast to the problem of mixed pixels which occurs where image spatial resolution is too coarse, the use of VHR imagery can give rise to a situation where “over-sampling” of the scene occurs. With this latter within-feature variation (occurring where image resolution too fine), pixel-based classification accuracy may be reduced (Lu *et al*., 2010, Aplin and Smith, 2011, Liu *et al*., 2013, Huang and Zhang, 2013). Certainly, the use of statistical parametric classifiers such as the maximum likelihood (ML) algorithm (Lu and Weng, 2004, Myint *et al*., 2011, Qin *et al*., 2015), which makes certain statistical assumptions about the data, is a challenge when used to classify VHR imagery. Specifically, ML classification tends to work well where training data are relatively noise-free and spectrally distinct between classes, such as where coarse spatial resolution imagery is used to classify general land cover classes. Where training data are rather noisier, such as where fine resolution imagery is used to map complex environments, (i.e., urban areas), the ML classifier can be considerably less accurate (Myint *et al*., 2011; Lu *et al*., 2011). Referring to statistical aspects, pixel-based approaches as a whole have limitations when it comes to urban analysis using VHR imagery (Campbell, 2007; Myint *et al*., 2011). Non-parametric pixel-based classifiers such as support vector machines (SVMs) seem well-suited to VHR urban classification since they are better able to handle noisy training data, compared to for instance the ML classifier (Mountrakis *et al*., 2011; Poursanidis *et al*., 2015).

That said, to fully exploit the VHR data (and bearing in mind the H-resolution model (Strahler *et al.,* 1986) that tends to prevail) there has been a movement away from traditional per-pixel-based land cover mapping approaches (Blaschke *et al*., 2014; Poursanidis *et al*., 2015). In particular, object-based image analyses have been explored with the premise that this approach directly addresses, and to an extent overcomes, the problem of within-feature variation and its attendant (pixel-based) misclassification (Lu and Weng, 2007; Zhou *et al*., 2008; Li and Shao, 2014). An image object is defined as a group of pixels sharing similar spectral and textural properties (Huang and Ni, 2008). The fundamental assumption of OBIA is that relatively homogeneous patches construct each image and that the size of the object is larger than a pixel. Object-based classification, which exploits spatial, textural and topological (as well as spectral) information (Walker and Blaschke, 2008, Bhaskaran *et al*., 2010; Blaschke *et al*., 2011) may facilitate highly accurate mapping of complex urban environments using VHR imagery. Moreover, object-based classification has grown in popularity, whereby objects are allocated to land cover classes to represent real-world features instead of somewhat arbitrary pixel structures (Aplin and Smith, 2011). Other practitioners have tested spatial indices and wavelet-based approaches to enhance classification accuracy (Myint and Mesev, 2012).

1.3 Exploiting VHR Data

Although VHR sensors may offer advantages for mapping urban areas, to-date their wide use has been hindered. The principal reason limiting the usage of

VHR imagery is the high cost (Lei *et al*., 2014; Franny *et al*., 2015; Chan, 2015; Kuffer *et al*., 2016). According to Land Info (2016), archive and new tasking of WorldView-2/3 imagery costs US$19 and US$29 per km^2 respectively, and other processed data products may cost several times more. This means that while a Landsat image (covering an area of 3425 km^2; 185 x 185 km) has no cost, WV-2 imagery for the same area costs at least $65075 (US$19 x 3425). The Landsat programme offers images with a 30m resolution, which are freely available via GloVis, EarthExplorer, or via the LandsatLook Viewer websites within 24 hours of reception. Furthermore, the French government has announced that SPOT (Pour l'Observation de la Terre) (with 10m spatial resolution) archive imagery (older than five years) would be accessible freely for non-commercial users (Selding, 2014). Costs would increase if a temporal record is required and the likelihood of attaining that is much reduced with VHR sensors. Furthermore, historical image archives are much shorter when using VHR data; IKONOS as the first very fine resolution optical satellite has only been available since 1999 (De Laet *et al.,* 2007), while the first sensor of the Landsat programme was launched in 1972. Also, the temporal resolution of VHR sensors are limited compared to the medium resolution sensors (i.e., Landsat). For example, while it takes Landsat 16 days to complete a full earth scan, the revisit time for VHR sensor is significantly higher (i.e., 144 for IKONOS). This can be especially problematic for VHR sensors for monitoring applications. Finally, a full scene of VHR image is significantly smaller in spatial extent compared to medium resolution imagery. For example, a full scene of Landsat imagery covers an area of 185 x 185 km on the ground, while this coverage is 16.4 x 16.4

km for WV-2. As a result, there are several WV-2 images are required provide a similar coverage to a single Landsat image. This leads to increasing the costs of projects when considering the use of VHR imagery.

This dichotomy between cost and spatial resolution (and associated temporal resolution constraints) underpins the central premise of this thesis. This thesis explores how medium resolution data (such as Landsat sensors) could be exploited for urban land cover mapping in a way that allows the full archive of the data to be used. The approach investigated here is to adopt some form of spatial resolution enhancement method. Methods based on image sharpening, especially if the sensor operates at more than one spatial resolution (Mather, 2004) or super-resolution analyses (Ling *et al.,* 2010) are ways forward. The training database in conventional multi-images is typically required to encompass a significant number of combinations of the high and low-resolution image (Jiji and Chaudhuri, 2006; Kasetkasem, Arora, Varshney, 2005; Yang *et al.,* 2012). These conventional super-resolution analysis methods may meet the requirement for general image processing applications (i.e., images from CCTV for face recognition applications). However, since the majority of remotely sensed sensors have long revisit periods, it is a challenging task to obtain a series of images during a short period. Therefore, rather than using a multi-frame approach, the single-frame method is more typically practised for remote sensing (Atkinson *et al.,* 1997; Boucher and Kyriakidis, 2006; Powell *et al.,* 2007; Ge *et al.,* 2009). These methods are numerically limited only to small increases in the resolution, usually by factors smaller than 2 (Baker and Kanade, 2002; Glasner *et al.,* 2009; Peleg and Elad, 2014).

This thesis explores the use of a new super-resolution approach called sparse representation as the basis for improving the spatial resolution of medium resolution data as a precursor for their utility in land cover mapping of urban areas. The key issue here is that VHR data are used in conjunction with a sparse representation algorithm applied to the medium resolution imagery to produce a new image with the spatial resolution of the VHR data but with geographical and temporal coverage beyond that of the VHR data. This keeps costs down especially for studying of large area over long time periods.

1.4 Research objectives

The principal aim of this research is to determine how land cover mapping of urban areas using medium resolution remotely sensed imagery could be enhanced to increase spatial and/or thematic accuracy. To achieve this goal, the work is divided into two main sections.

In the first section, three influential factors on classification accuracy will be evaluated for the extent of an entire urban area. These factors include image properties (specifically spatial resolution and spectral characteristics) and classifier type.

Having established the requirement for increased spatial resolution relative to Landsat data (of medium spatial resolution) for land cover mapping of urban areas, the second part of this thesis focuses on enhancing the spatial resolution of original, medium resolution remotely sensed data to be classified as land cover. Landsat sensor data will be the focus of the investigation, and the method of sparse representation will be explored in two ways. First, sparse

representation will be applied directly to the original resolution of the multispectral Landsat sensor data; and second, sparse representation will be applied on pan-sharpened Landsat data. This sparse representation work has two steps. The initial step will demonstrate how much the accuracy of resultant land cover classification is improved, and the final step will explore how flexible the sparse representation method is and how it could be used to exploit the Landsat archive more fully. The objectives of this thesis are therefore to:

- Establish the importance of enhancing the spatial resolution of remotely sensed data for land cover mapping of an urban area with respect to spectral resolution and classifier type;
- Illustrate the effectiveness of sparse representation as a method of enhancing the spatial resolution of Landsat data;
- Determine parameters around the use of the sparse representation method.

1.5 Outline

The outline of the thesis is illustrated in Figure 1.1. Following this introductory chapter (1), there are two background (literature review) chapters (2 and 4), three analysis chapters (3, 5 and 6), and a closing discussion and conclusions chapter (7). Chapter 2, the first literature review chapter, elaborates the importance of three factors which are influential in the classification of the urban area: spatial resolution, spectral characteristics, and classification approach.

Chapter 3 presents a study to evaluate how the three factors presented in Chapter 2 – spatial resolution, the number of spectral bands, classification approach – influence land cover classification accuracy in an urban environment. In total, 36 land cover maps were generated by performing maximum likelihood (ML), support vector machine (SVM), and OBIA\NN classification on twelve data sets of Nottingham formed by a combination of four spatial resolutions (2m, 4m, 10m, and 30m) and three spectral band sets (four bands, six bands, and eight bands).

Chapter 4 has two main parts. The first part provides an overview of super-resolution techniques, starting with a critical evaluation of conventional super-resolution methods and ending with an appraisal of the contemporary sparse representation approach. The second part outlines the methodology developed here to employ sparse representation for enhancing Landsat image data in urban land cover mapping.

Chapter 5, the second analysis chapter, explores the potential of the sparse representation method to enhance the capability of Landsat OLI imagery of Nottingham, UK. Sparse representation analysis will be conducted on both original, multispectral Landsat OLI imagery and pan-sharpened OLI imagery. Subsequently, the spatially enhanced OLI imagery will be classified using both pixel-based and object-based SVM approaches.

Chapter 6, the final analysis chapter, assesses the potential of the sparse representation method to enhance the capability of Landsat data when there is a spatial and temporal separation between the training image and Landsat datasests.

Chapter 7 provides a comprehensive discussion of the results and research achievements. This chapter considers the original research questions and indicates the extent to which the thesis objectives have been achieved. Then, future work opportunities are outlined and, finally, a conclusions section is presented which summarises this doctoral research projects, its key findings and highlights the contribution to the remote sensing discipline.

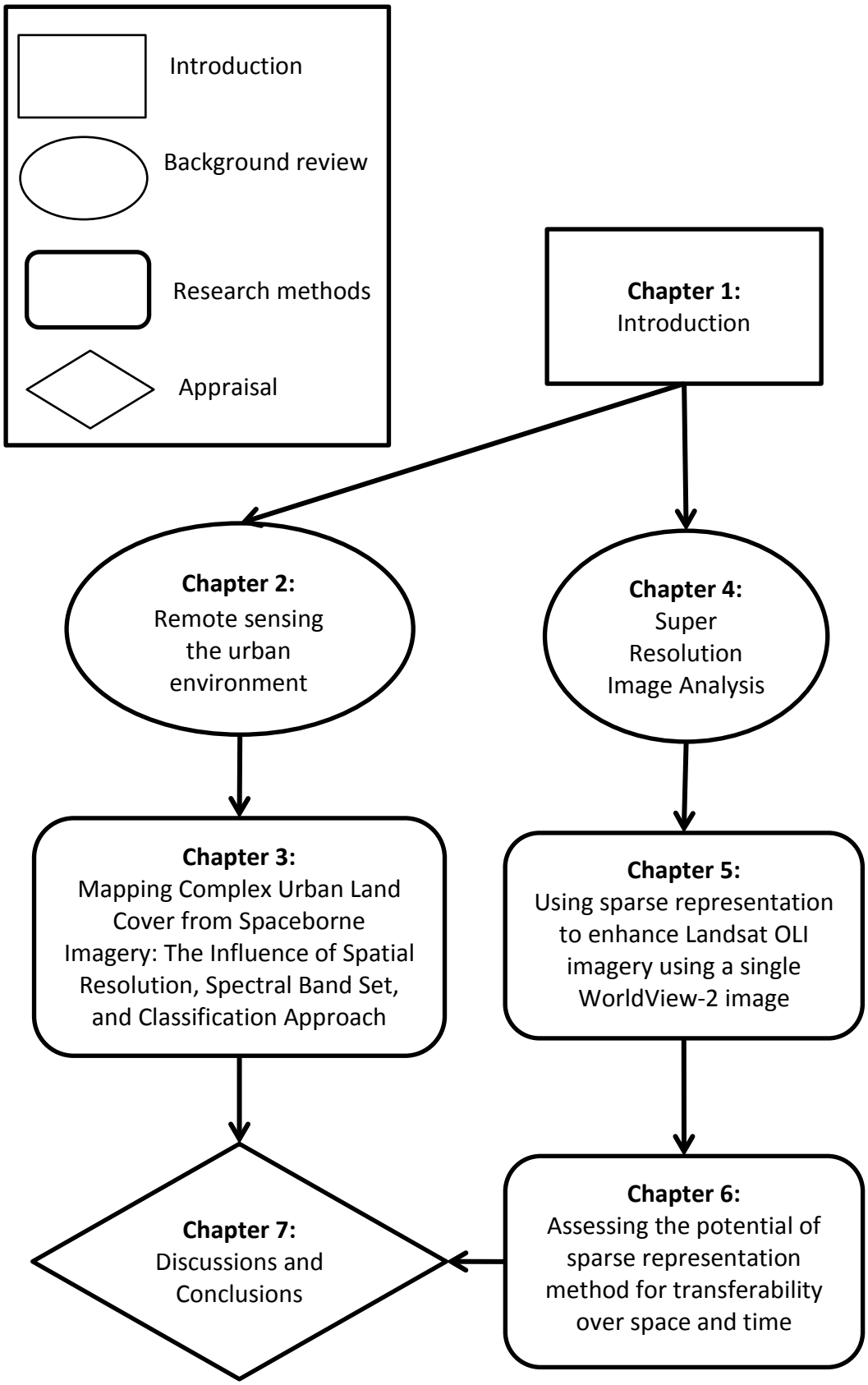


Figure 1.1. The structure of thesis.

Chapter 2: Remote Sensing and Urban Environment

2.1 Overview

According to the United Nations, the twenty-first century is the first urban century (Weng and Quattrochi, 2007). Due to shifting population from rural areas, the population of urban areas increases, leading to urban growth. From an environmental perspective, the dynamic process of urban growth causes a fundamental alteration to the functions, structure and dynamics of environmental systems at a wide range of scales (Bihamta *et al.,* 2015; Liu and Phinn, 2010). Due to its capability to monitor natural resources and urban assets, remote sensing technology has been used for urban mapping applications. However, early such studies were limited by the relatively coarse spatial resolution of available satellite sensors. The relatively recent availability of VHR imagery focuses attention on remote sensing for more accurate analysis of urban features. Further, compared to first and second generation optical satellite sensors (i.e., Landsat sensor products with 30m resolution), the newer sensors offer multispectral imagery at finer spatial and spectral resolution (i.e., eight spectral bands of WV-2 (WorldView-2) with 2m resolution). Moreover, recent advances in remote sensing have not been limited by the capabilities of the sensors. In addition, more advanced classification algorithms have been developed to support the use of remotely sensed data.

This chapter reviews the current literature pertaining to the use of remote sensing for the study of urban areas. In particular, the influences of spatial

resolution will be considered, in conjunction with advancements with spectral capabilities, as well as the classification routines available for mapping urban land cover. The first part of this Chapter outlines the phenomenon of urbanisation and its potential environmental threats. Then, remote sensing is introduced as a reliable technology to address the need for mapping and monitoring urban areas. Next, a review of previous urban studies using multispectral sensors with medium spatial resolution will be provided, while coarse resolution will be introduced as a possible constraint for mapping urban areas accurately. The availability of the most recent VHR sensors with finer spatial and spectral resolution will then be discussed, since these instruments may potentially increase the accuracy of urban mapping. Finally, advancements in classification approaches will be reviewed. This part of the Chapter highlights the choice of the classifier as an important factor in urban mapping accuracy.

2.2 Urbanisation and its environmental effects

Rising human population, industrialisation and growing interest in accessing urban facilities are causing more people to live in cities. Cities have become the centre for commercial activities, the host for most global manufacturing, and are popular residential zones. Urbanisation in England has increased significantly since the middle of last century, by approximately 7% (Perry and Nawaz, 2008). In China, urbanisation has occurred at a fast rate, whereby the proportion of urban population increased from 20% in 1982 to more than 50% in 2014 (United Nations, 2015). As a developing country in West Africa, the percentage of urban population in Ghana was only 15% in 1950; this figure

increased to 52% in 2010, and it is expected that 75% of Ghana's population will live in urban areas by 2050 (Coulter *et al.,* 2016). Urbanisation as a global phenomenon may be beneficial for many city dwellers, but there is a direct relationship between fast urbanisation and environmental impacts at regional and global scale (Douglas, 2002; Ban *et al.,* 2010; Hu *et al.,* 2016).

Rapid global urbanisation can lead to environmental problems such as deforestation, climate change (Yang, 2011; Elsayed, 2012), hydrology fragmentation (Mancipe-Munoz *et al.,* 2014), and biogeochemistry changes across a range of local, regional and global scales (Grimm *et al.,* 2008). According to Angel *et al. (*2011), increasing urban land cover by 10% in a country can increase the country's total CO_2 Emissions by more than 11%. As one of the consequences of urban growth, the amount and distribution of impervious surfaces (i.e., building's roofs, roads, pavements) increases which deteriorate hydrological status, including water chemistry and biological conditions (Paul and Meyer, 2001). For example, the quality of runoff can be deteriorated through a high percentage of impervious surfaces (Arnold Jr and Gibson, 1996). Particularly, buildings' roof runoff in urban areas is one of the most substantial pollutants within stream and rivers (Gobel *et al.,* 2007; Yusop *et al.,* 2007). For instance, according to Eglimanand Menendez, (2001) and Kanarek, (2001) asbestos rooftops can cause Mesothelioma cancer.

2.3 Urban remote sensing

Monitoring the urban structure can help to identify the relevant trends and threats to both urban and global ecosystems. This information can contribute

to understanding the status of urban dynamics correctly which can potentially alleviate the effects of urbanisation on the environment (Aburas *et al.,* 2016; Dahiya, 2016). Also, knowledge of urban structures and distributions are key elements to determine the use of urban energy and emissions (Frolking *et al.,* 2013), to aid ecosystem protection, and to enable sustainable development of urban areas (Yao *et al.,* 2016).

Traditionally, aerial photograph interpretation and field surveys were popular methods to acquire information on mapping ground features within urban areas. Unfortunately, these two methods are time-consuming, costly, and unable to provide complete coverage of large areas (Gounaridis and Koukoulas, 2016). However, satellite observations of the Earth's surface can be used to document urban features over large areas (Bhaskaran *et al.,* 2010). In fact, remote sensing technology is a suitable option for this task, by providing detailed spatial information for environmental monitoring applications in urban areas (Blaschke, 2010).

Remote sensing has been recognised as a reliable source to derive information on the urban environment in a timely and detailed manner, aiding planning and management activities (Yang, 2007). Several groups benefit from using remote sensing technology to map urban areas. Urban researchers employ remote sensing to extract urban structure information for studying urban geometry. This can help with the development of theories and models of urban morphology. Remote sensing can derive urban land cover information as the primary boundary condition used in many spatially distributed models within

the broad scope of environmental studies. Further, remote sensing offers a practical and economical means to study change within urban areas frequently.

2.3.1 Urban remote sensing challenges

Mapping urban land does, however, pose a significant challenge for remote sensing due to the prevailing high spatial and spectral variability, coupled with the irregular size, shape, and orientation of objects (Nichol, 2009). Typically, urban surfaces comprise a wide range of different land cover types which can be summarised in four main categories: impervious surfaces; soil; wetlands (i.e., lakes, pool, and rivers); and vegetated areas (i.e., grassland and other tree and shrub species). The variety of landscape elements in urban areas can create spectral confusion among different vegetation types (i.e., various trees species) and human-made features (i.e., various roof types) (Hamedianfar, 2015). Another characteristic of urban environments is their dynamism, whereby some human-made structures and green spaces undergo quick physical changes. Shadowing is another common challenge to remote sensing in urban areas due to the presence of tall buildings and large trees (Priem, and Canters, 2016). Seasonal variation presents another challenge for urban remote sensing, especially for mapping vegetation features (Voss and Sugumaran 2008). Despite these challenges, remote sensing is the only practical and cost-effective technology to acquire a unique abridged view of the Earth's environment without labour intensive and exhaustive field surveys (Taubenbock *et al.,* 2010; Xu, 2013). Remotely sensed data is suitable for monitoring applications due to the frequent data acquisition. Further, due to

the digital nature of remote sensing data, image products can be integrated with other forms of digital spatial data for integrated analysis. As a successful method of acquiring geospatial data, remote sensing technology has been applied to many urban applications including: green space monitoring (Rougier *et al.,* 2016; Hashemi, 2016), urban creep (Allitt, 2009), urban growth (Li *et al.,* 2014, Megahed *et al.,* 2015; Li *et al.,* 2015), impervious mapping (Zhang *et al.,* 2016), urban heat islands (Chen *et al.,* 2016), urban water management (Li *et al.,* 2016), and evaluation of human health (Gong *et al.,* 2012).

2.4 Image spatial resolution

2.4.1 Medium spatial resolution multispectral imagery

Land cover classification represents one of the most common forms of remote sensing image analysis (Franklin and Wulder, 2002; Alvarez *et al.,* 2003). This process depicts the natural and anthropogenic objects that can be observed on the Earth's surface like grass, trees, soil, buildings' roof types, roads, and water (Holland *et al.,* 2006). Land cover mapping has become significant for national plans to resolve haphazard urbanisation issues such as uncontrolled development, deteriorating environmental quality, and loss of prime agricultural lands. As an intrinsic element, land cover is used for most of the remote sensing analysis (Aplin, 2004).

Since 1972, with the availability of the Landsat MSS (Multispectral Scanner System) sensor, remote sensing started to receive attention as a mapping tool for urban areas. By advancing the technology, newer models of Landsat have become operational, such as TM (Thematic Mapper), ETM (Enhanced Thematic

Mapper), ETM+ (Enhanced Thematic Mapper plus), and OLI (Operational Land Imager). While the most recent versions of Landsat (i.e., ETM+ and OLI) offer a panchromatic band with 15m spatial resolution, the spatial resolution of the multispectral bands remains 30m. Beside Landsat products, the French satellite SPOT became available from 1986. SPOT-1 was the first of a series of SPOT satellite sensors offering multispectral imagery with 20m spatial resolution, and panchromatic imagery with 10m spatial resolution. The advancement in SPOT sensors continued, and the next generation of SPOT offered imagery with finer spatial resolution: 10m and 5m respectively for multispectral and panchromatic imagery.

Regarding previous studies on urban mapping using medium resolution multispectral data sets, Guindon *et al.* (2004) tested three Landsat products (Landsat MSS, Landsat TM, and Landsat ETM+) to map the urban area of Ottawa in Canada. The entire study area was categorised into five classes: water, forest, herbaceous, residential, and commercial/industrial. For example, the producer accuracies were obtained for 78% and 73% for residential and commercial/industrial classes, respectively. Wu, (2004) was unsuccessful in mapping the metropolitan region of Columbus Ohio in the US according to the VIS (vegetation, impervious surfaces, and soil) model using Landsat ETM+. An overall accuracy of 45% was achieved.

Although spaceborne remote sensing has been employed for urban land cover classification over several decades, early work was limited by the relatively coarse spatial resolution of available sensors, perhaps most commonly Landsat Thematic Mapper (TM) and its 30m resolution multispectral imagery (Guindon

et al., 2004; Small and Lu, 2006). Medium resolution multispectral images are suitable to provide generalised land cover classes (Somg *et al.,* 2005; Chen *et al.,* 2007). The level of spatial information offered by Landsat products is insufficient to detect small urban features which create significant challenges for mapping detailed urban areas accurately (Small and Lu, 2006; Zhou and Troy, 2008). The inadequate level of spatial detail in medium resolution multispectral imagery for detailed urban mapping potentially leads to land cover mixing within the pixels (Guindon *et al.,* 2004). Therefore, applying remote sensing to answer urban questions has been limited (Hu *et al.,* 2016).

2.4.2 Fine spatial resolution multispectral imagery

Detailed land cover information is crucial for mapping and managing complex urban environments across local and regional scales (Lu and Weng, 2007; Zhou and Qiu, 2015). For example, urban and landscape designers require detailed inventories of urban vegetation to reach informed decisions about conservation, planting, and maintenance of urban trees. Hu *et al.* (2016) and Zhou and Troy, (2008) highlighted the need for using fine resolution data sets to produce more accurate land cover maps for urban management. Advances in remote sensing technology have led to the availability of VHR imagery and the development of more advanced digital processing and analysis techniques. Such finer resolution imagery can be acquired through airborne hyperspectral imagery and multispectral sensors. Several studies demonstrated the capability of airborne hyperspectral with a fine resolution (also is referred as H-resolution) to map urban features including Herold *et al.* (2004); Heiden *et al.*

(2007); Taherzadeh and Sharifi, (2011). However, the limited coverage and high cost of data acquisition are the key challenges to widespread use of the airborne hyperspectral data set (Hamedianfar, 2015).

On the other hand, advances in technology and the availability of finer resolution multispectral satellite sensor imagery (i.e. IKONOS, QuickBird, GeoEye-1, WV-2, and WV-3) (Table2.1) enables identification of small urban features, which contributes to accurately mapping urban areas (Hamedianfar and Shafri, 2015). Improving the sensor's capability to record data with finer spatial resolution has opened up new possibilities for remote sensors to identify and classify complex and heterogeneous urban environments (Weng, 2012). Several studies successfully used VHR imagery for various urban planning applications, including tree canopy mapping (Pu and Landry, 2012), green space analysis (Lang *et al.,* 2008), urban hydrologic modelling (Han and Burian, 2009), and urban land-use mapping (Herold *et al.,* 2003).

For example, fine resolution sensors, including QuickBird, IKONOS, and the most recent VHR sensors WV-2 and WV-3 capture ground features with an improved spatial resolution that can potentially enhance the accuracy of mapping detailed urban areas (Table 2.1).

Satellite sensor parameters	IKONOS	QuickBird	GeoEye-1	WV-2	WV-3
Launch date	September 1999	October 2001	September 2008	October 2009	August 2014
Design life per year	> 15	5	7	7.25	7.25
Expected life per year	>30	13	15	>1	>1
Nominal altitude (km)	681	482	770	770	620
Spatial resolution (m)	4	2.4	2	2	1.5
Spectral bands	blue, green, red, and NIR	blue, green, red, and NIR	blue, green, red, and NIR	coastal blue, blue, green, yellow, red, red edge, NIR1, and NIR2	coastal blue, blue, green, yellow, red, red edge, NIR1, and NIR2
Radiometric resolution	8 bit	11 bit	11 bit	11 bit	11 bit
Panchromatic band spatial resolution (m)	0.82	0.61	0.46	0.46	0.31
Swath width (km)	11	16.4	15.2	16.4	13

Table 2.1. VHR satellite sensor specifications for IKONOS, QuickBird, GeoEye-1, WV-2, and WV-3.

Aplin (2003) conducted a study to compare the accuracy of urban mapping using IKONOS with 4m spatial resolution and SPOT with 10m spatial resolution. The town of Arundel in West Sussex was classified into four classes: grassland, woodland, asphalt, and roof tiles. In this study, the advantage of using 4m data sets was demonstrated to map Arundel Town compared to the 10m data set. Pu and Landry, (2012) explored the potential of WV-2 to map seven urban tree species in the City of Tampa, Florida, USA. IKONOS imagery was compared to WV-2 to demonstrate the advantages of WV-2, with three different spectral

band sets including WV-2 with the conventional four spectral bands set, WV-2 with four new spectral bands (coastal blue, yellow, red edge, NIR2), and WV-2 with full eight spectral band set. The result showed significant improvement in mapping all tree species using all three combinations of WV-2 imagery. The overall average accuracy was increased by 16% to 18% using WV-2 imagery compared to IKONOS imagery. They suggested that this improvement in the mapping accuracy can be addressed by the finer spatial and an enhanced spectral data set of WV-2 imagery. As such, previous research suggests that VHR imagery, like WV-2 (with 2m spatial resolution and eight multispectral wavebands), can potentially increase the capability of remote sensing for conducting detailed studies of the Earth's surface (Lu and Weng, 2007; Blaschke, 2010). Therefore, using finer spatial resolution can improve the overall accuracy of mapping urban areas because of reducing per-pixel spectral heterogeneity (Myint *et al.,* 2011) and mixed-pixel problem as seen in traditional remote sensing (Lu and Weng, 2007).

2. 5 Image spectral characteristics

Although using imagery with finer spatial resolution can help to identify small features in urban areas, the task of urban mapping can still be a challenge. Typically, urban areas are formed from a wide range of intra-urban land covers, comprising artificial and natural surface materials. Some of these features of interest contain the same composition materials (i.e., human-made features or tree species) which lead to a high degree of spectral confusion (Myint *et al.,* 2011). Zhou and Troy (2008) and Hamedianfar and Shafri, (2014) identified the

within-class spectral diversity among urban features as a significant challenge to extract detailed urban features. Dealing with this complexity is particularly challenging because of insufficient knowledge about the spectral properties of urban materials (Ben-Dor *et al.,* 2001; Zhou and Troy 2008). The inadequate level of spectral information to deal with the spectral heterogeneity of urban materials can result in inaccurate mapping (Whiteside *et al.,* 2011; Blaschke *et al.,* 2014). Herold *et al.,* (2003) highlighted the limited number of four spectral bands as a possible constraint of the first generation of H-resolution multispectral sensors (i.e., IKONOS and QuickBird) for application to urban mapping.

The advancements in multispectral sensors were not limited to improvement in spatial resolution only, WV-2 and WV-3 offer four new spectral bands (i.e., coastal blue, yellow, red edge, and NIR2) (Table 2.1) compared to the four conventional spectral bands (blue, green, red, NIR). This enhancement in spatial and spectral capabilities can potentially enable detailed characterisation of very small urban features (Duque De Pinho *et al.,* 2012; Hamedianfar and Shafri, 2015).

2.6 Land cover classification

Notwithstanding these spectrally and spatially advanced satellite sensors, difficulties remain for urban mapping. For instance, selecting appropriate classification techniques is a crucial part of the urban land cover mapping.

Image classification is recognised widely among scientists and researchers as a common approach to categorise pixels based on the natural signature into predefined classes.

This chapter will review the most prevalent classification approaches currently used by the remote sensing community. Parametric hard pixel-based classification techniques, such as ML, will be discussed before exploring non-parametric pixel-based classification to address the unsuitability of statistical assumptions of parametric classification. The mixed pixel problem with be discussed before finally reviewing multi-stage object-based image analysis/Nearest Neighbour (OBIA/NN) approaches as a means of overcoming the pixel based classification limitations and drawing conclusions on the most suitable approach to be used in this study.

2.6.1 Pixel-based classification

Traditionally, the per-pixel concept was the dominant classification technique to derive a thematic land-cover map automatically from remotely sensed data (Gao *et al.,* 2004). Pixel-based classifications assign each pixel to a class based exclusively on the pixel number (Gao *et al.,* 2004). There are different types of approaches based on:

- Familiarisation with the training data (supervised and unsupervised classification),
- Whether output is a definitive decision about land cover class (hard and soft classification),
- Statistical distribution of data (parametric and non-parametric),

2.6.1.1 Hard classification

Traditionally, urban classification has been conducted using hard (also known as crisp) pixel-based classifiers, where each pixel is assigned to a single land cover class (Blaschke *et al.,* 2014; Poursanidis *et al.,* 2015). Hard-pixel based approaches are relatively straightforward in structure and regarding computational analysis, and over last four decades have been the most common classification technique used (Gao *et al.,* 2004). Hard classification assigns each pixel to the class that it most closely resembles, thus making the assumption that the pixel covers a homogenous land cover type (Schowengerdt, 1997). In fact, hard classification methods categorise each pixel into the class which has the greatest likelihood of being correct. The pixel can belong to one class and one class only, no matter whether the pixel is entirely homogenous or made up of varying percentages of other classes. For example, if 90% of a pixel is covered by vegetation, using a hard classifier may not significantly affect accuracy. However, if just 55% of the pixel comprises vegetation the accuracy of mapping these heterogeneous surfaces can reduce significantly. The accuracy of this approach is dependent upon spatial resolution, variance amongst classes and homogeneity of the land cover that the pixel is representing. These will be discussed fully in the following sections.

2.6.1.2 Unsupervised and supervised classification

The concept of using training data for the classifier introduces the theories of supervised and unsupervised classification. Unlike unsupervised classification supervised classification, requires external training from the user, who decides

which pixels belong to which class by providing a sample of the pixels from each class and thus training the classifier (Foody *et al.,* 2007).

Because unsupervised classification does not require any earlier knowledge about the area and human input, this technique is quicker, cheaper, and simpler than supervised methods (Goetz *et al.,* 2003). Further, the result of unsupervised approach may reveal some hidden characteristics of data, while the pixels in the same class may not necessarily illustrate identical features. Therefore, the result of the unsupervised technique does not necessarily correspond to the classes that users are interested in.

A common assumption of supervised analysis is that the classes of interest have been fully defined (Congalton and Green, 2009), but potential failure to fulfil this assumption involves errors in the classification procedure (Foody, 2002). Supervised approaches are more involved due to the challenge of capturing and identifying training data. Further, the quality and quantity of training data are a key consideration in a supervised approach. Goetz *et al.* (2003) have asserted that the difficulty in selecting training data is a primary factor which results in land cover misclassification.

Because supervised classification techniques predefine the classes of interest for the user, Goetz *et al.* (2003) and Shan *et al.* (2010) recommend this approach as more suitable and accurate than the unsupervised approach. Overall, many studies have demonstrated that supervised approaches have the potential for better accuracy than unsupervised approaches (Lu and Weng, 2007; Shan *et al.,* 2010).

2.6.1.3 Maximum likelihood classifier

Historically most pixel-based supervised classification analysis has employed statistical parametric classifiers such as the ML (Lu and Weng, 2004; Myint *et al.,* 2011; Qin *et al.,* 2015; Myint *et al.,* 2011; Qin *et al.,* 2015). Many researchers consider the ML classifier as a benchmark for evaluating their data sets and classification schemes (Lu and Weng, 2004; Yan *et al.,* 2006; Dixon and Candade, 2008; Erener and Duzgun, 2009; Kavzoglu and Colkesen, 2009; Otukei and Blaschke, 2010; Myint *et al.,*2011; Qin *et al.,* 2015). This supervised classifier is one of the most widely used classification techniques among the remote sensing community for per-pixel classification, and one which often produces acceptable results (Ippoliti-Ramilo *et al.,* 2003; Keuchel *et al.,* 2003; Pal and Mather, 2003; Bhaskaran *et al.,* 2010).

This supervised method measures the variances and covariances of the training sample classes and then each class is depicted as a covariance matrix and the mean vector, by assuming the training samples are normally distributed (Mather and Koch, 2011). The ML algorithm estimates the statistical (Bayesian) probability function of each class for the individual pixel based on the inputted training data. As a final step, every pixel is assigned to the class to which it most probably belongs.

Huang *et al.* (2002) stated that because of its straightforward algorithm structure the ML classifier is relatively straightforward for computational analysis. Although ML classification is a valid method, it makes the statistical assumption that the data are normally distributed (Swain and Davis, 1979), but the nature of VHR imagery can mean that it is difficult to honour these

assumptions. In fact, when using VHR remotely sensed data, the ML assumption regarding the normal distribution of data is often not met, especially in complex sites that include noise (Lu and Weng, 2007). In fact, because of urban complexity, failure to meet this assumption possibly leads to classification errors. Specifically, ML classification tends to work well where the training data are relatively "clean", such as where coarse spatial resolution imagery is used to classify general land cover classes. In situations where the training data are noisier, such as when finer spatial resolution imagery is used to map complex (i.e., urban environments), the ML classifier can be considerably less accurate (Myint *et al.,* 2011; Lu *et al.,* 2011). Finally, parametric methods have limitations to integrate with ancillary data including spatial and contextual attributes, and non-statistical information, into the classification process (Lu and Weng, 2007).

2.6.1.4 Support Vector Machine classifier

The assumption of normal data distribution in ML classification rarely occurs when VHR remotely sensed data is used especially in complex urban environments which include noise (Lu and Weng, 2007). In contrast to the parametric assumption, non-parametric methods do not rely on statistical assumptions to measure class separation. As a result, non-parametric classifiers may provide better classification results than parametric classifiers in complex landscapes (Foody, 2002). Further, according to Mather and Koch (2011), non-parametric classifiers are specifically appropriate for integration of non-remotely sensed data into the classification procedure. Therefore, non-parametric pixel-based classifiers such as SVM seem well-suited to VHR urban

classification since they are better able to handle noisy training data, compared to for instance the ML classifier (Mountrakis *et al.,* 2011; Poursanidis *et al.,* 2015).

SVM is as an advanced, non-parametric pixel-based classifier that has been recognized as one of the successful classifiers for data classification among the remote sensing community (Huang *et al.,* 2002; Pal and Mather, 2005; Boyd *et al.,* 2006; Chi *et al.,* 2008; Dixon and Candade, 2008; Foody, 2008; Kavzoglu and Colkesen, 2009; Otukei and Blaschke, 2010;). This operates based on statistical learning theory by grouping theoretically superior machine learning algorithms. The aim of these machine learning algorithms is to determine the position of decision boundaries that yield the optimal separation of classes by employing optimisation algorithms (Huang *et al.,* 2002).

This classifier has several advantages compared to other methods, such as the ML classifier. Firstly, because it is a non-parametric classifier, it can handle complex sites and can integrate non-parametric ancillary data (Rogan and Chen, 2004). According to Pal and Mather (2004), SVM could handle identification of nonlinear decision surfaces. In addition, SVM has low processing times in both training and running the classification (Foody and Arora, 1997). However, according to Huang *et al.* (2002), SVM as a competitive approach is comparable to other available machine learning algorithms, especially for labelling high-dimensional data sets. Moreover, due to independent dimensionality behaviour of SVM, it does not require as much training data compared to an ML classifier (Dixon and Candade, 2008). It is particularly advocated for use with high-dimensional data sets primarily

because it is believed that the Hughes effect does not constrain the decision making (Chi *et al.,* 2008; Tuia *et al.,* 2009; Plaza *et al.,* 2009). Although the effectiveness of SVM largely depends on the values of a few user-defined parameters, including data set size, kernel parameter setting and class separability, the use of the SVM as a classifier should benefit complex classification problems (for example where fine spatial resolution imagery is used to map detailed classification schemes in heterogeneous environments), and can perform better than ML classification for urban environments using VHR imagery (Pal and Mather, 2004; Dixon and Candade, 2008). Therefore, SVM seems well-suited to VHR urban classification since they are better able to handle noisy training data, compared to for instance the ML classifier (Mountrakis *et al.,* 2011; Poursanidis *et al.,* 2015).

2.6.1.5 Limitation of hard pixel-based classification

Because the majority of individual pixels represent more than a single land cover type, using hard pixel-based classification methods often leads to misclassification. Lu and Weng (2007) stated that the estimated areas by hard classification could generate significant errors, specifically by using coarse spatial resolution imagery.

Mapping complex urban areas using hard pixel-based classification approaches has faced challenges due to the mixed pixel problem (Aplin and Atkinson, 2001; Tatem *et al.,* 2003; Ling *et al.,* 2012), within-feature variation which occurs for VHR images (Lu *et al.,* 2010; Aplin and Smith, 2011; Liu *et al.,* 2013; Huang and Zhang, 2013), and noise which is the cause of the 'salt and pepper' issue (Aplin, 2003; Lu and Weng, 2007; Aplin and Smith, 2011). Another drawback of pixel-

based classifiers is the exclusion of spatial, textural, and topological relationships during the classification process (Bhaskaran *et al.,* 2010; Myint *et al.,* 2011).

The mixed pixels problem is the main drawback of hard pixel-based classifiers. One of the conditions which increases the chance of mixed pixels is to use coarser spatial resolution images for identifying the target features. Furthermore, using hard pixel-based classification over VHR imagery poses a new challenge to map different vegetation types in complex urban environments. For instance, the spectral response from a tree can be influenced by background objects and variation in canopy illumination, which can decrease the accuracy when utilising per-pixel based approaches (Quackenbush *et al.,* 2000).

2.6.1.6 Soft pixel based classification

As a type of pixel based classifier, soft (also known as fuzzy) classification has been proposed as a solution to the mixed pixel problem by allowing a pixel to have partial and multiple class memberships (Tatem *et al.,* 2003; Lillesand *et al.,* 2008; Ling *et al.* 2013; Holland and Aplin, 2013).

The soft pixel based classifier, developed in the 1990's, has become increasingly prevalent in the intervening years. Tso and Mather (2009) stated that the prime difference between soft and hard classifiers is the definition of the membership function. If the decision is multi-valued, that means each pixel can be assigned to more than one class. The classifier allocates each pixel by a membership grade for an individual class in which the membership grade is this value ranging between 0 and 1 (Campbell and Wynne, 2011; Mather and Koch,

2011). Several previous studies have proven that accuracy could be improved by using a soft classification compared to the hard method when using medium spatial resolution data (Foody, 1999). Artificial neural networks, fuzzy logic, decision tree, and genetic algorithms are several examples of soft classification techniques.

Often, the land cover map generated from this classification approach provides more meaningful information at the sub-pixel scale compared to hard per-pixel classification methods (Lu and Weng, 2007), although this approach does not deliver spatial information for land-cover within each pixel (Aplin and Atkinson, 2001). In other words, soft classification fails to account for the spatial distribution of class components within a pixel (Tatem *et al.,* 2003; Ling *et al.,* 2013; Holland and Aplin, 2013). Therefore, whereas soft classification conveys more realistic representation of land-cover than hard classification, the generated land-cover map from soft classification still contains a significant degree of spatial uncertainty.

2.6.2 Object-based image analysis

2.6.2.1 Moving from pixel to object

Most pixel-based classification techniques rely on spectral information but classifying urban features require additional features beyond spectral information using high-resolution imagery. This is because the size of a pixel (i.e., WV-2 with $4m^2$) is significantly less than the size of a building. Therefore, classifiying these areas using VHR imagery requires a classification approach which exploits spatial contextual information (Vatsavai, 2011).

Recent developments in remote sensing have not merely been limited to enhanced image resolution. Significant advancements on classifier algorithms have occurred over the past decades, especially with regard to shifts from pixel to segment, which has opened up new interest towards enhancing urban mapping. An image object is defined as a group of contiguous pixels sharing similar spectral and textural properties (Huang and Ni, 2008). Several researchers, including Flanders *et al.* (2003) and Platt and Rapoza (2008), have used this classification approach, operating at the spatial scale of objects, as a solution to the difficulties linked to non-contextual pixel-based classification. The idea to classify objects stems from the fact that most image data exhibits characteristic textures that are neglected in conventional classifications.

This method comprises two main steps including segmentation (detecting the edges of homogeneous objects) and then labelling each object to a pre-defined class. The fundamental assumption of OBIA is that relatively homogeneous patches construct each image and that the size of the object is larger than a pixel. In other words, OBIA operates by the grouping of neighbouring pixels that create objects or patches with some spatial coherence (Yan *et al.,* 2006; Lu and Weng, 2007; Bhaskaran *et al.,* 2010; Mather and Koch, 2011; Myint *et al.,* 2013). Pixels are grouped together into objects using spectral and spatial information similar to the way humans recognise colour, tone, texture and shape. These spatial and spectral properties can then be used to classify the land cover in the image by OBIA (Blaschke, 2010).

2.6.2.2 OBIA relationship with spatial resolution

Figure 2.1 indicates the suitability of classification approaches based on the relationship between objects (ground feature) and spatial resolution. In the coarse resolution imagery, the size of pixel is significantly larger than objects (figure2.1 (a)). The Figure 2.1 (b) describes the condition when pixel and objects have similar size. In this case, pixel-based classifiers are appropriate however, figure 2.1 (c) represents high-resolution pixels in which pixels are significantly smaller than the object. In this case individual objects formed from several pixels in which it is suitable for OBIA. The figure 2.1 indicates the OBIA is effective when VHR data are used, (referred to situation C) where pixels are significantly smaller than objects (Blaschke, 2010).

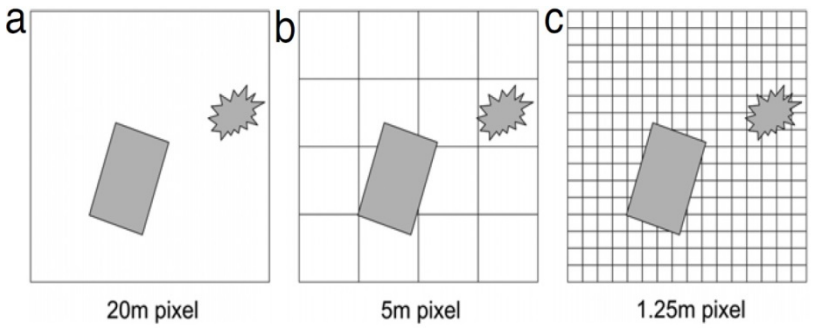


Figure 2.1. The relationship between objects and pixels as spatial resolution varies (Blaschke, 2010).

The development of OBIA has been linked closely with the emergence of VHR imagery (Lu and Weng, 2007; Blaschke, 2010; Cai and Liu, 2013).

2.6.2.3 Previous studies comparing pixel- and object-based classification

A large number of comparative studies have been conducted to evaluate the accuracy of OBIA relative to non-contextual pixel-based classification (Laliberte

et al., 2004; Chubey *et al.* ,2006; Cleve *et al.,* 2008, Platt and Rapoza, 2008; Lu *et al.,* 2011; Newman *et al.,* 2011). Pu *et al.* (2011) compared the accuracy of OBIA when mapping five vegetation classes (such as broadleaved, needle-leaved and palm trees) with IKONOS imagery in urban areas demonstrating the superiority of OBIA over pixel-based classification.

Another example of using OBIA is the creation of the UK Land Cover Map (LCM, 2007), which contains around 10 million objects. In this project, the Land Cover Map was produced by running OBIA on Landsat and SPOT imagery (a total of 37 images) and including generalised Ordnance Survey MasterMap vector map data in the segmentation process (Morton *et al.,* 2011).

Tehrany *et al.* (2014) compared the accuracy of decision tree pixel-based classification against OBIA on Selangor's urban areas (in Malaysia) using SPOT imagery with 10m spatial resolution. The evaluation of the results showed that the OBIA improved the total accuracy from 61.4% in tree pixel-based classification to 94.4% using OBIA. Beside impervious features, other studies reported that the OBIA can potentially improve the accuracy of mapping urban green areas (Yu *et al.,* 2006; Lu and Weng, 2007; Blaschke, 2010; Myint *et al.,* 2011, Kim *et al.,* 2011; Pu *et al.,* 2011; Rasmussen *et al.,* 2011).

2.6.2.4 Pros and cons of OBIA

OBIA has been explored with the premise that this approach directly addresses, and to an extent overcomes, the problem of within-feature variation and the mixed-pixel problem of hard pixel-based classification (Lu and Weng, 2007; Zhou *et al.,* 2008; Li and Shao, 2014). As a consequence, compared to pixel-based approaches, using OBIA can potentially increase the accuracy of mapping

complex urban areas (Dean and Smith, 2003; Yan *et al.,* 2006; Thomas *et al.,* 2003a; Blaschke, 2010; Myint *et al.,* 2011).

OBIA incorporates textural information (texture, shape, and size) besides spectral information in the classification process (Walker and Blaschke, 2008; Bhaskaran *et al.,* 2010; Blaschke *et al.,* 2011). Many researchers believe that adding the textural information can potentially improve the accuracy of land cover mapping (Blaschke and Strobl 2001; Zhou and Robson 2001; Dean and Smith 2003; Pizzolato and Haertel 2003; Thomas *et al.,* 2003b; Liu and Xia, 2010; Myint *et al.,* 2011). For example, OBIA may be able to distinguish between an urban park and a rural field even though both areas may have identical spectral characteristics because their surrounding areas differ. In addition, because image objects are similar to humans' conceptual organisation of the landscape, OBIA offers a more realistic representation of real world by considering the object instead of the pixel as their concepts of spatial extent (Hay and Castilla, 2006; Aplin and Smith, 2011). Furthermore, the output map generated by OBIA techniques can be used within a GIS which is preferable for many users.

Although OBIA has some advantages over pixel-based methods, it contains several challenges too. The major constraint of this approach is to select appropriate segmentation parameters which can have negative effects on the accuracy of results (Benz *et al.,* 2004). The segmentation elements are defined by human visual interpretation, which assesses the size, shape, and texture of sights, areas or objects; this requires significant effort and time and may not be

highly accurate when considering the limitation of human eyesight and the subjectivity of interpretation of features by individuals.

Two types of errors often exist in image segmentation, over-segmentation, and under-segmentation (Liu and Xia, 2010). Over-segmentation occurs when one semantic object is partitioned into multiple smaller image objects, while under-segmentation occurs when different semantic objects are grouped into one large image object (Liu and Xia, 2010).

Features extracted from mis-segmented image objects with over-segmentation or under-segmentation errors do not represent the properties of real objects on the Earth's surface (i.e. shape and area), so they may not be useful and could even reduce classification accuracy if not chosen appropriately (Song *et al.,* 2005; Kim *et al.,* 2011; Myint *et al.,* 2013).

Regarding over-segmentation, each over-segmented image object ideally corresponds to one single class, so it is potentially possible to classify all pixels in an over-segmented image object to their true class. That means the highest possible accuracy that can be achieved in the subsequent classification is 100%. However, because each under-segmented image object overlaps with multiple classes, it is impossible to classify all image objects to their true classes, but the image object is classified into only a single class. As a consequence, without consideration of object classifier, the highest possible classification accuracy that can be achieved for each under-segmented image object is always less than 100%. In this sense, under-segmentation necessarily introduces errors to the subsequent classification, whereas over-segmentation may not (Liu and Xia, 2010).

2.6.2.5 Multi-scale/stage OBIA

Due to the complexity of urban areas, fulfilling accurate segmentation is a challenging task. The urban area is a complex environment which involves the mixture of various objects of different shapes, sizes, and types (i.e., single trees, large grass areas, and shadows from buildings and other objects). The scale parameter (especially for eCognition Software) determines the object's size (according to spatial heterogeneity) whereby selecting larger scale values leads to formation of larger size objects. To achieve the most realistic and accurate segmentation, it is important that each segment (also object) represent its corresponding feature. For example, the size of segmentation required for lake features is larger than the size of segmentation required for house features. The common approach for OBIA is to allocate a single scale which fits best to all objects within the image. However, using a single scale parameter for an entire image can potentially lead to over- and under-segmentation issues which potentially reduce the accuracy of segmentation and ultimately the final map.

Because single scale procedures can not always achieve balanced segmentation outcomes for all classes of interest, multi-stage OBIA (also known as multi-scale) have been introduced to minimise the occurrence of over/under segmentation for mapping urban areas (Liu *et al.,* 2013; Piotr *et al.,* 2014). For example, specific scale parameters may be suitable for certain classes (i.e., large areas of grassland), but lead to considerable under/over segmentation of other classes (i.e., buildings). Since multi-stage OBIA allows

different parameter settings for different classes, this can achieve optimum classification outcomes for all classes (Karl and Maurer, 2010).

Advances on sensor resolutions and classification approach have opened a new window for mapping urban areas. Therefore, it is necessary to conduct a comprehensive study to test how these recent improvements in sensor resolutions and analysis methods, enhance the performance of mapping urban areas. The next chapter will assess the influence of varying spatial resolution and spectral band set and classification approach on the accuracy of detailed urban mapping.

Chapter 3: Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set, and Classification Approach

3.1 Overview

Warner *et al.* (2009) stated that the resolution requirements of remotely sensed data should be determined by a project's specific goals. Often, however, compromises are made. This chapter presents work undertaken to demonstrate the influence of spatial resolution on the accuracy of land cover mapping of urban areas. By building on our theoretical understanding of how image and classifier characteristics influence land cover accuracy, an exhaustive and rigorous practical experiment to compare the influence of spatial resolution, spectral band set and classification approach for mapping complex urban environments is presented. This is a departure from the prevailing literature in that, despite the significant advances in sensor technologies and classification procedures, studies of land over-classification of urban areas studies have often tended to be limited in their scope, perhaps comparing only one or two variables (i.e., spatial resolution or classification approach); and/or adopting general and unambitious thematic classification scheme; and/or working with small image data sets. For instance, many studies have conducted fairly basic, procedural comparisons between pixel-based and object-based classification (Myint *et al*., 2011; Aplin and Smith, 2011; Liu *et al*.,

2013), but these have not necessarily considered other influential variables such as input spatial resolution or spectral bands, or classification algorithm. Some practitioners have (sensibly) adopted VHR imagery, including WV-2 data with its enhanced spectral properties, for mapping urban environments, however commonly these attempt only broad differentiation between a few general land cover classes (Bouziani *et al*., 2010; Jebur *et al*., 2014; Ziaei *et al*., 2014) rather than detailed distinction of many specific classes. In effect, these studies seem content to replicate the sort of classification scheme used for decades with medium spatial resolution imagery, rather than attempting to exploit the full information content of VHR imagery and create very thematically-detailed urban land cover maps. Also, most test VHR data sets tend to be very small, often only a few hectares (Hamedianfar *et al*., 2014), and this can limit the strength of the scientific findings since there is no consideration of spatial extrapolation or transferability of the approach. That is, the results may be very parochial and depend strongly on local context.

In meeting the first objective of the thesis, this chapter provides a rigorous test (through the generation of 36 land cover maps with comprehensive accuracy assessment and the McNemar test) of the latest VHR imagery (WV-2) for urban classification, demonstrating how its components contribute to output mapping accuracy. With a particular focus on spatial resolution, the analysis is iterated using a series of different spatial resolutions and spectral band sets to simulate imagery ranging from traditional medium spatial resolution satellite sensors such as Landsat TM to state-of-the-art VHR sensors like WV-2, adapting an approach developed by Aplin (2003). Moreover, given the influential role

that classification approach plays on output accuracy, and how this is linked intrinsically with image specifications, all image data sets are classified using parametric and non-parametric pixel-based, and object-based, classifiers. Unlike earlier work, this study adopts a detailed thematic resolution, including many specific land cover classes rather than few general categories. This enables a fuller and more robust assessment of the WV-2 data, but also delivers helpful, practical information for urban planners and other user communities on the level of thematic detail that can be achieved when mapping complex urban environments from VHR data. Finally, the analysis is conducted using a relatively large image covering approximately $121\mathrm{km}^2$ of the city of Nottingham, UK, and its environs. This means that urban land cover information is generated at a scale of practical value and relevance (the whole city-scale), unlike many earlier experiments that have been limited to very small, local areas.

This Chapter is organised as follows: Section two explains the research material including the study area, data and classification scheme used. Section three highlights specific research methods by explaining pre-processing steps, spectral indices, and reference data sets. Section four depicts the classification procedure including pixel-based training, pixel-based classification, and OBIA. Section five describes the accuracy assessment and statistical test procedures carried out on the land-cover maps. Section six presents and discusses the results. Finally, section seven summarises this chapter.

3.2 Research materials

3.2.1. Study area

The city of Nottingham was used as a study area for this study. Nottingham is located to the North East of the East Midlands of England in the UK (Figure 3.1). It covers approximately 121 km^2 and with a longitude of 1° 14‘ to 1° 40‘ E and latitude of 52° 54‘ to 53°. The topography is fairly flat, with altitude generally around 100m. With fairly flat topography, the ground elevation in the Nottingham City barely exceeds more than 200m. With an estimated population of 310,840 in 2011, Nottingham was calculated to be the eighth largest urban area in the UK’s 2011 Census (Census, 2011). Nottingham is a relatively typical UK city, in that it comprises a mixture of residential, industrial and commercial land use, and therefore represents a good test for urban land cover mapping methodologies. The land cover can be broadly categorised into various types of anthropogenic features (i.e., asphalt, concrete, roof materials) intersecting with the semi-natural environment (i.e., vegetation (grass, trees), bare soil and water). While its city centre is dense with more man-made structures, the suburbs contain more vegetation and dispersed impermeable surfaces.



Figure 3.1. Nottingham, UK study area location and WorldView-2 image (© DigitalGlobe, Inc. All Rights Reserved).

3.2.2 Data and classification scheme

A WorldView-2 (WV-2) imagery of the city of Nottingham was captured on 26 June 2014 at 11:48 am local time and supplied in 11 bit data format (digital number range from 0 to 2047), at a spatial resolution of 2 m and with eight spectral wavebands: coastal, blue, green, yellow, red, red edge, NIR1, and NIR2 (Figure 3.2).

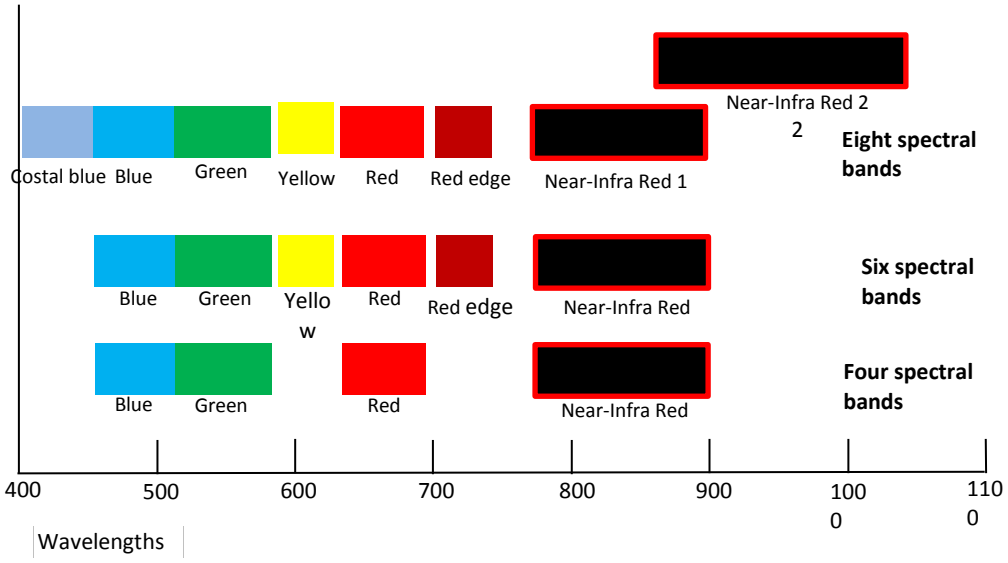


Figure 3.2. WV-2 spectral wavebands (top line) and four and six spectral bands data sets used for comparative analysis.

The original WV-2 image with 2m spatial resolution and eight spectral bands was modified to create twelve data sets with varying spatial and spectral properties for comparative classification analysis. First, the imagery was degraded successively to a series of coarser spatial resolutions: 4m, 10m, and 30m using Nearest Neighbour re-sampling technique used to degrade the WV-2 imagery with 2m resolution to 4, 10, and 30m. The advantage of nearest

neighbour compared to other re-resampling techniques (i.e., bilinear and cubic) is that it keeps the original pixels' values. These particular values were chosen to approximate the spatial properties of commonly used satellite sensors, ranging from state-of-the-art very fine resolution imagery to traditional medium resolution imagery. For instance, while 2m represents WV-2, 4m matches earlier VHR imagery from IKONOS, 10m matches the new Sentinel-2 MultiSpectral Instrument (MSI), and 30m matches Landsat TM or OLI.

Second, two additional spectral band subsets were created from the eight original bands. This was a simple process that just involved deselecting spectral bands as required; a four-bands subset was created using the blue, green, red and NIR1 bands, and a six-band subset was created using these four plus the red edge and NIR2 bands (Figure 3.2).

Again, the aim here was to compare a range of spectral band sets, and where possible approximate the spectral properties of commonly used satellite sensors. The original eight bands WV-2 image represents state-of-the-art VHR remote sensing, but also shares some spectral innovations with other recently developed sensors. For instance, Landsat OLI, Sentinel-2 MSI, and RapidEye use certain novel bands, including, in common with WV-2, coastal (OLI) and red edge (RapidEye). The four-band subset represents a conventional and widely used visible/near infrared band set. For instance, other VHR sensors such as IKONOS and GeoEye-1 use these four bands; and many medium resolution sensors, including some early Landsat instruments, typically use three or four visible and near infrared bands. The six bands subset is less direct in matching

real-world sensors, but represents an intermediate step between four and eight band data sets, and also specifically targets spectral bands of value for characterising terrestrial features. In total, twelve spatial/spectral image combinations (four spatial resolutions, three spectral band sets) were used for comparative classification analysis (see Figure 3.3).

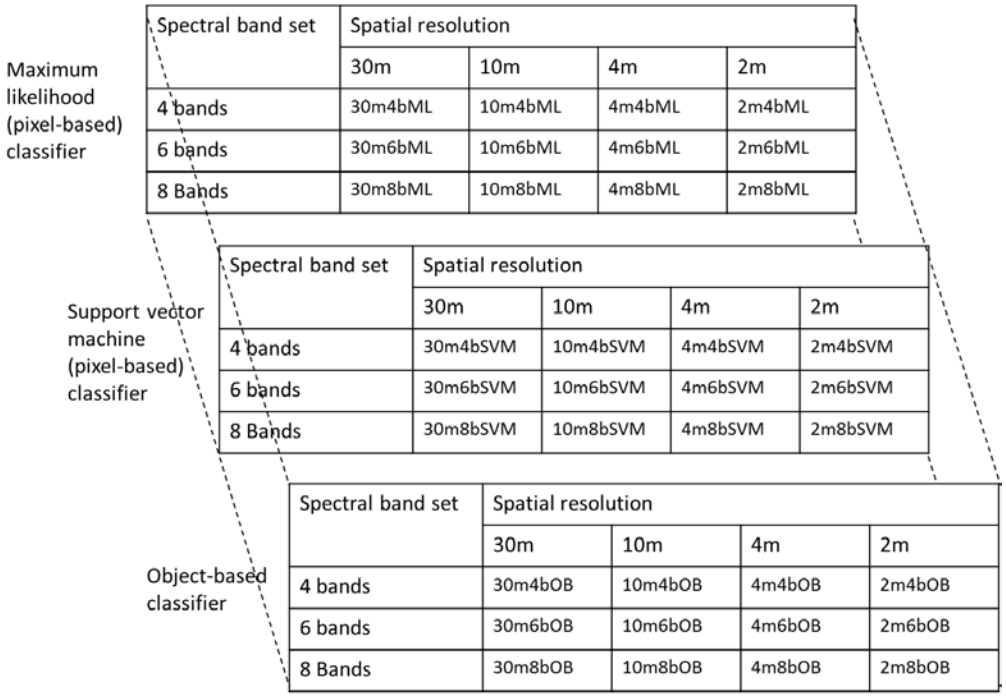


Figure 3.3. The 36 image data set/classifier combinations (4 spatial resolutions × 3 spectral band sets × 3 classifiers) used for comparative classification analysis.

The central urban core is more built-up and less vegetated than the outlying residential areas, though this varies considerably across the city and its districts. A classification scheme was developed to capture the detailed spatial heterogeneity of the urban land cover throughout Nottingham. In total, eleven classes were identified: asphalt, concrete roof, clay roof, slate roof, metal roof,

grass, broad-leaved trees, needle-leaved trees, bare soil, water and shadow (Table 3.1).

Class	Description
Asphalt	Urban ground surfaces covered in asphalt such as roads and car parks.
Concrete roof	Predominantly residential buildings covered in dark grey concrete tiles.
Clay roof	Predominantly residential buildings covered in red clay tiles.
Slate roof	Predominantly residential buildings covered in light grey slate tiles.
Metal roof	Predominantly industrial buildings covered in white metal panels.
Grass	Areas of grassland such as urban parks and lawns, plus surrounding rural agriculture.
Broad-leaved tree	Patches of broad-leaved deciduous trees.
Needle-leaved tree	Patches of needle-leaved evergreen trees.
Bare soil	Open areas covered by bare soil.
Water	Water bodies including lakes, rivers, ponds and canals.
Shadow	Areas of shadow cast from tall structures such as buildings and trees.

Table 3.1. Classification scheme of Nottingham urban land cover with eleven members,

3.3 Research methods

3.3.1 Pre-processing

Image pre-processing requirements were minimal for two reasons. First, a single source data set was used whereby all comparative outputs were derived from the original WV-2 image, and this meant that geometric distortion was of relatively little consequence. Nonetheless, the image's geometric fidelity was examined manually by cross-referencing the image with ancillary map data; geometric accuracy proved relatively high in general. Second, the analysis involved thematic classification and the accuracy of output land cover maps was assessed independently (of the original spectral imagery). Since all 12 data sets generated from a single imagery, the influence of atmospheric distortion

were of little consequence. Additionally, external sources of data sets through field survey and VHR online imagery (i.e., Google map) were used to create reference data set for training and assessment.

3.3.2 Spectral indices

The entire twelve data sets were supplemented with certain spectral indices, with the underlying intention to increase the accuracy of the resultant classifications. Specifically, the data sets with four and six spectral band sets were supplemented with a normalized difference vegetation index (NDVI, $\frac{(NIR1-red)}{(NIR1-red)}$) (Hamedianfar and Shafri, 2015) layer (so they in effect became five and seven band data sets, respectively). Those four data sets with eight spectral bands set were supplemented with NDVI, Normalized Difference Bare Soil Index (NDBSI, $\frac{(green-yellow)}{(green-yellow)}$) and Normalized Difference Brick Roof Index (NDBRI, $\frac{(yellow-green)}{(yellow+green)}$) layers (Zhou *et al*., 2012). Note, NDBSI and NDBRI could be calculated for the eight-band set, but not the four or six band sets, because only the full eight band sets included a yellow band. These particular spectral indices were added through trial-and-error whereby many indices were tested and these three proved useful to aid identification of vegetation, soil and roof classes. The indices were added initially at the object-based classification stage (described below), but to ensure a fair comparison between all classification analysis, the same input data layers (i.e., spectral bands plus indices) were used for all classifiers.

3.3.3 Reference data

Reference data were collected from a range of sources to create an independent data set for training and testing the classification analysis. Field land cover survey was conducted at locations throughout the study area in May 2013, matching the anniversary date of original image acquisition. Free online spatial data resources such as Google Street View and Bing Maps were used to supplement field survey (Gu *et al*., 2015; Tsutsumida and Comber, 2015), whereby secondary ground photographs and images were browsed to identify the land cover classes present at sample locations. Detailed vector map data—specifically MasterMap data (Survey, 2015) created by Ordnance Survey—were also consulted and cross-referenced with the imagery to gain a fuller appreciation of the land cover and land use present throughout the study area. Reference data sources were compiled and triangulated to create a comprehensive reference data set of land cover at locations throughout the study area. This data set was split into two parts; one used to create training class samples for classification analysis, and the other used to test the accuracy of the output land cover maps.

3.4 Land cover classification

Three different classification approaches were tested:

1. Maximum likelihood classification: a parametric pixel-based approach,
2. Support vector machine classification: a non-parametric pixel-based approach,

3. Object-based classification, using non-parametric nearest neighbour approach.

3.4.1 Pixel-based training

The first step in supervised land cover classification is generally class training. Selecting the most suitable training data is one of the essential steps in supervised classification procedure. The quality of training data has a significant role in determining the success of the classification (Congalton and Green, 2009). The training process was conducted for pixel-based classification, and this involved laborious trial-and-error, iterating training samples to optimise classification accuracy.

Mather and Koch (2011) recommended the minimum number of training samples per each class as 10ρ to 30ρ training samples per class, where ρ = a number of spectral bands. Since NIDVI was considered as an additional spectral band for data sets with four and six spectral bands, the training samples requirement ranges from 50 to 150 and 70 to 210 per class for data sets with four and six spectral bands, respectively. However, NDVI, NSDBI and NSDBRI were considered for data set with eight spectral bands leading the training samples requirement ranges from 110 to 330 per class. Since the classification scheme has eleven members, in this study, the minimum number of training samples for data sets with four, six, and eight spectral bands would be 550, 770, and 1210, respectively.

Also, training samples were selected randomly from locations throughout the study area, thereby avoiding any spatial bias that can be caused where training

samples are spatially clustered. Two different training sample strategies were employed since data sets with different spatial resolutions had different requirements. Firstly, a single pixel was used to represent each training point for the 4, 10, and 30m data sets. However, a block of 3 × 3 pixels was chosen to form each individual training sample for 2m spatial resolution data sets. A recent study by Holland and Aplin, (2013) proved the effectiveness of this approach for VHR data. Using a group of pixels rather than a single pixel can provide some representation of the natural variation within land cover structures and McCoy, (2005) explained that it may help exclude noisy and unrepresentative individual pixels. In total, for the 2m resolution image, 479 training samples were selected, each representing a block of nine (3 × 3) pixels, so 4311 pixels overall. This is well above the minimum requirement specified according to (Mather and Koch, 2011).

After completing class training for the 2m spatial resolution imagery, the process was repeated successively for data sets with 4m, 10m and 30m spatial resolution. The aim was to use the same or similar training points at the different resolutions to ensure direct comparability between results, though some slight modifications were necessary. First, because of the spatial averaging implicit to coarsening resolution, it was neither desirable nor possible to maintain 3 × 3 blocks of pixels as training samples, so individual pixels were used instead. As spatial resolution becomes coarser, pixels cover larger areas on the Earth's surface. So, individual pixels are less likely to represent very small, unrepresentative features. Also, as resolution coarsens, it becomes harder to identify homogenous training samples that extend over 3

× 3 pixels; for a 30m spatial resolution image, training samples would need to cover almost a hectare in size, and this is unlikely and uncommon in an urban environment. Second, for accurate classification results (where hard training, as opposed to soft or fuzzy training, is used), training classes should be pure, or as pure as possible. That is, each training sample should represent only its designated land cover class, not a mixture of classes. Clearly, as spatial resolution coarsens, it becomes harder to identify pure pixels as training samples since there is more pixel mixing in general. Here, to ensure training samples were as pure as possible, each original (2m imagery) training point was inspected to determine whether or not it represented a pure land cover class at the coarser spatial resolution. Only those samples that were deemed pure were retained for classification; others were discarded. This had the effect that the total number of training samples reduced successively at each coarser spatial resolution (4m resolution = 412 training samples, 10m = 299, 30 m = 254), meaning it was not always possible to achieve the recommended number. Nonetheless, relatively large samples were maintained for all classifications, and this approach enabled direct comparability between results. Further, various statistical tests were conducted to ensure the suitability of training classes.

Class training was conducted with care to investigate and ensure the spectral separability of classes. The spectral separability of classes was assessed using Transformed Divergence (TD) analysis which examines how well pairs of classes are statistically separated from each other. TD results can range between 0 (meaning a pair of classes is not separable) and 2 (expressing the two classes

are perfectly separated). A value more than 1.9 indicates that the training points have good separability, according to Richards and Jia (1999). Initially, the spectral similarity was observed between several classes including slate roof versus asphalt, board-leaved trees versus needle-leaved trees, and concrete roof tile versus clay roof tile. Consequently, training classes were inspected and refined, with the removal and addition of points as necessary. Eventually, through repeated evaluation of TD values and refinement of training classes, all training class sets achieved satisfactory spectral separability.

3.4.2 Pixel-based classification

Initially, ML classification was performed on the twelve image data sets, using the training data as described above. The ML algorithm is perhaps the most commonly used image classification approach (Pu *et al*., 2011; Song *et al*., 2005) and is now widely-known and well-understood (i.e., see Mather and Koch, (2011) for a full description), so only brief detail is provided. The main intention of using ML classification here was to demonstrate the potential of a conventional parametric pixel-based classifier as a benchmark against which other, newer classification approaches could be compared. Though ML classification is effective where its assumptions of data normality are met, it may be that the inherent "noisiness" (i.e., spatial heterogeneity) of VHR image pixels renders this form of data unsuitable for parametric classification.

Next, a support vector machine (SVM) classification was performed on the twelve image data sets. The SVM is a non-probabilistic binary linear classifier which, through the operation of the kernel trick, determines the radial position

of decision boundaries (support vectors) that yield the optimal separation of classes (Huang *et al.,* 2002; Cortes and Vapnik, 1995). SVMs are increasingly used in image classification, often increasing classification accuracy over traditional approaches (Hamedianfar *et al.,* 2014; Otukei and Blaschke, 2010; Boyd *et al.,* 2006). Parameter settings for the SVM classifier were chosen through consideration of prevailing theory and literature where available, plus trial-and-error testing, ultimately leading to optimum classification outputs. SVM classification was conducted using ENVI image processing software (Solutions, 2015). A radial basis function nonlinear (Gaussian) kernel method was used (Jebur *et al.,* 2014; Hamedianfar *et al.,* 2014; Keerthi and Lin, 2003) because it deals with non-linear problems (Marjanovic *et al.,* 2011) and can be used for various applications (Wang and Wang, 2015; Jebur *et al.,* 2014). This kernel requires two main parameters to be determined: gamma and penalty. Gamma expresses the degree of influence of training samples in the classification process (as gamma increases, influence decreases), and penalty controls the trade-off between misclassification of training samples and simplicity of the decision surface (Pedregosa *et al*., 2011). After extensive trial-and-error testing, gamma and penalty were set at 0.5 and 500 respectively. These values were chosen because they values resulted in the highest classification accuracies.

3.4.3 Object-Based Image Analysis (OBIA/NN)

3.4.3.1 Segmentation

Following pixel-based classification, object-based classification was performed on the twelve image data sets. Object-based classification operates at the scale of identifiable objects or patches in the landscape, rather than pixels. Usually, these objects are derived directly from remotely sensed imagery, whereby spectrally similar neighbouring pixels are grouped together to form objects (Blaschke *et al.,* 2011; Mather and Koch, 2011).

Object-based classification technique involves two main steps; segmentation and classification. Segmentation as the first step of OBIA divides imagery to series of patches which group neighbouring pixels with some spatial coherence (Yan *et al*., 2006; Lu and Weng, 2007; Bhaskaran *et al*., 2010; Mather and Koch, 2011). Following segmentation, classification is conducted on the segmented objects. Each object is classified by its pixels' spectral information, but this can also be supplemented by additional discriminating variables such as object size and shape.

The study developed a multi-stage (on eCognition software) OBIA classification procedure to determine the optimum OBIA approach. Meinel and Neubert (2004) compared widely used segmentation algorithms and identified eCognition as the most successful algorithm for creating accurate segments. Multi-resolution segmentation algorithm was the main core of the segmentation which was initially developed by Benz *et al.* (2004). Various spatial parameters can influence this process. The three most important

parameters of interest are scale, shape, and compactness. These three parameters determine segmented objects by, respectively, the object's size (determined by spatial heterogeneity), its regularity of form (i.e., the complexity of an object's boundary configuration) and how closely packed the object's pixels are (through comparison of the object to a circle). However, defining accurate segments can be challenging because this process relies on human input to determine segmentation parameters.

The final segmentation algorithm was spectral difference segmentation performed after conducting multiresolution segmentation. It works by merging neighbouring image objects if the difference between their layer mean intensities is lower than a user defined threshold value (Trimble, 2013).

In the case of the 2m spatial resolution data sets, an examination of residential buildings' roofs reveals that these objects represent the majority of buildings tended to be represented by two polygons instead of one; the sunlit side of the roof and the shaded side of the roof. Studying the spectral information of these two objects emphasised that they were spectrally different. So, to negate this potential inaccuracy, each roof object was classified into two sub-classes (sunlit roof, shadow roof) and then merged at a later stage (Figure 3.4).

In the first stage of the OBIA procedure, imagery was divided into two – vegetation and non-vegetation – groups. The vegetation category contained grass, needle-leaved trees, and broad-leaved trees, whereas the other eight classes were categorised as non-vegetation (concrete roof tiles, clay roof tiles, slate roof tiles, metal roof, asphalt, bare soil, water, and shadow). Figure 3.4 outlines the OBIA\NN procedure for the 2m spatial resolution data sets. The

reader can refer to the Appendix to view a more generic OBIA procedure diagram used with the 4m, 10m, and 30m spatial resolution data sets.

Following the first stage, vegetation features were divided into their constituent classes (stage 2a), and separately non-vegetation features were divided into their constituent classes (stage 2b). While the optimum shape and compactness settings remained consistent between input data sets (see Table 3.2 below), the scale value had a significant impact on segmentation outcome (Li and Shao, 2014).

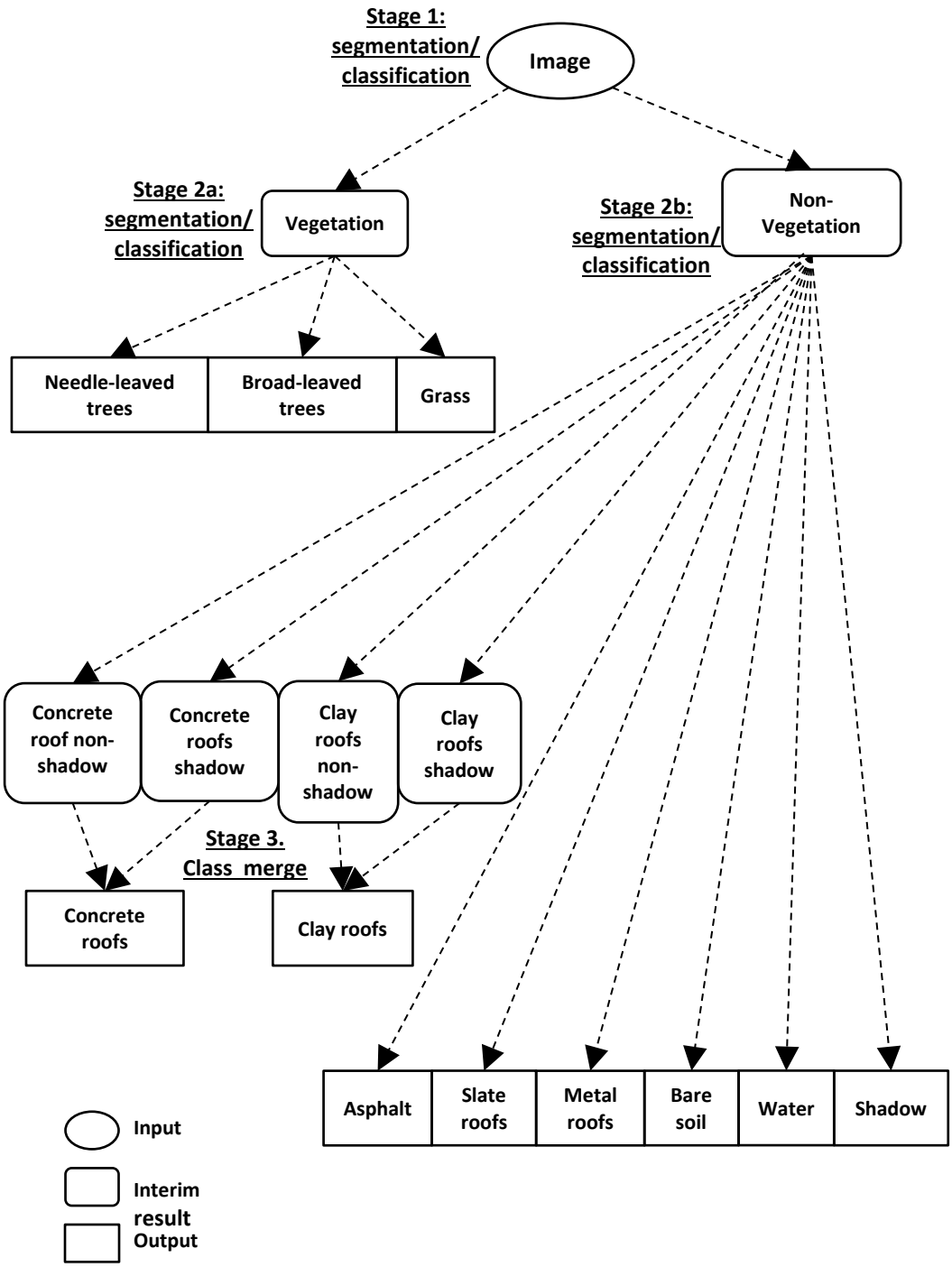


Figure 3.4. The multi-stage OBIA\NN procedure (this chart especially demonstrates the OBIA\NN of WV-2 imagery with 2m spatial resolution).

Table 3.2 depicts the parameter settings that were adopted in the OBIA/NN procedure. Notably, it is interesting to compare settings between the different spatial resolution inputs (30m, 10m, 4m, 2m), bearing in mind that in each case

sensitivity testing was used to optimise the classification outcome. It can be seen that the shape and compactness settings are consistent throughout all twelve classifications, whereas the optimum scale setting increased consistently from 30m to 2m resolution. Increasing the scale parameter increases average segment size, and this follows current context whereby the higher scale settings at finer resolutions offset smaller pixel sizes, leading to consistently sized objects (i.e., consistent between varying input spatial resolution). Also, the optimum scale setting was consistently larger for vegetation classes (i.e., large parcels of grassland and woodland) than non-vegetation classes (i.e., small urban features such as buildings and roads). Finally, following segmentation, a merging procedure can be used to combine spectrally similar objects thus refining the final segmented output. Here, this proved helpful only in the case of the input 2m spatial resolution data, since the complexity of this imagery led inevitably to some degree of over-segmentation.

Input Image Spatial Resolution (m)	Stage 1. Separation of Vegetation and Non-vegetation		Stage 2a. Identification of Individual Vegetation Classes		Stage 2b. Identification of Individual Non-vegetation Classes	
	Segmentation Parameters (Scale, Shape, Compactness)	Spectral Merging Threshold	Segmentation Parameters (Scale, Shape, Compactness)	Spectral Merging Threshold	Segmentation Parameters (Scale, Shape, Compactness)	Spectral Merging Threshold
30	6, 0.3, 0.8	NA	7, 0.3, 0.8	NA	5, 0.3, 0.8	NA
10	10, 0.3, 0.8	NA	12, 0.3, 0.8	NA	6, 0.3, 0.8	NA
4	20, 0.3, 0.8	NA	30, 0.3, 0.8	NA	12, 0.3, 0.8	NA
2	25, 0.3, 0.8	20	35, 0.3, 0.8	35	17, 0.3, 0.8	10

Table 3.2. Segmentation parameter settings for multi-stage object-based classification.

3.4.3.2 Assessing the quality of segmentation

A key factor for segmentation is how well-segmented outputs correspond to real-world features. Clinton (2004) recommended the combination of visual assessment and calculation of a series of quality segmentation equations, to select the most accurate image objects for the classification stage. Initially, visual assessment was undertaken to evaluate the quality of segmentation. This process was conducted by comparing the generated segmented images with reference objects, which resulted from OS Mastermap Topography layer.

A sensitivity test was performed to compare a range of scale parameter settings and assess their accuracy using a combination of objective metrics, as described by (Clinton *et al*., 2010), and human assessment. 40 objects were selected randomly per class and compared against reference data acquired from the MasterMap vector coverage and field survey. Weidner (2008) recommended to consider a series of metrics rather than relying on a single metric for quantitavily assessing the extracted objects, here five different metrics were used from Clinton *et al*. (2010) to check segmentation accuracy. The metrics employed were the Area Fit Index (AFI) which shows how closely segments overlap reference objects; two Relative Area (RA) measures, RAsub and RAsuper, which indicate over- and under-segmentation respectively; the Quality Rate (QR) which is an area-based measure that includes consideration of false positives when determining segmentation success; and the D-index which is a combined metric that considers both over- and under-segmentation to indicate how closely objects produced match ideal segmentation output.

Readers can refer to Clinton *et al*. (2010) for further information on these metrics.

Collectively, the five metrics provided a robust and varied test of segmentation accuracy. Nonetheless, the somewhat arbitrary nature of accuracy metrics' units and their sometimes conflicting outcomes (Weidner, 2008) means that a visual check can also be useful (Yang *et al*., 2014; Myint *et al*., 2013; Zhang *et al*., 2008). Pal and Pal (1993) claimed that human interpretation represents the most effective means of assessing segmentation output, supported later by Benz *et al*. (2004). Therefore, ultimately, both quantitative metrics and qualitative assessment were used in combination to determine final scale settings (shown in Table 3.2).

3.4.3.3 Classifying extracted objects

Following segmentation, a non-parametric nearest neighbour (NN) classifier was used to label objects to the most appropriate class. The NN classification method is also a non-parametric rule and is therefore independent of a normal distribution. The nearest neighbour approach allows unlimited applicability of the classification system to other areas, requiring only the additional selection or modification of new objects (training samples) until a satisfactory result is obtained (Ivits & Koch, 2002). The candidate pixel assigned to the class which is closer (shorter Euclidean distance) to feature space of each class. To ensure a fair comparison between pixel-based and object-based classification, the same training samples were used in all cases. As well as the straightforward spectral information provided by the different band sets, classification accuracy was enhanced (determined through trial-and-error) with additional

discriminatory variables. Certain spatial object characteristics were incorporated, including area, shape and length/width ratio; and various spectral indices (described above) were also used: NDVI for all spectral band sets, plus NSDBI and NSDBRI for the eight spectral bands set only. Here, therefore, the concrete and clay roof classes were each first classified as two separate sub-classes (i.e., non-shadowed clay roof, shadowed clay roof) and then later combined to form a single (i.e., clay roof) class (Figure 3.4, stage 3).

3.5 Accuracy assessment and statistical testing

3.5.1 Error matrix

Following classification, the 36 output land cover maps were tested against reference data to calculate their accuracy. The confusion (or error) matrix is a common land cover classification accuracy assessment method. This evaluation approach is carried out by comparing predicted and actual class labels from selected testing points. As a result, overall, users, and producers accuracy was calculated for the 36 land cover maps. To enable direct comparison between the three different classification approaches, it was necessary to adopt a common means of accuracy assessment, so point-based checking was conducted. It is important to note that alternative object-based approaches are now available for use with OBIA outputs (Stehman and Wickham, 2011; Congalton and Green, 2009; Zhan *et al*., 2005; Radouxa *et al*., 2011), and these can have the benefit of providing an exact match between analysis data (i.e., classified objects) and reference data (i.e., vector map features). In total, 438 sample points were checked, with an average of 40 points per class. The same

points were used for all classification outputs, ensuring direct comparability between results.

3.5.2 McNemar statistical Test

While confusion matrices provide a useful means of comparing classification results, they can only provide an "estimate" of classification accuracy (based on the sample of points used), and therefore only tentative conclusions can be drawn (Foody, 2004). This is especially the case where differences in accuracy are marginal, for instance, a few percentage points apart. For example, it may be unwise to assert that a land cover map with an accuracy of, say, 93% is definitively more accurate than a map with 89% accuracy. This 4% difference may, in fact, be a statistical artefact of the sample of test points. Because of this, Janssen and Vanderwel (1994) stated that, accuracy statements should be provided with confidence limits enabling comparison in a statistically rigorous manner. As a result, the McNemar test was used to compare classification outputs and indicate the statistical significance of any difference in results (Foody, 2004). That is, in the example above, the McNemar test could provide information on whether or not a difference of 4% is statistically significant. As Foody (2004) notes, this is a non-parametric test that is focused on the binary distinction between correct and incorrect class allocations of two classification outputs. The McNemar test calculates the z value:

$$z = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \qquad (1)$$

In this formula, f_{21} indicates the total number of ground data correct in land cover map 2 and incorrect in land cover map 1, while f_{12} represents the

frequency reference point incorrect in thematic map 2 and correct in thematic map 1. If $|z| \geq 3.3$, it means that there is a significant difference between two thematic maps at the 99% confidence level (Rees, 1987). Here, a fully rigorous and exhaustive approach was adopted for expressing the statistical significance of classification output differences. The McNemar test was conducted on every possible pair of classified land cover maps. With 36 original maps, this meant 630 paired combinations. The results, expressed as a matrix, enable straightforward comparison between all classifications, clearly identifying those classification pairs that are significantly different and those that are not.

3.6 Results

3.6.1 General overview

In total, 36 land cover maps were generated, and extracts of all 36 classified outputs are provided in Figure 3.5. It should be noted that this figure should be interpreted with some caution since it shows only one small area and is not therefore fully representative of land cover throughout the entire study area. Nonetheless, the figure clearly shows the most significant and consistent pattern evident throughout the results: classification improves as the spatial resolution becomes finer.

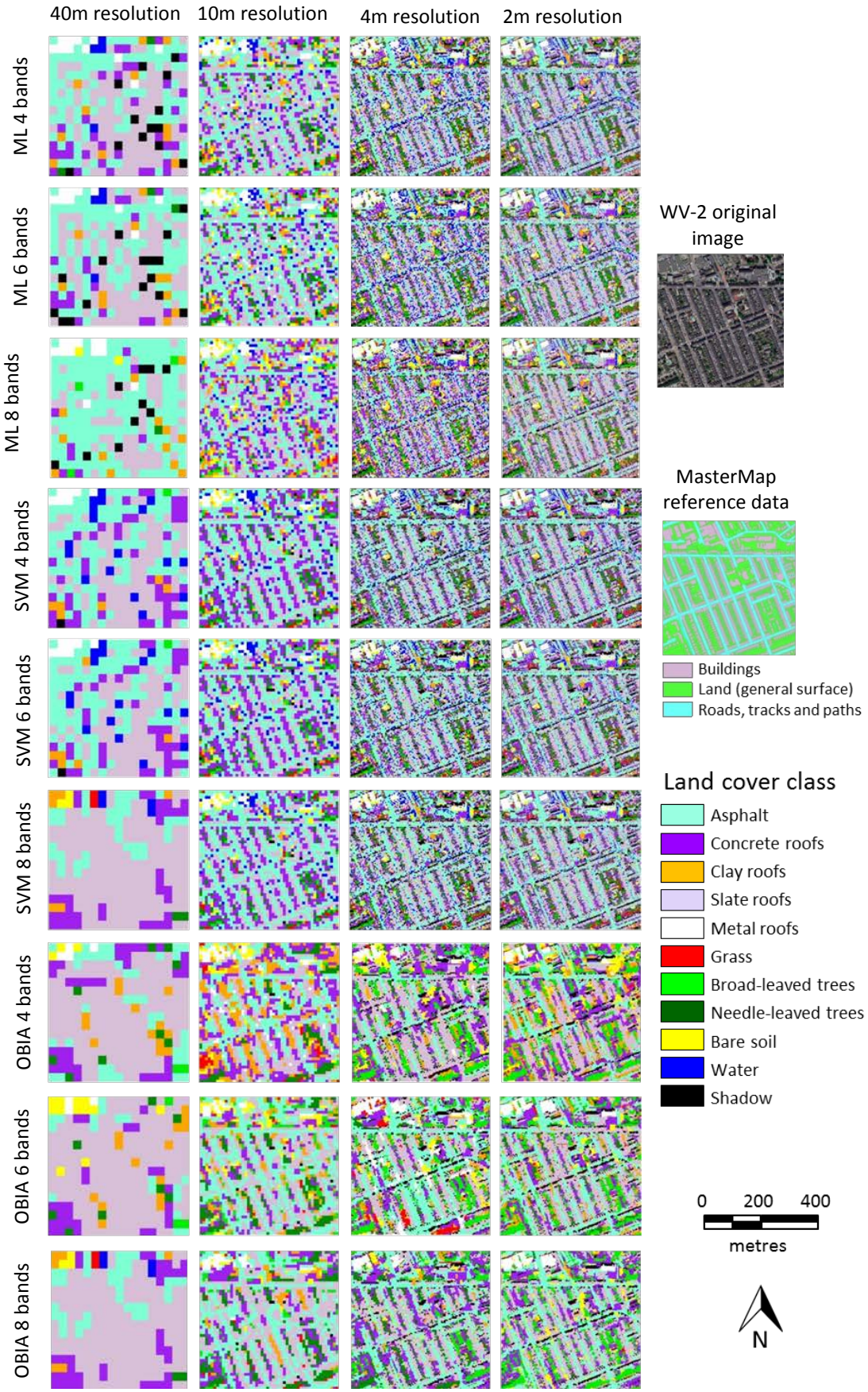


Figure 3.5. Land cover maps (detail) for the 36 data set/classifier combinations, plus the WorldView-2 image (© DigitalGlobe, Inc. All Rights Reserved) and MasterMap vector data (© Crown Copyright and Database Right 2015. Ordnance Survey (Digimap Licence).

A summary of classification accuracies for all 36 land cover maps is provided in Figure 3.6. This figure presents raw overall classification accuracies and enables direct assessment of the differences between classifications.

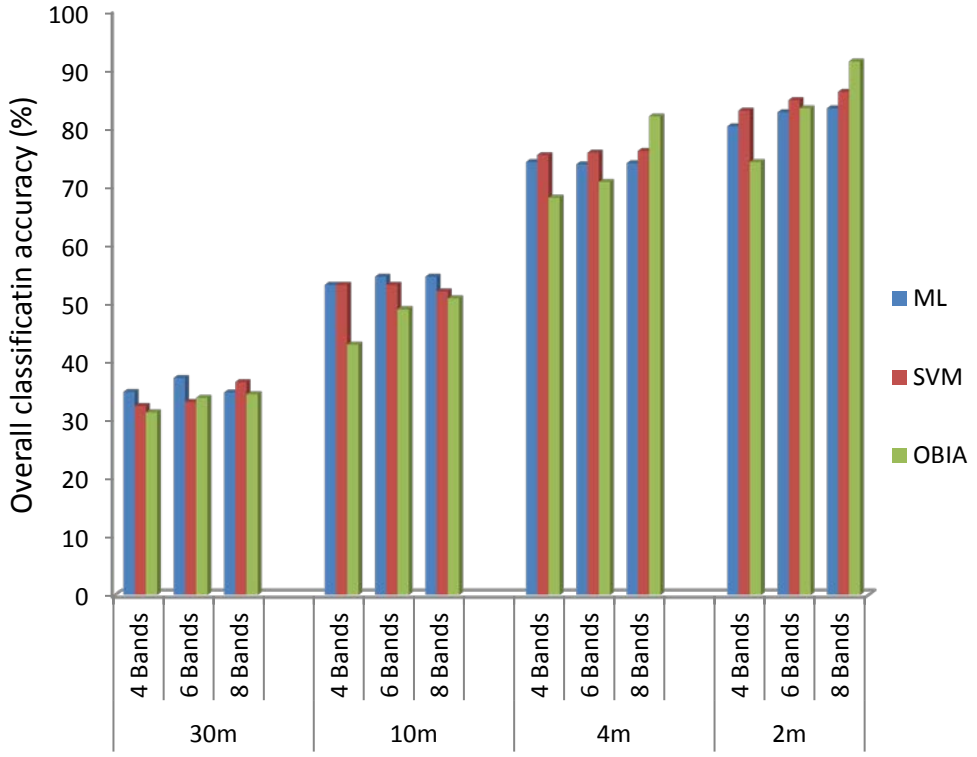


Figure 3.6. The overall classification accuracies for the 36 data set/classifier combinations (ML, SVM, and OBIA/NN).

While the full set of statistical classification results are summarised above (figure 3.5), six full error matrices are first presented to give some examples of class-level detail (Tables, 3.2 – 3.7).

To provide contrast and show the full range of classification accuracies, two confusion matrices were chosen from each of the classifiers used. Each of these three classifiers was represented by two different spatial and spectral resolution properties. As such tables 3.2 and 3.3 depict the resultant error matrixes of performing OBIA/NN on images with the full eight spectral bands,

with 2m and 4m spatial resolution respectively. The highest overall classification accuracy (91%) was achieved by arguably the most sophisticated data set/classifier combination, using the most advanced spatial and spectral characteristics of WV-2 imagery (2m spatial resolution and full eight spectral bands) and state-of-the-art OBIA (Table 3.2). This classification enabled relatively accurate mapping of all classes, with only minor confusion between vegetation classes and between concrete and other impervious classes.

The error matrixes of performing SVM classification on data sets with six spectral bands with 2m and 4m spatial resolution are presented in Tables 3.4 and 3.5 respectively. The accuracy of ML classifier on data sets with four spectral bands and 10m and 30m spatial resolution are described in Table 3.6 and 3.7 respectively. It is clear that the less sophisticated data set/classifier combination (using relatively coarse 30m resolution and only four basic spectral bands) is entirely inadequate in classifying such detailed urban land cover classes (Table 3.7). Few classes are mapped with any success, and overall classification accuracy is only 35%. Perhaps the most significant factor here, as will be discussed below, is the coarse spatial resolution, which prevents accurate identification of small urban features. Finally, OBIA achieve relatively poorer accuracies by using fewer spectral bands. This is possibly due to a reduction in the spectral information that is considered when generating the segments.

OBIA\NN on WV-2 with 2m spatial resolution and eight spectral bands

Overall Accu: 91% Kappa Coeff: 0.9	Reference data											
Classification result / **Class**	**Bare soil**	**Asphalt**	**Water**	**Broad L trees ***	**Needle L tree***	**Clay roof**	**Metal roof**	**Slate roof**	**Shadow**	**Concrete roof**	**Grass**	**User accuracy %**
Bare soil	**29**	0	0	0	0	1	1	0	0	0	0	**94**
Asphalt	0	**37**	0	0	0	1	0	1	0	1	0	**93**
Water	0	0	**43**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	0	0	**40**	8	0	0	0	0	0	5	**75**
Needle L trees*	0	0	0	2	**30**	0	0	0	0	0	0	**94**
Clay roof	0	0	0	0	0	**34**	0	1	0	1	0	**94**
Metal roof	0	0	0	0	0	0	**39**	0	0	0	0	**100**
Slate roof	0	1	0	0	0	0	0	**35**	0	1	0	**95**
Shadow	0	0	0	0	0	0	0	0	**39**	0	0	**100**
Concrete roof	0	4	0	0	0	4	0	3	0	**38**	0	**78**
Grass	1	1	0	0	0	0	0	0	1	0	**36**	**92**
Producer accuracy %	**97**	**86**	**100**	**95**	**79**	**85**	**98**	**88**	**98**	**93**	**88**	

Table 3.3. Error matrix for performing OBIA/NN on WV-2 with 2m spatial resolution and eight spectral bands.

OBIA\NN on WV-2 with 4m spatial resolution and eight spectral bands

Overall Accu: 82% Kappa Coeff: 0.8	Reference data											
Classification result / **Class**	**Bare soil**	**Asphalt**	**Water**	**Broad L trees ***	**Needle L tree***	**Clay roof**	**Metal roof**	**Slate roof**	**Shadow**	**Concrete roof**	**Grass**	**User accuracy %**
Bare soil	**23**	0	0	0	0	3	2	0	0	0	0	**82**
Asphalt	0	**34**	0	0	0	1	1	1	1	2	0	**85**
Water	0	0	**41**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	0	0	**34**	1	0	0	0	0	1	5	**83**
Needle L trees*	0	0	0	8	**37**	0	0	0	1	1	3	**74**
Clay roof	5	1	0	0	0	**28**	0	0	0	0	0	**82**
Metal roof	0	0	0	0	0	0	**36**	0	0	0	0	**100**
Slate roof	0	1	1	0	0	1	0	**31**	5	3	0	**74**
Shadow	0	0	0	0	0	1	0	0	**29**	0	0	**97**
Concrete roof	1	5	0	0	0	6	1	10	4	**33**	0	**55**
Grass	1	1	0	0	0	0	0	0	0	1	**32**	**91**
Producer accuracy %	**77**	**81**	**98**	**81**	**97**	**70**	**90**	**74**	**73**	**80**	**80**	

Table 3.4. Error matrix for performing OBIA/NN on WV-2 with 4m spatial resolution and eight spectral bands.

SVM on WV-2 with 2m spatial resolution and six spectral bands

Overall Accu: 85% Kappa Coeff: 0.83 Class	Reference data											
Classification result	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**28**	0	0	0	0	0	0	0	0	0	0	**100**
Asphalt	2	**38**	5	0	0	1	0	2	0	0	0	**79**
Water	0	0	**31**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	0	0	**23**	1	0	0	0	0	0	2	**88**
Needle L trees*	0	0	0	19	**24**	0	0	0	0	0	0	**56**
Clay roof	0	0	0	0	0	**34**	0	0	0	1	0	**97**
Metal roof	0	0	0	0	0	0	**40**	0	0	0	0	**100**
Slate roof	0	4	4	0	0	2	0	**35**	1	0	0	**76**
Shadow	0	0	3	0	13	0	0	1	**39**	0	0	**70**
Concrete roof	0	1	0	0	0	3	0	2	0	**40**	0	**87**
Grass	0	0	0	0	0	0	0	0	0	0	**39**	**100**
Producer accuracy %	**93**	**88**	**72**	**55**	**63**	**85**	**100**	**88**	**98**	**98**	**95**	

Table 3.5. Error matrix for performing SVM on WV-2 with 2m spatial resolution and six spectral bands.

SVM on degraded WV-2 with 4m spatial resolution and six spectral bands

Overall Accu: 76% Kappa Coeff: 0.73 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Classification result												
Bare soil	**26**	0	0	0	0	1	4	0	0	1	0	**81**
Asphalt	1	**35**	8	0	0	1	1	6	1	3	0	**63**
Water	0	0	**29**	0	0	0	1	0	1	0	0	**94**
Broad L trees*	0	0	0	**21**	1	0	0	0	0	0	1	**91**
Needle L trees*	0	0	0	15	**22**	0	0	0	0	0	1	**58**
Clay roof	2	1	0	0	0	**30**	0	0	0	2	0	**86**
Metal roof	0	0	0	0	0	0	**34**	0	0	0	0	**100**
Slate roof	0	1	3	0	0	1	0	**26**	1	0	0	**81**
Shadow	0	0	2	1	15	1	0	2	**35**	0	0	**63**
Concrete roof	0	5	0	0	0	6	0	8	2	**35**	0	**63**
Grass	1	0	0	5	0	0	0	0	0	0	**38**	**86**
Producer accuracy %	**87**	**86**	**69**	**50**	**58**	**75**	**85**	**62**	**88**	**85**	**95**	

Table 3.6. Error matrix for performing SVM on degraded WV-2 with 4m spatial resolution and six spectral bands.

ML on degraded WV-2 with 10m spatial resolution and four spectral bands

Overall Accu: 53% Kappa Coeff: 0.48 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**20**	1	0	0	0	7	3	0	0	4	3	**53**
Asphalt	1	**26**	4	6	0	2	2	11	6	3	3	**41**
Water	0	0	**35**	0	0	0	0	0	6	0	0	**85**
Broad L trees*	0	2	0	**16**	12	0	0	0	1	5	4	**40**
Needle L trees*	2	1	1	7	**20**	1	0	3	5	3	4	**43**
Clay roof	4	1	1	4	1	**24**	0	6	3	11	1	**43**
Metal roof	0	0	0	0	0	0	**35**	0	3	0	0	**92**
Slate roof	1	5	0	0	0	4	0	**16**	8	3	0	**43**
Shadow	0	0	1	0	3	0	0	0	**2**	0	0	**33**
Concrete roof	0	6	0	2	0	2	0	5	6	**12**	0	**36**
Grass	2	0	0	7	1	0	0	0	0	0	**25**	**71**
Producer accuracy %	**67**	**62**	**83**	**38**	**54**	**60**	**88**	**39**	**5**	**29**	**63**	

Classification result

Table 3.7. Error matrix for performing ML on degraded WV-2 with 10m spatial resolution and four spectral bands.

ML on degraded WV-2 with 30m spatial resolution and four spectral bands

Overall Accu: 35% Kappa Coeff: 0.28 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**16**	3	2	1	9	1	7	0	2	3	6	**32**
Asphalt	3	**14**	6	7	0	2	5	7	12	9	2	**21**
Water	0	0	**22**	0	0	0	0	0	4	1	2	**76**
Broad L trees*	3	0	1	**13**	16	1	0	1	1	2	12	**26**
Needle L trees*	0	0	5	4	**2**	0	0	0	1	0	1	**15**
Clay roof	0	6	0	4	3	**18**	2	6	5	6	5	**33**
Metal roof	2	0	0	0	0	0	**25**	2	6	0	0	**71**
Slate roof	0	3	0	0	0	1	0	**12**	2	3	0	**57**
Shadow	2	0	3	2	1	2	0	2	**0**	0	0	**0**
Concrete roof	2	16	3	4	1	15	1	11	7	**18**	0	**23**
Grass	2	0	0	7	5	0	0	0	0	1	**12**	**44**
Producer accuracy %	**53**	**33**	**52**	**31**	**5**	**45**	**63**	**29**	**0**	**42**	**30**	

Classification result

Table 3.8. Error matrix for performing ML on degraded WV-2 with 30m spatial resolution and four spectral bands.

However, this does not indicate which of these differences are statistically significant. Therefore, a full matrix of z values (calculated from the McNemar test of statistical significance) between all 630 classification pairs is also presented, in Figure 3.5. Statistically significant differences (i.e., z values ≥ 3.3, at the 99% confidence level) are bolded. It is noted that the figure clearly contains a large volume of information and requires careful interpretation, but it is included here to enable comprehensive and exhaustive comparison between data set/classifier combinations.

			ML												SVM												OBIA											
			30m			10m			4m			2m			30m			10m			4m			2m			30m			10m			4m			2m		
			4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b	4b	6b	8b
ML	30m	4b																																				
		6b	2.1																																			
		8b	1.4	3.2																																		
	10m	4b	3.9	2.6	4.9																																	
		6b	4.3	3.0	5.2	0.9																																
		8b	3.8	2.6	5.0	0.2	0.9																															
	4m	4b	8.7	7.8	9.3	6.2	6.0	6.2																														
		6b	8.8	7.9	9.2	6.3	6.1	6.3	0.0																													
		8b	8.6	7.9	9.4	6.0	5.7	6.2	0.0	0.0																												
	2m	4b	9.6	8.9	10.2	7.1	7.0	8.5	1.9	1.9	1.8																											
		6b	10.5	9.7	11.0	8.3	8.1	8.4	3.5	3.5	3.2	2.3																										
		8b	10.9	10.1	9.3	3.9	8.4	7.2	6.2	3.6	3.7	2.5	0.6																									
SVM	30m	4b	0.4	1.2	1.5	3.4	3.8	3.4	8.3	8.3	8.3	9.4	10.2	10.5																								
		6b	0.6	0.9	1.8	3.3	3.7	3.3	8.2	8.2	8.2	9.3	10.2	10.4	0.6																							
		8b	2.5	0.7	3.5	2.1	2.5	2.0	7.4	7.3	7.3	8.5	9.3	9.6	3.0	3.0																						
	10m	4b	3.8	2.6	4.7	0.0	0.6	0.2	6.3	6.3	6.1	7.3	8.4	8.7	3.5	3.4	2.1																					
		6b	3.9	2.6	4.8	0.0	0.6	0.2	6.3	6.4	6.1	7.3	8.4	8.9	3.5	3.4	2.1	0.0																				
		8b	3.2	1.9	4.2	1.0	1.6	0.9	6.7	6.8	6.7	7.8	8.9	9.4	2.8	2.7	1.4	1.6	1.7																			
	4m	4b	8.7	8.0	9.4	6.4	6.1	6.2	0.6	0.5	0.6	1.4	2.8	3.3	8.6	8.5	7.6	6.5	6.5	7.1																		
		6b	8.9	8.2	9.6	6.7	6.4	6.5	0.9	0.9	1.0	1.1	2.6	3.1	8.7	8.7	7.8	6.8	6.8	7.4	0.8																	
		8b	9.2	8.7	10.0	7.1	6.8	6.9	1.5	1.4	1.6	0.7	2.2	2.6	9.1	9.1	8.2	7.1	7.1	7.8	1.8	0.9																
	2m	4b	10.3	9.7	10.9	8.1	7.9	7.9	3.3	3.1	3.2	1.6	0.1	0.7	10.1	10.0	9.1	8.3	8.2	8.6	3.0	4.6	2.2															
		6b	11.1	10.4	11.6	8.7	8.5	8.5	4.4	4.2	4.2	3.1	1.2	0.6	10.8	10.7	9.9	8.9	8.7	9.3	4.2	3.8	3.4	2.1														
		8b	11.3	10.8	11.9	9.3	9.1	9.2	5.0	4.8	4.8	3.8	2.2	1.7	11.1	11.1	10.3	9.3	9.4	9.8	4.9	2.7	4.3	2.9	1.6													
OBIA	30m	4b	1.4	2.7	0.3	5.0	5.1	4.7	9.6	9.7	9.4	10.5	11.5	11.6	1.9	2.2	3.6	4.8	4.8	4.2	9.7	10.0	10.3	11.2	11.9	12.3												
		6b	0.7	2.0	0.4	4.2	4.4	4.1	8.7	8.7	8.8	9.8	10.6	10.8	1.1	1.3	2.8	4.3	4.2	3.6	9.1	9.4	9.6	10.6	11.1	11.5	0.9											
		8b	0.7	2.1	0.4	4.3	4.7	4.2	9.4	9.3	9.2	10.1	11.0	11.2	1.1	1.3	2.8	4.4	4.4	3.8	9.5	9.7	10.0	11.0	11.6	11.9	0.8	0.0										
	10m	4b	0.5	0.6	1.5	3.6	3.8	3.4	8.6	8.5	8.3	9.3	10.4	10.5	0.2	0.1	1.1	3.6	3.5	2.7	8.8	9.0	9.3	10.1	10.7	11.1	1.7	1.1	1.1									
		6b	2.8	1.6	3.8	1.3	1.6	1.1	6.9	6.8	6.7	7.7	8.7	9.1	2.5	2.3	1.1	1.3	1.3	0.4	7.1	7.4	7.7	8.4	9.0	9.6	4.2	3.5	3.5	3.0								
		8b	3.7	2.4	4.7	0.3	0.7	0.1	6.3	6.4	6.3	7.3	8.4	8.8	3.3	3.1	1.9	0.2	0.2	0.6	6.3	6.7	7.0	8.1	8.6	9.2	4.7	4.1	4.1	3.7	1.1							
	4m	4b	7.7	6.6	8.5	4.9	4.5	5.0	1.4	1.4	1.5	3.0	4.2	4.6	7.3	7.2	6.3	5.0	4.9	5.6	1.8	2.1	2.5	4.2	4.8	5.3	8.8	8.2	8.1	7.2	5.7	5.0						
		6b	7.4	6.4	8.4	4.6	4.2	4.8	1.7	1.7	1.8	3.2	4.4	4.9	7.0	6.9	5.9	4.8	4.7	5.3	2.2	2.4	2.9	4.4	5.3	5.8	8.4	7.8	7.7	7.3	5.5	5.0	4.0					
		8b	9.9	9.0	10.7	7.7	7.1	7.7	1.7	1.7	1.6	0.1	1.4	1.8	9.6	9.5	8.6	7.7	7.7	8.2	1.2	0.9	0.5	1.3	2.3	2.9	10.8	10.3	10.2	9.9	8.4	8.0	3.1	3.6				
	2m	4b	14.0	13.4	14.4	12.5	12.3	12.5	9.1	9.1	9.1	8.1	7.4	7.2	13.7	13.7	13.1	12.6	12.6	12.8	8.8	8.7	8.4	7.5	6.8	6.3	14.5	14.3	14.2	13.8	12.7	12.4	9.9	10.2	8.2			
		6b	11.2	10.4	11.9	9.6	9.2	9.5	4.3	4.2	4.4	2.8	1.8	1.2	11.2	11.2	10.4	9.5	9.4	10.0	3.9	3.7	3.3	1.8	0.8	0.1	12.5	12.0	12.0	11.5	10.3	9.8	5.8	6.2	3.1	6.1		
		8b	12.5	12.0	13.2	11.0	10.6	11.1	6.7	6.6	6.8	5.3	4.2	3.9	12.6	12.5	11.8	11.2	11.1	11.4	6.4	6.3	5.8	4.3	3.4	2.7	13.4	13.3	13.0	12.6	11.6	11.2	7.8	8.1	6.2	4.5	2.7	

Figure 3.7. Matrix of McNemar test results showing the statistical significance of differences between all classification pairs (ML = maximum likelihood, SVM = support vector machine, OBIA = object-based image analysis; 30m, 10m, 4m and 2m refer to spatial resolution; four, six and eight bands refer to the number of spectral bands).

3.6.2 Spatial enhancement

The results demonstrated a strong correlation between spatial resolution and overall accuracy. It is evident that as the spatial resolution became finer—from 30m, through 10m and 4m, to 2m—the overall accuracy of the entire 36 thematic maps increased consistently, with all spectral band sets and all classifiers (Figure 3.6).

Figure 3.6 demonstrates the overall accuracy of implementing ML, SVM, and OBIA/NN on data sets with 30m, 10m, 4m, and 2m resolution including the influence of classifier and number of spectral bands used. According to Figure 3.8, data sets with 2m spatial resolution (represented by purple bars) achieved the highest overall accuracy, followed by 4m, 10m, and 30m. The overall accuracy of data sets with 30m, 10m, 4m, and 2m spatial resolution was as 34%, 52%, 74%, and 83% (averaging nine classified images with different spectral resolution and classifier but similar spatial resolution). In other words, the overall accuracy of classified images when spatial resolution became finer from 30m to 10m, 10m to 4m, and 4m to 2m, increased by 16%, 22%, and 9% respectively. Despite these consistent improvement, the largest increase in overall accuracy occurred when the spatial resolution improved from 10m to 4m with the overall accuracy enhanced by 23% compared with 17% and 9% when the pixel size changes from 30m to 10m and 4m to 2m.

According to the McNemar statistical test, there were consistent and significant statistical differences using finer spatial resolution imagery (Figure 3.7). Figure 3.9 demonstrates the variation in the level of statistical difference of data sets with 10m, 4m, and 2m against data sets with 30m spatial resolution. It is

evident that there were statistically significant differences when spatial resolution improved from 30m to 10m, 4m, and 2m. The average z value between 30m and 10m, 10m and 4m, and 4m and 2m reveals a constant significant at 99% level of confidence differences of 4.8, 5.1, 3.6 respectively.

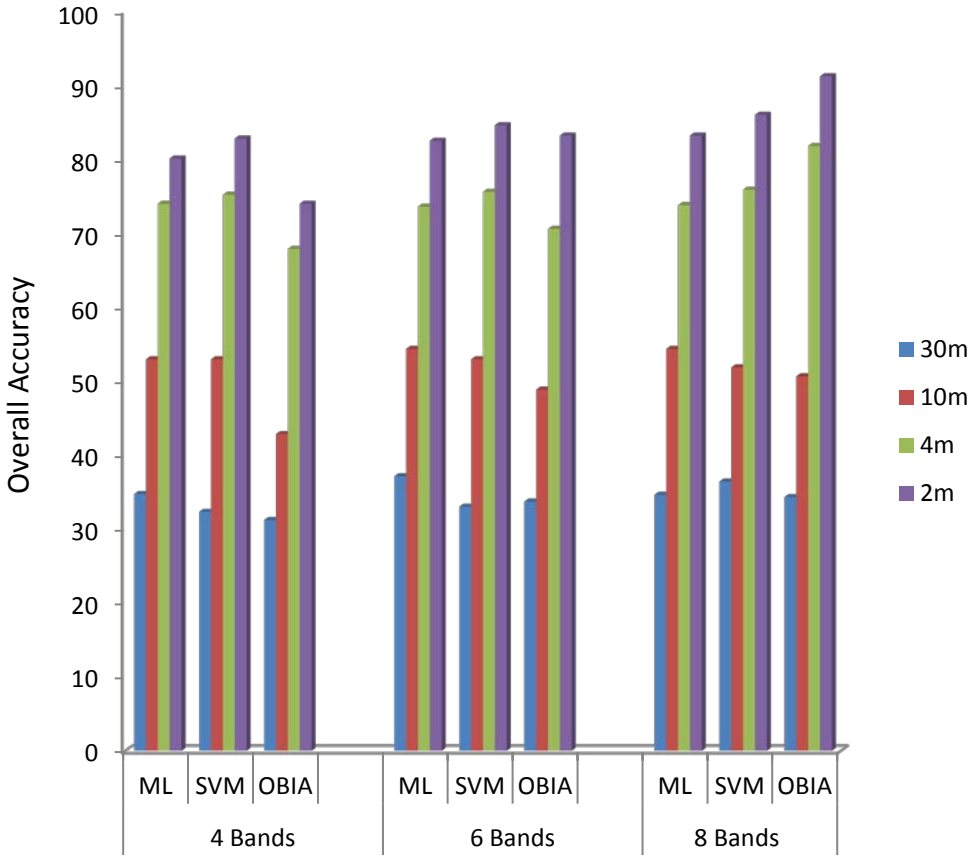


Figure 3.8. The overall accuracy of 36 land cover maps by highlighting the influence of spatial resolution.

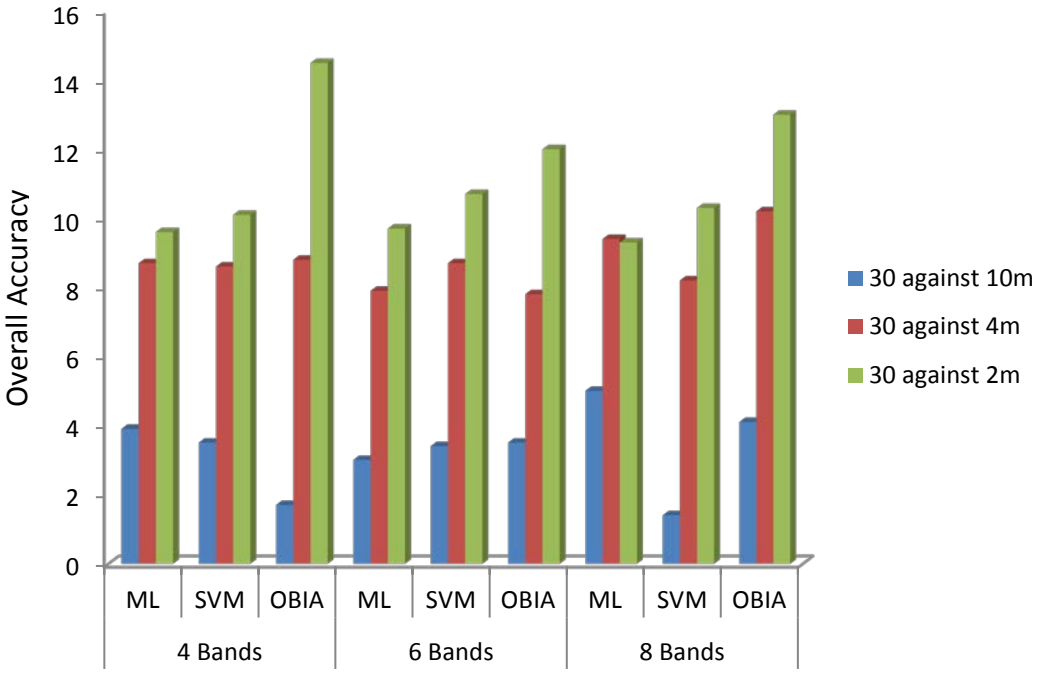


Figure 3.9. The McNemar's statistical test, comparing 30m spatial resolution against 10m, 4m, and 2m spatial resolution.

3.6.3 Spectral enhancement

The relationship between the number of spectral bands and classification accuracy is less marked than that of spatial resolution. Figure 3.8 indicates the overall accuracy of 36 land cover maps by highlighting the influence of three different spectral band sets: four, six, and eight spectral bands.

The results show that enhancing spectral resolution in general terms can lead to modest increases in classification accuracy. However, more complex spectral data sets (i.e. those with six and eight spectral bands compared to those with four spectral bands) do not necessary increase classification accuracy, especially in the 10m to 30m spatial range resolution when pixel based classifiers (ML and SVM) were implemented. For instance, for the 30m imagery, the average accuracy increases only very slightly from 33% for four bands to

35% for six bands and 35% for eight bands. It should be noted that the influence of the number of spectral bands on classification accuracy also depends on the classifier used, as will be depicted in the following section.

Since this study evaluates the influence of varying spectral band set on the overall accuracy of classification using fixed training data sets, Pal and Foody, (2010) highlighted Hughes phenomenon as a possible factor which can effect classification accuracy. Hughes phenomenon describes how classification accuracy initially increases gradually when spectral bands are added in turn to a data set, but then starts to decrease by further increasing the number of spectral bands (Ma *et al.,* 2013).

According to Pal and Foody (2010) the influence of Hughes phenomenon becomes evident when the number of spectral bands exceeds 30, however in this study the maximum number of spectral bands was 11 (8 multi-spectral WV-2 plus three indices). As a result, the effect of the Hughes phenomenon should be very limited in this study.

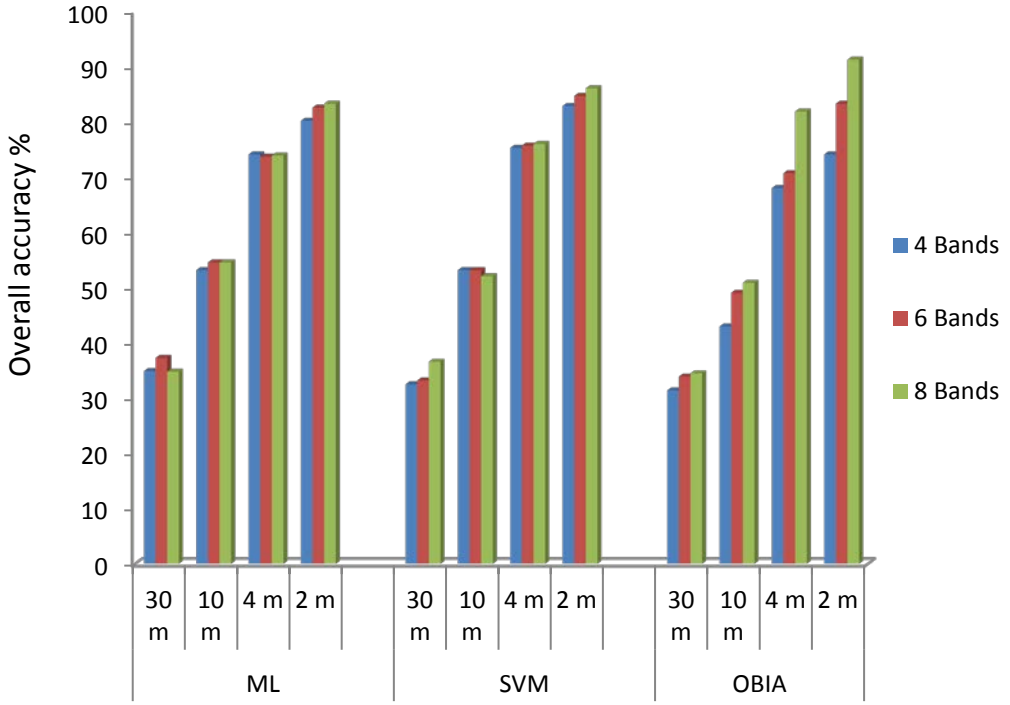


Figure 3.10. The influence of using various spectral bands on overall classification accuracy of 36 land cover maps.

Figure 3.10 indicates the variation in overall accuracy of data sets with four, six, and eight spectral bands. Despite the inconsistent relationship between spectral resolution and classifier types, there was a consistent and direct relationship between spectral and spatial resolution. The result shows that the influence of using more spectral bands became more significant as spatial resolution became finer. For instance, for the 2m resolution imagery, average accuracy (i.e., the average of all three classifiers) increases from 79% when using four bands, to 84% when using six bands and 87% when using eight bands (Figure 3.11). This figure shows that the influence of the number of spectral bands becomes more evident when spatial resolution became finer.

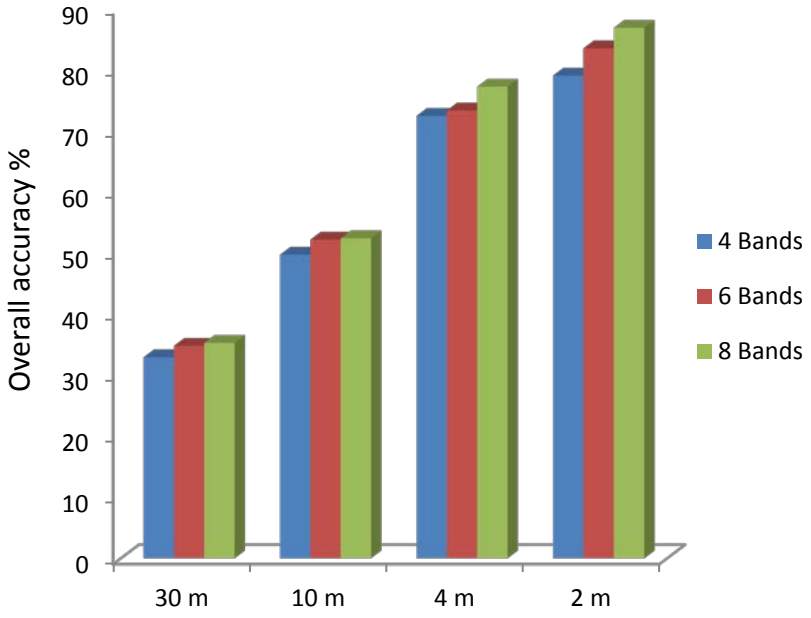


Figure 3.11. The relationship between spectral and spatial resolution.

3.6.4 Classification approach

The influence of choice of classifier on classification accuracy is more complex than that of spatial resolution or spectral band set. The results show that the classifier can have a noticeable effect on accuracy, but only when considered in combination with spatial resolution and/or number of spectral bands (Figure 3.10). Generally, within the medium spatial scale (10m and 30m), ML performed slightly better, followed by SVM and OBIA/NN. In particular, ML achieved the finest overall accuracy using four and six spectral bands. According to the McNemar test, there was no significant difference among ML and both SVM and OBIA/NN at the medium spatial scale. The poor accuracy of OBIA/NN and SVM in the medium resolution range can be attributed to the inadequate level of spatial detail to map detailed urban features accurately.

At the highest spatial resolutions (4m and 2m), patterns related to the choice of classifier change from those observed at the coarser resolutions, and also become more defined (Figure 3.10). In fact, moving from medium to VHR scale (4m and 2m) showed the superiority of non-parametric SVM and OBIA/NN classifiers. Rogan and Chen (2004) explained that because detailed urban areas are not distributed normally in the data they contain, non-parametric approaches such as SVM can better handle the classification of these complex sites. According to the McNemar test, there was a significant statistical difference in 99% level of significant between SVM and ML in the VHR range (2m and 4m).

Compared to OBIA/NN, SVM achieved a higher overall accuracy in 4m and 2m spatial resolution using four and six spectral bands. However, OBIA/NN became a superior classifier in 4m and 2m spatial resolution using eight full bands. For example, the overall accuracy of performing the OBIA/NN on 2m imagery was obtained as 91% improving the accuracy of mapping by 6% compared to the averaging of both ML and SVM classifiers. There were significant statistical differences between OBIA/NN and both pixel-based classifiers for data sets with 4 and 2m spatial resolution.

According to figure 3.12, OBIA/NN showed the highest sensitivity to spectral resolution variation. For example, the overall accuracy of performing OBIA/NN on data sets with the full eight spectral bands achieved 35 %, 51%, 82%, and 91% for 30m, 10m, 4m, and 2m resolution respectively. According to the McNemar test, comparing data sets with six and eight spectral band sets led to statistically significant differences at the 99% level of confidence only when

OBIA/NN was performed. This means that there was poor accuracy of OBIA/NN by using four and six spectral bands, while OBIA/NN show its superiority by consideration of full eight spectral bands. In particular, at 2m spatial resolution, OBIA/NN improved the overall accuracy by using six bands over four bands and also eight bands over six bands by 9% and 8% respectively.

Involving additional spectral bands also increases the complexity of the data sets which the ML algorithm may not be able to deal with. While ML showed inconsistent improvement when adding more spectral bands, SVM was more reliable and consistent towards using finer spectral resolution data sets.

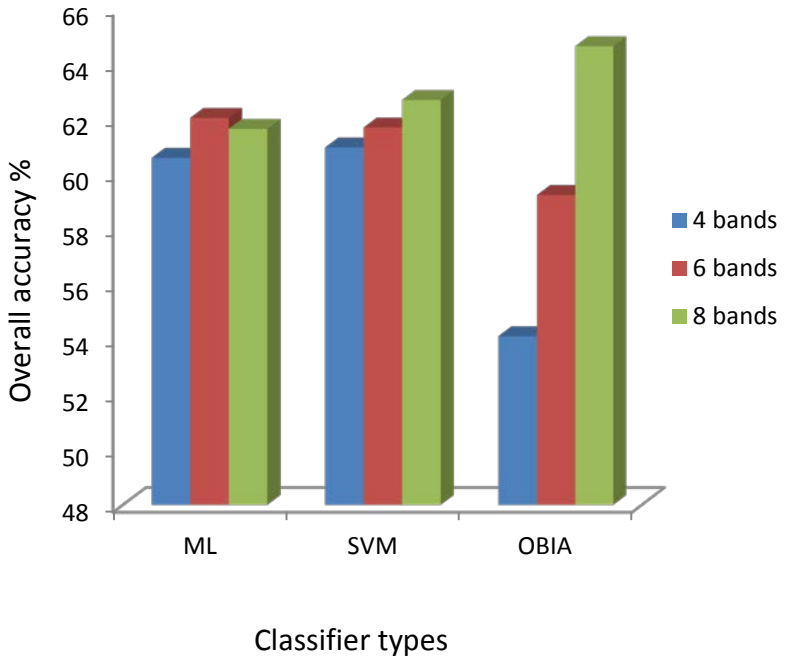


Figure 3.12. Overall classification accuracies: the relationship between spectral resolution and classification approach.

SVM as a non-parametric pixel-based classifier is more suitable to deal with mapping complex urban areas, compared to a parametric classifier like ML. This non-parametric classifier performed better than ML at 2m, 4m, and 10m spatial

resolution. For instance, SVM increases the overall accuracy of 2m imagery by 5% in contrast to the parametric classifier. However, with a very low overall accuracy of 35% both classifiers performed similarly.

As figure 3.10 shows, OBIA\NN is superior only at 2m and 4m spatial resolution when all eight bands were used. Including more spectral bands potentially increases the capability of classifier for more accure classification, by improving the quality of the segments. For the four and six band data sets, SVM achieves the best performance at 2m and 4m resolution, while ML is best at 10m and 30m resolution.

3.7 Summary

The findings presented in this chapter reaffirm the contention that accurate and detailed mapping of complex urban environments requires spatially detailed data, and contemporary VHR imagery holds considerable value for the mapping community. The results of this research verify the importance of third generation sensors like IKONOS and WV-2 with 4m and 2m spatial resolution respectively to map complex urban areas over second-generation sensors including SPOT and Landsat with 10m and 30 m spatial resolution.

Further, it is demonstrated that enhanced spectral information can aid distinction of detailed thematic classes in complex environments. Notably, here, the additional bands offered by contemporary VHR sensors such as WV-2 and WV-3 may offer some advantage over early VHR sensors such as IKONOS. The choice of the classifier can influence the accuracy with which complex urban environments are mapped, but due consideration should also be given

to image characteristics. Contemporary classification approaches such as pixel-based SVM and OBIA/NN perhaps hold considerable potential here where state-of-the-art VHR imagery (i.e., with enhanced spectral capabilities) are available. In particular, OBIA/NN is likely to improve the accuracy if there is an adequate spatial detail provided based of the level of the details. Therefore, this result showed the requirement of using VHR data sets and superiority of non-parametric including SVM and OBIA/NN over parametric (i.e., ML) classifier for mapping the complex urban areas.

Despite the superiority of VHR imagery over medium resolution imagery (like Landsat imagery) to accurately map detailed urban areas, the wide usage of VHR imagery is hindered because of a series of reasons including the high cost of VHR imagery. On the other hand, while medium resolution imagery like Landsat and equivalent SPOT (with 30m and 10m spatial resolution) are accessible free of charge, their coarse spatial resolution limits the use for a detailed urban mapping application.

The following chapters investigate how (free) medium resolution imagery may be spatially enhanced for detailed land cover classification. In particular, the next chapter provides a comprehensive review of the literature of super-resolution analysis.

Chapter 4: Super Resolution Image Analysis

4.1 Overview

While sensors with spatial resolution more than 4m are considered to offer VHR imagery, remote sensors with spatial resolution between 4 and 10m offer high resolution imagery. The medium resolution sensors (spatial resolution between 10 and 30m) have been used for various environmental urban applications. The key point of Chapter 3 was the necessity of using the finer spatial resolution imagery to map urban areas. This Chapter introduces super-resolution analysis as a method to enhance the spatial resolution of medium resolution sensor (like Landsat data) for subsequent land cover mapping of urban areas in detail demanded by the complexity of these environments.

Tsai and Huang (1984) initially introduced the super-resolution concept. Super-resolution analysis is an image processing technique which recovers one or a series of fine resolution image(s) from one or several coarse resolution images(s) (Peleg and Elad, 2014; Li *et al.,* 2013a). It offers to overcome some of the inherent resolution limitations of relatively low-cost imaging sensors (i.e., coarse satellite imagery, cell phone cameras or surveillance cameras) (Yang *et al.,* 2012).

This chapter reviews the background to super-resolution mapping to provide an underpinning introduction for its subsequent use in this thesis. At first, super-resolution mapping and restitution will be introduced as two main fields of super-resolution analysis. Then, each technique will be elaborated including

their methods. Next, sparse representation will be presented as a state of the art super-resolution method which is capable of enhancing the spatial resolution of medium resolution imagery using a single VHR image. The final section explains how the sparse representation method works. The main concept and assumptions of sparse representation will be explained, with its two phases of training and reconstruction summarised.

4.2 Super-resolution mapping (SRM)

The topic of super-resolution analysis can be categorised into two fields of super-resolution restitution (SRR) and super-resolution mapping (SRM). Unlike the SRR that only enhances the spatial resolution of imagery, SRM is a land cover classification technique that produces thematic classifications at a sub-pixel scale (Atkinson and Aplin, 2004). The SRM methods estimate the spatial distribution of land cover within mixed pixels of remotely sensed images with the proportion of the classes determined from a soft classification technique into a finer scale hard classification map (Foody, 2002; Mertens *et al.,* 2003; Atkinson, 2009). In fact, SRM methods can derive a finer spatial resolution than the size of the pixel of the coarse spatial resolution image (Ricotta and Avena, 1999; Woodcock and Gopal, 2000; Foody *et al.,* 2005).

The Section 2.2.6 discussed soft (fuzzy) classification and explored the mixed pixel problem of pixel-based classification. Although soft classification may be used to estimate the class composition of image pixels, the spatial distribution of each class in these mixed pixels cannot be determined. Tatem *et al.,*(2003)

and Atkinson (2009) described the SRM as a capable method to partially answer the limitation of soft classification and mixed-pixel problem.

4.2.1 Types of SRM methods

Several types of SRM methods have been introduced to enhance medium spatial resolution data for remotely sensed applications through two main groups of spatial correlation and knowledge-based (Boucher, 2009; Ling *et al.,* 2013).

The spatial correlation method exploits the spatial correlation of the fine scale attributes. This method benefits from the first law of geography: "*everything is related to everything else, but near things are more related than distant things*" (Tobler, 1970). However, this assumption may not be necessarily correct under all conditions and observations (Tatem *et al.,* 2001; Mertens *et al.,* 2003, Atkinson; 2009).

Knowledge-based methods rely on the accurate identification of boundary features that separate the mixed pixels into pure components at finer spatial resolution. These techniques use knowledge-based information or *a priori* spatial model (like a variogram) to identify accurately the boundaries of features in which split the mixed pixels into pure components at fine spatial resolution. However, the knowledge-based SRM techniques are merely limited to localise large and homogenous fields with straight boundaries (Muad, 2011).

Several types of SRM have been developed to address the sub-pixel mapping issue, including the Hopfield neural network (Tatem *et al.,* 2001; Tatem *et al.,* 2002; Tatem *et al.,* 2003), the genetic algorithm (Mertens *et al.,* 2003), linear

optimization (Verhoeye and De Wulf, 2002), the Markov random field (Kasetkasem *et al.,* 2005), pixel swapping (Atkinson, 2005, Makido *et al.,* 2007), and spatial attraction models (Mertens *et al.,* 2006; Gu *et al.,* 2008), Neural network predicted wavelet coefficients (Mertens *et al.,* 2004), and Geostatistics indicator (Boucher and Kyriakidis, 2006).

4.2.3 Pros and cons of SRM

SRM methods can generate more accurate and informative land cover maps for medium and coarse spatial resolution images (Tian and Ma, 2011a; Foody *et al.,* 2005; Li *et al.,* 2011). Therefore, they can potentially address the constraint of an adequate level of spatial resolution in medium resolution data sets in many fields of research including mapping complex urban sites (Lu and Weng, 2007; Ling *et al.,* 2011; Ardila *et al.,* 2012).

Because different spatial dependence description methods lead to different sub-pixel land cover maps by considering the relationship between spatial resolutions of pixels and land cover areas (Atkinson, 2009), defining the spatial dependence of land cover classes is the major problem of SRM in which applies for particular types of sub-pixel classification. Additionally, although SRM can effectively increase the spatial resolution of an image, due to the noisiness (i.e., spatial heterogeneity) at fine resolution scale, per-pixel classification methods are not always accurate (Yan *et al.,* 2006). Ling *et al.* (2012) have stated that maximising the spatial dependence does not represent the real world conditions. Next, while SRM improves land cover mapping, SRR merely enhances the spatial resolution of an input coarse resolution image. This means

that, because the output of SRR is the enhanced image, it offers the choice of classifier including any pixel-based or object-based classification algorithms.

4.3 Super-resolution restitution

SRR (also referred as reconstruction) is defined as a technique to construct a fine spatial resolution image from multiple or single coarse spatial resolution images by recovering the high frequency components (like texture and edges) of image content (Yang and Huang, 2010). Due to an inadequate number of observations and the unknown registration parameters, SRR is typically severely ill-conditioned. However, various regularisation techniques have been proposed to address the inversion of this ill-posed problem (Hardie *et al.,* 1997; Tipping and Bishop, 2002; Farsiu *et al.,* 2004[b]). The traditional SRR methods combine complementary information from different images based on sub-pixel shifts or different parameter information, while the more recent SRR methods focus on extracting relationships among neighbourhood pixels or learning prior pattern structures from image databases that store significant amounts of L\H-resolution image pairs (Kasetkasem *et al*., 2005; Jiji and Chaudhuri 2006). SRR methods aim to discover the prior knowledge from the coarse resolution imagery to recover the lost high frequency information. In other words, SRR is a recovery procedure to regain the information lost during image acquisition. For other examples of SRR in remote sensing applications, the readers can refer to following sources (Akgun *et al.,* 2005; Merino Teresa and Nunez, 2007; Hu *et al.,* 2009; Shen *et al.,* 2009).

4.4 Image observation model

Because SRR aims to reconstruct the lost information of image during image acquisition, it is important to understand the procedure for capture image. The SRR method attempts to resolve the inverse problem, which is defined as a linear forward model. Therefore, four issues need to be solved: motion, blur, noise and down-sampling (figure 4.1).

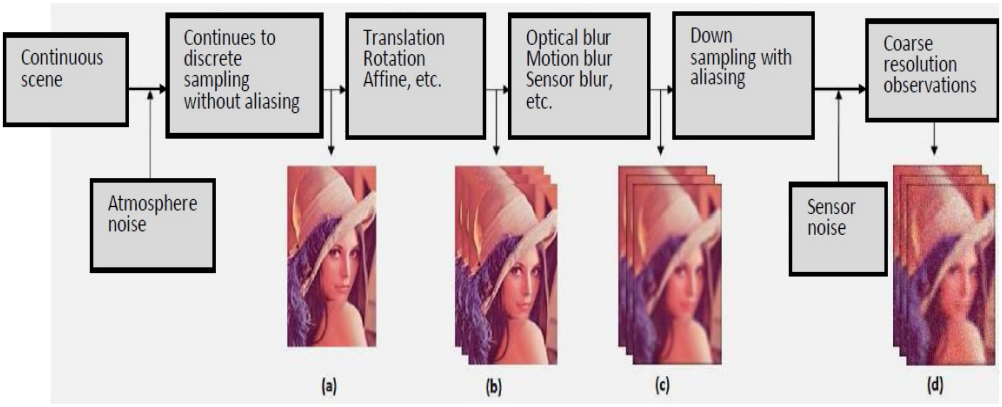


Figure 4.1. Image observation model which demonstrates factors impacting the quality of the image during acquisition (Yang and Huang, 2010). It highlights the key components of signal distortion including discrete sampling, motion, blur, and down-sampling. It is possible to retrieve the high-quality image by inverting this process.

Figure 4.1 depicts the entire process, which occurs during image recording procedure by a sensor as shown. Initially, the continuous scene is converted to a discrete scene composed of pixel units without aliasing (Figure 4.1(a). Because of the technical limitations of the sensor, image aliasing phenomenon occurs which values of adjacent pixels are not continuous, causing the edge of the objects to appear relatively sharp (Figure 4.2).

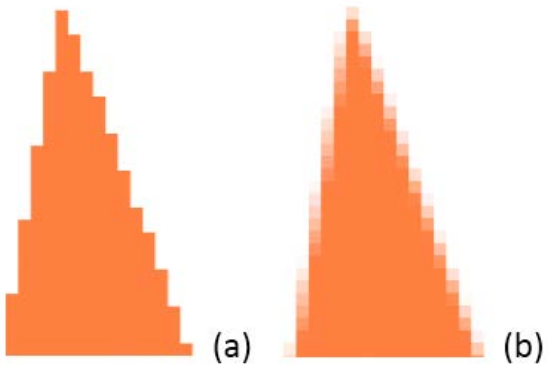


Figure 4.2. Image aliasing: (a) is the condition without aliasing, (b) is the condition with aliasing (Feng, 2016).

It can be assumed that the produced image (Figure 4.1(a)) could be considered as the fine spatial resolution image, which required to be recovered through SRR procedure. In the case of a multi-frame condition in which several images are acquired for the same scene, there is a strong possibility of some tiny motion differences among each image (frame). This movement can occur due to the camera's motions like translation, rotation, and shift (Figure 4.1(b)). Next, as a result of noise and hardware error, the sensor blur can arise during the image acquisition procedure (Figure 4.1(c)). As a final step, the shifted, blurred sampling images can be down-sampled, while sensor noise has been added to the image (Figure 4.1(d)). For a more comprehensive review on image observation model, readers can refer to Farsiu *et al.,* 2004a; Yang and Huang, 2010; Tian and Ma, 2011a.

4.5 Super-resolution and the reconstruction phase

Typically super-resolution restitution (SRR) can be divided into two steps, including image registration or training (depends on types of the method) and reconstruction. The majority of methods developed are based on the

reconstruction phase, which can be separated into the two categories of frequency and spatial domain.

4.5.1 Frequency domain

The frequency domain algorithm was introduced by Tsai and Huang (1984) for Landsat TM imagery. The concept of frequency domain method can be summarised into the following three principles:

- the shifting property of the Fourier transform,
- the aliasing relationship between the continuous Fourier transform of an original high-resolution image and the discrete Fourier transform of observed coarse resolution images (Park *et al.,* 2003), and
- the assumption that an original high-resolution image is bandlimited (Li *et al.,* 2013b).

As an extension of the frequency domain algorithm, Kim *et al.* (1990) developed a method for a blurred and noisy image by assuming all coarse resolution images have the same blur and noise properties. Kim and Su (1993), improved this method and considered different blurs for each coarse resolution image. To reduce the effects of registration errors Bose *et al.* (1993) introduced the recursive total least squares method. Later on, a new method based on discrete cosine transform was developed by Rhee and Kang, (1999) which reduces the memory requirements and computational costs.

The Wavelet transform, a more recent frequency domain method, decomposes signals into components at different scales or resolutions. Nguyen and Milanfar, (2000) stated that this characteristic of wavelet transform method

could help the signal trends at various scales to be isolated and studied. Later on, Robinson *et al.* (2010) proposed a Fourier-wavelet method, whereas Naik and Patel (2013) developed a single-image SRR method in the wavelet domain. The major advantage of this approach is the theoretical simplicity, regarding the relationship between high and coarse resolution images (Park *et al.,* 2003). However, every coarse resolution image is assumed to be affected by a translational motion and a linear space-invariant blur. Though this method is convenient for parallel implementation, which is capable of reducing hardware complexity, the observation model is restricted to only global translational motion and blur (Park *et al.,* 2003). This seems like a limitation of the frequency domain approaches for handling various situations. Because the data are uncorrelated in the frequency domain, it is challenging to apply a priori knowledge for regularisation (Park *et al.,* 2003).

4.5.2 Spatial domain

Initially, super-resolution methods based on the spatial domain were introduced by Stark and Oskoui (1989), known as projection onto convex sets. Typically, this method initially estimates a fine spatial resolution image. Then a series of coarse spatial resolution images simulated corresponding to an input coarse spatial resolution image. Next, those simulated images shift the pixel of an input image in both the horizontal and vertical directions. An initial estimation of fine spatial resolution image is enhanced by computing the difference between the simulated images and the input image. This process is repeated iteratively to minimise the error function. In particular, for a single

image, super-resolution is reduced to a problem of image de-blurring. The spatial domain method converges quickly by fixing stable pixels. In a situation where the pixel value is similar in two successive iterations, the pixel would not be considered in the next iteration which quickens the convergence process (Muad, 2011).

Compared to the frequency method, because of the possibility to involve prior constraints on the solution space and the effect of noise in the data, the spatial domain method improves the quality of the image (Babu and Murthy, 2011). While spatial domain methods offer more flexibility compared to frequency domain counterparts, they required iterative computation which in return requires considerable computation, leading to slow convergence (Babu and Murthy, 2011). Furthermore, because projection onto convex sets optimises a merely constraint-based objective, it does not converge to a unique solution (Park *et al.,* 2003). Regarding the different approaches of acquiring the prior knowledge, super-resolution methods operated in the spatial domain can be divided into two categories: motion and learning-based.

4.5.2.1 Motion-based

Motion-based super-resolution techniques are a group of traditional super-resolution method, which relies on the sub-pixel shifts among input coarse resolution images. They operate by accumulating information on high frequency details (for example, texture and edges) and projecting all the coarse images to a sub-pixel grid (Cristóbal *et al.,* 2008; Tian and Ma, 2011[b]; Hardeep *et al.,* 2013). This approach focuses on multi-frame conditions and uses sub-pixel shifts between each coarse spatial resolution frame to estimate the fine

spatial resolution image. The Motion-based algorithms required conducting motion estimation, which is based on image registration. The success of motion-based methods relies on the quality of image registration. For further super-resolution algorithms based on motion estimation, readers can refer to Bose and Ahuja (2006); Callico *et al.* (2008); Protter and Elad (2009).

Motion-based methods require a series of input images to be captured within a short period. This requirement is essential and significantly increases the chance in which each image is almost the same as the others. However, this requires very fine temporal resolution which is difficult for remotely sensed application with a relatively low temporal resolution. Furthermore, these approaches require significant amounts of training samples so as to include as many patterns for the possible encountered conditions as possible (Feng, 2016).

Typically, motion-based methods can be categorised into two groups: interpolation and regularisation-based approaches. Interpolation-based methods are the first solution of motion-based super-resolution which involve the sub-pixel shifts between frames through an interpolation process. The iterative back-projection and projection onto convex sets are two popular motion-based image interpolations developed according to the input coarse resolution images by a set-theoretic algorithm (Patti *et al.,* 1997).

Though interpolation-based super-resolution techniques are a prevalent option to involve auxiliary information, regularisation-based methods are capable of incorporating the prior knowledge. This is particularly useful since SRR can be treated as an ill-posed inverse problem. According to Bayesian

theory, the extracted information from the coarse spatial resolution images about the unknown fine spatial resolution image is limited in the probability distribution of the unknown fine resolution images (Tian and Ma, 2011[a]). However, the success of regularisation techniques relies on statistical model assumptions.

4.5.2.2 Learning-based

Learning-based methods require prior knowledge which is extracted from training data to improve the spatial resolution of the input coarse resolution image (Chaudhuri and Joshi, 2005; Siu *et al.,* 2012). They generate new information based on generative models (Babu and Murthy, 2011). Training images in learning-based methods can be normal images or a set of image pairs of which the images in each pair cover the same area but at difference spatial resolutions. Learning-based (also referred as example-based) methods are more suitable for remotely sensed applications because they are mainly employed for super-resolution applications using single input images and fewer limitations for data preparation (Yang and Huang, 2010).

The type of learning-based method is defined according to their training procedure. Several learning-based methods treat the training process as an optimisation problem and use various optimisation algorithms including matching pursuit and orthogonal matching pursuit to extract the prior information from the training images (Chen *et al.,* 1989; Mallat and Zhang, 1993; Tropp, 2004). However, others use different algorithms for their training procedure, such as the K-singular value decomposition algorithm (K-SVD) which K represents k-means clustering method (Aharon *et al.,* 2006; Zeyde *et al.,*

2012). Typically, learning-based methods can be divided into three groups: image analogy, neighbourhood embedding, and sparse representation. For further details on super-resolution methods, readers can refer to Park *et al.,* (2003); Pickup (2007); Babu and Murthy (2011); Muad (2011); Feng (2016).

Most methods in the super-resolution field, are strictly reconstruction based in which they perform primarily on a uniform and non-uniform sampling theorems. Table 4.1 represents a result of comparison between frequency and spatial domain in various aspects. While these methods do not create any information, learning-based methods can create new information based on generative models (Babu and Murthy, 2011).

	Frequency domain	**Spatial domain**
Observation model	Frequency domain	Spatial domain
Motion models	Global translation	Almost unlimited
Noise model	Limited	Very flexible
Degradation model	De-aliasing	De-aliasing A-priori information
Super-resolution Mechanism	Low	High
Computation requirements	Limited	Good
A-priori information	Limited	Excellent
Regularisation	Poor	Excellent
App. performance	Limited	Wide
Applicability	Good	Almost unlimited

Table 4.1. A general comparison of frequency and spatial domain methods (Babu and Murthy, 2011).

4.6 Super-resolution and the number of training samples

4.6.1 Multi image and example-based

Super-resolution can be broadly classified into two families including the classic multi-image and Example-based super-resolution. Conventional methods normally require multiple images (also referred as multi-frames) as training (or registration for some methods) phase that are acquired a series of images from the same scene, typically aligned with sub-pixel accuracy (Yang *et al.,* 2008).

Multi-image methods typically refer to video and multi-sensor observation, often employed in surveillance or for forensic purposes, such as when specific object require magnification (for example, licence plates of cars or faces), where their algorithms are often an extension of a single image (Borman and Stevenson, 1998; Chaudhuri and Joshi, 2005). They essentially perform as single image super-resolution on a continuous series of input coarse resolution images.

Due to the possibility to acquire multiple images in medical imaging, multi-image super-resolution is well developed in the medical fields. This particular method employs computed tomography and magnetic resonance imaging, while the resolution quality is rather limited (Park *et al.,* 2003). This method operates in the way which each input image imposes a set of linear constraints on the unknown high-resolution intensity values (for a comprehensive review see Park *et al.* (2003) and Farsiu *et al.* (2004a)). If the number of coarse resolution images meets the requirement (at subpixel shifts), a series of

equations would be determined to recover the high-resolution image. However, this problem is typically severely ill-conditioned due to the insufficient number of observations and the unknown registration parameters (Peleg and Elad, 2014).

Since the majority of remote sensors have long revisit periods, it is a challenging task to obtain a series of images during a short period. Therefore, rather than using a multi-frame approach, the single-frame method is more typically practised for remote sensing (Atkinson *et al.,* 1997; Boucher and Kyriakidis, 2006; Powell *et al.,* 2007; Ge *et al.,* 2009). Regarding the practicality, these methods are numerically limited only to small increases in the resolution, usually by factors smaller than 2 (Baker and Kanade, 2002; Glasner *et al.,* 2009; Peleg and Elad, 2014). According to Yang *et al.* (2012) the training database in conventional multi-images is typically required to encompass a significant number of combinations of high and low-resolution image patch pairs to represent a generic image, precisely which can potentially lead to a computationally intensive process. Finally, due to the possibility of platform vibration and air turbulence, these classical multi-image methods can fail to overcome the irregular local warps and transformation in low-resolution remote sensing images (Li *et al.,* 2013[b]).

4.6.2 Example-based super-resolution

4.6.2.1 Multi images

The Example-based super-resolution has been developed to address the limitations of multi-image super-resolution methods. The aim of example-

based super-resolution techniques are to learn the co-occurrence of local structures between the high and low-resolution image from an external training database (Chang *et al.,* 2004; Sun *et al.,* 2008). Despite using multiple images from similar sensed sub-pixel misalignments, a new form of example-based has been introduced recently that uses a single image as a source of training. These types of super-resolution methods rely on minimum training images, which satisfy most of the remote sensing applications. Single image super-resolution methods went through significant developments by exploiting the self-similarity properties of local image patches across and within different spatial scales in the same image (Zhang *et al.,* 2010; Freedman and Fattal, 2011). Typically, single image super-resolution methods perform a separate deblurring process that is an ill-posed problem (Protter *et al.,* 2009).

4.6.2.2 Single image super-resolution

Bicubic interpolation has been studied as a popular and basic conventional single image linear interpolator. In recent years new methods have been proposed which outperform bicubic interpolation, including image smoothness and geometric regularity of image structures (Zhang and Wu, 2008), gradient profile priors (Sun *et al.,* 2008), self-similarity of image patches within and across different scales in the same image (Freedman and Fattal, 2011; Yang *et al.,* 2013), and sparsity either in the wavelet domain (Mallat and Yu, 2010; Adler *et al.,* 2010), through Gaussian mixture models (Yu *et al.,* 2012) or through an analysis operator (Hawe *et al.,* 2013).

In recent years, as learning-based methods, single image sparse representation has been developed which is more suitable for remote sensing applications

(Huang and Song, 2012; Song and Huang, 2013; Jiang *et al.,* 2014). The interesting point of this method is that the training image can represent the different location from coarse resolution image that is required to be scaled up. Further, there is a possibility this single image super-resolution performs successfully in cases when there is a temporal gap between training and coarse resolution image. Because the training image would be the only source to generate a dictionary, which would be representative of the input coarse resolution image, literature highlights that the training image should hold a similar statistical nature as the coarse resolution images (Yang *et al.,* 2010). For example, this can mean that if the coarse resolution represents complex urban areas formed from various impervious and vegetated features, the training image should process similar textural diversity.

Adler *et al.* (2010) and Peleg and Elad (2014) compared the accuracy of sparse representation compared to Bicubic interpolation. They demonstrated the superiority of using state-of-the-art, sparse representation over bicubic interpolation.

Because the sparse representation computation is mainly based on linear programming, it is reasonably efficient and scalable (Yang *et al.,* 2008). Another benefit of sparse representation is it's computational adaptively to identify the most relevant patches to train the dictionary. This advantage leads to superior accuracy, both qualitatively and quantitatively, compared to methods proposed by Chang *et al.* (2004) who incorporate a pre-defined number of nearest neighbours, generating sharper edges and clearer textures.

Inspired by the recent compressive sensing theories (Cand`es, 2006), Yang *et al.,*(2010) proposed using sparse representation to recover the High-resolution image patch which is capable of generating both photorealistic textures and sharp edges from a single input image. Song *et al.,*(2015[a]) demonstrated the capability of sparse representation research to enhance Landsat TM (30m spatial resolution), using SPOT (with 10m spatial resolution) and a classification scheme with five members. Before moving to sparse representation, the dictionary as a key part of sparse representation will be explained.

The soft (fuzzy) classification is the first step to perform SRM. Because this thesis focus on SRR rather than SRM, the idea of using soft classification in this study dispersed.

4.7 Dictionary

Since the dictionary is an essential part of sparse representation concept, initially the first section reviews the basic concept and the advance in the field of dictionary learning.

A sparse dictionary represents an image structure by holding the most distinctive elements (also called atoms) of the image. In other words, the dictionary is a summarised version of the image, which represents underlying structures of image patches. A dictionary has been considered for sparse representation studies because it is a smaller representation of an image which makes it quicker for computation and fits the sparse representation assumption.

The dictionary learning can be defined as a selection of series of patches from training image or images which hold the most distinct information. The aim of dictionary training is to determine a sparse representation (also referred as sparse coding) of high-resolution training images as a form of a linear combination of dictionary elements. In other words, a dictionary learning problem is a matrix factorisation in which the goal is to factorise a training data matrix that can be represented as the product of a dictionary and a sparse coefficient matrix (training data matrix ≈ dictionary × sparse coefficient matrix) (Sadeghi *et al.,* 2013).

The size of dictionary varies from the small size of 128, 256, 512, to larger as 1024, 2048, and even greater. The choice of selecting dictionary size should be referred to the complexity of training images. While increasing the dimensionality elements, members of the dictionary require more computational facilities; an image with a simple structure (like images of human portray) can be represented by the smaller sized dictionary. However, because it is important that the dictionary represents the structure of the image more realistically, it is recommended to allocate a larger size dictionary for images with more complex structure (like remotely sensed VHR images of urban areas). Figure 4.3, shows an example of trained dictionary with 512 atoms.

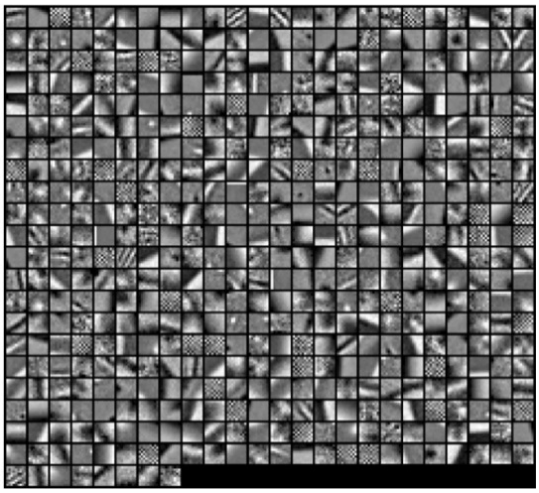


Figure 4.3. A trained dictionary with total 512 atoms which trained with each atom of size 9 (Yang et al., 2010),

4.7.1 Types of dictionaries

Regarding the pre-knowledge on the training process, dictionary types can be categorised into two groups: pre-specified (also referred as non-adoptive) and machine learning (also referred as adoptive) dictionaries. There are several types of pre-specified dictionaries including Fourier, Gabor, Discrete Cosine Transform, and wavelet. These non-adaptive dictionaries operate based on mathematical models of the data. Despite being simple and computationally fast, pre-specified dictionaries are not able to efficiently (sparsely) represent a given class of signals.

On the other hand, machine learning based dictionaries are trained according to training data sets, which belong to the signal class of interest. The dictionary learning field has been through extensive investigation within the last decade (Rubinstein *et al.,* 2010; Tosic, 2011). According to Aharon *et al.,* (2006) training based dictionaries mostly outperform the pre-specified dictionaries in various applications such as image compression and enhancement and classification

tasks (Mairal *et al.,* 2012). Another advantage of training-based over pre-specified dictionaries is that they require a lower number of patches, leading to faster computation (Yang *et al.,* 2008). Furthermore, because trained dictionaries are adaptive to the signals of interest, they are commonly used in state of the art signal recovery tasks, such as denoising (Aharon *et al.,* 2006; Hou *et al.,* 2016), inpainting (Mairal *et al.,* 2008[a]; Filipovic and Kopriva, 2011) and super-resolution (Nasrollahi and Moeslund, 2014).

In contrast to the traditional dictionaries, which are limited to single atoms where the dictionary matrix has the same dimension in the number of signal dimension (rows) and atoms (columns), the over-complete dictionary has been introduced with more atoms than the dimensions of the signal. Adopting over-complete dictionaries offer the flexibility to represent a much wider range of signal phenomena (Rubinstein *et al.,* 2010; Nakanishi-Ohno *et al.,* 2016).

4.7.2 Joint dictionary

As a recent advancement in the field of dictionary training, Yang *et al.* (2010) proposed a joint dictionary training method as over-complete and training-based dictionaries. These joint dictionaries are formed by a small set of randomly chosen high and low-resolution patches, which hold unique information. These high and low-resolution patches are extracted from training images. Joint dictionary learning essentially concatenates the two feature spaces and converts the problem to the standard sparse coding in a single feature space. For example, initially, 100,000 patches are extracted from the training image or image. According to the size of the dictionary, the most useful

patches, which hold distinctive information (like texture), would be chosen for training the dictionary. This number of patches is considerably smaller than that needed by other learning-based methods (Sun *et al.,* 2003) leading to faster computation. Within other training methods, both low- and high-resolution dictionaries are trained separately which increases the required time for the computational process. However, the joint dictionary learning method trains both high and low-resolution dictionaries together with the feature representation and prediction model for image classification in a bilevel formulation, which leads to an optimised computation process (Yang *et al.,* 2012).

4.8 Sparse representation

4.8.1 Sparse representation principle

This section explains the principle and main concepts of performing the sparse representation technique. It starts by describing the basic concept of sparsity, and then the mathematical background of sparse representation technique and its fundamental assumptions will be explained. This research explores using sparse representation method, developed by Yang *et al.* (2010) to recover the lost high-frequency details within input coarse resolution imagery.

Sparsity can be defined as a term of representing some phenomenon with as few variables as possible. The sparsity coding aims to find succinct representations of stimuli, by discovering basic functions which capture high-level patterns in the input data (Lee *et al.,* 2007). Sparsity is represented by input basis vectors in a matrix. These matrices contain only one column in

which most of its members are zero, and the others have small values. Despite a considerable amount of recent research on measuring sparsity, finding spare representation values in the image is still a very difficult computational problem (Lee *et al.,* 2007).

Sparsity principle was mainly considered in signal processing studies in which most of the sparsity methodology has been devised for signal processing. The problems which are related to the signal are often from the same data sources, such as natural images (Mairal and Bach, 2012). In fact, image processing is one of the applications of sparsity in which an image can be represented as a sparse representation over well-chosen dictionary (Elad, 2010).

Sparse and redundant representation modelling has been shown to be a powerful and efficient tool for signal analysis and processing (Sadeghi *et al.,* 2013).

Sparse representation has been used widely in the signal processing field, although this is a fresh topic in image processing. Spare representation has been applied in a variety of recent image processing studies, including: image denoising (Giryes and Elad, 2014; Ojha *et al.,* 2015; Azzari and Foi, 2016), inpainting (Zoran *et al.,* 2011; Chatterjee and Milanfar, 2012; Papyan and Elad, 2016), demosaicing (Menon and Calvagno, 2011), pan-sharpening (Yin, 2015), restoration (Mairal *et al.,* 2008[a]), and Super-resolution (Yand *et al*, 2008; Yang *et al* 2010; Zeyde *et al*, 2011).

4.8.2 Compressed sensing

Sparse representation for image scale-up relies on the principle of compressed sensing theory. This theory assumes that under mild conditions (when training images represent similar structure of coarse resolution image), the sparse representation can be precisely recovered from the down-sampled signals (observation or vector data) (Yang *et al.,* 2010). As a result, linear relationships among high-resolution signals can be correctly recovered from their low dimensional projection (Candes and Tao, 2006). More recent studies have used compressed sensing theory for image scale-up analysis (Zeyde *et al.,* 2012).

4.8.3 Assumptions

Assumption 1: First of all, this method assumes that a coarse resolution image is, in fact, a fine resolution image which has gone through the blur and down-scaling with additive noise (figure 4.4). This assumption in literature refers to reconstruction constraint (Zeyde *et al.,* 2012).

As figure 4.4 describes, sparse representation is an inverse problem which aims to enhance the coarse resolution image (30m) spatially to its fine version (2m). The equation of reconstructing constraint assumption is presented below:

$$x_l = \mathbf{Z\ B}\ y_h + \mathbf{n} \quad \mathbf{(1)}$$

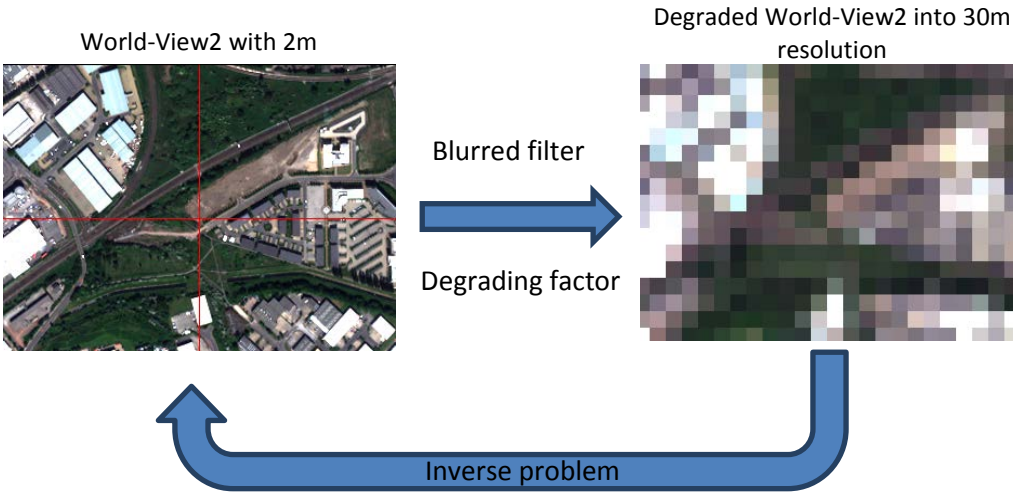


Figure 4.4. The diagram summarises the sparse representation as an inverse problem.

x_l represents a coarse resolution image, while y_h refers to the fine resolution image. B is blur factor applied as a low-pass filter and Z acts as zoom factor by discarding rows/columns from x_l. Finally, n represents an additive zero-mean white and homogeneous Gaussian noise (Cheng, 2015). Next, y_l is generated through as spatial enhancement of x_l to the similar size of y_h. y_l holds spatial details of x_l but with similar pixel size of y_h.

$$y_l = I\,Z\,B\,y_h + Q\,n \longrightarrow y_l = I\,Z\,B\,y_h + n' \quad (2)$$

To prevent potential complexity between x_l and y_h which is created as result of using various resolutions. In addition, Q acts as a simple interpolation operator ($R^{N_l} \longrightarrow R^{N_h}$) to simplify the overall recovery algorithm. It assumes the image z_l is scaled-up by a simple interpolation (e.g. bicubic interpolation) that fills-in the missing rows/columns, returning to the size of y_h.

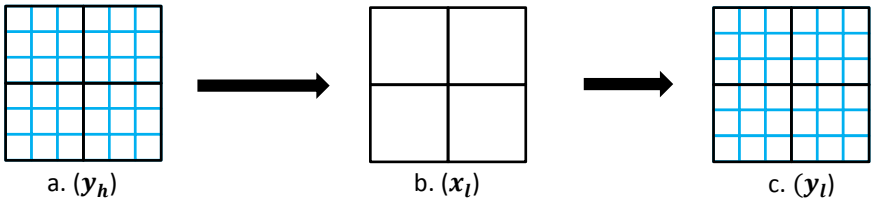


Figure 4.5. a) A fine resolution image (y_h) degraded by zoom factor 3 creating. B) A coarse resolution image (x_l). C). referred as performing bicubic interpolation to up sample it according the y_h scale.

Because the sparse representation operates locally, it performs based on patches.

$$L^{all} = \mathbf{I\ Z\ B} \quad \Rightarrow \quad y_l = L^{all}\, y_h + \mathbf{n'} \quad \mathbf{(3)}$$

Patches are generated as result of subsetting an imagery to a smaller size. Here, each patch is assumed to be represented by multiplying an extracted operator on an image.

$$P_k^l = R_k\, y_l \quad \in R^n \qquad \mathbf{(4)}$$

For example, P_k^l represents a patch from coarse resolution image (y_l) in the location of k and with size of $\sqrt{n} \times \sqrt{n}$ **.** The locations of K are assumed to be the centroids which are surrounded by true pixels within the coarse resolution image. The aim is to estimate the equivalent of spatially enhanced patches from y_h (p_k^h) resolution from patches, which are extracted from x_l (p_k^l).

Assumption 2: According to the sparse signal representation studies conducted by Liu *et al.* (2007) and Mairal *et al.* (2008[b]), the linear relationships among high-resolution signals can precisely recover from their relative low-dimensional projections. This assumption is extended into the field of image processing, in which Donoho (2006) claimed that whether a dictionary meets an appropriate near-isometry condition, for a wide variety of matrices, any

sufficiently sparse linear representation (s_k) of a fine resolution patch (p_k^h) can be recovered (almost) perfectly from the coarse resolution patch (p_k^l).

Assumption 3: The next step introduces the sparsity prior by assuming that the fine resolution image patches (p_k^h) can be well-presented as a sparse linear combination of elements from appropriately over-complete dictionary ($D_h \in R^{n\times m}$) (Yang *et al.,* 2010).

$$\boldsymbol{P_k^h \approx D_h\, s_k}\ , \qquad \boldsymbol{s_k} \in R^n\ and\ \ \|\boldsymbol{s_k}\|_0 \ll \boldsymbol{n} \quad \textbf{(5)}$$

Patches are defined as a small subset of imagery (formula (5)), and according to formula (3) and (4), it can result in that

$$\boldsymbol{P_k^l = L\, P_k^h + n'_k} \quad \textbf{(6)}$$

L defines as a local operator which being apportioned of L^{all} for individual patch, while n' represents the additive noise in relevant patch. Based on sparsity prior assumption, local fine resolution patch can be represented by multiplying its fine resolution over-complete dictionary to local sparse representation vector. If equation (5) ($P_k^h = D_h\ s_k$) would be multiplied by L, it can be result as:

$$\boldsymbol{L\, P_k^h = L\, D_h\, s_k} \qquad \textbf{(7)}$$

Since the coarse resolution patch defined by multiplying local operator into fine resolution patch plus its relevant additive noise (7), it can be as:

$$P_k^l = L\, P_k^h + n'_k \ (7) \ \Longrightarrow\ L\, P_k^h = P_k^l - n'_k \ (8) \quad \text{and} \quad L\, P_k^h = L\, D_h\, s_k$$

$$\Longrightarrow\ \boldsymbol{L\, D_h\, s_k = P_k^l - n'_k} \ \Longrightarrow\ \|\boldsymbol{P_k^l} - \boldsymbol{L\, D_h\, s_k}\|_2 \le \varepsilon \quad \textbf{(9)}$$

ε is a proportion of n'_k.

It concluded with above formula (9) that, coarse resolution patch (P_k^l) can be defined by multiplying the sparse vector (s_k) into local operator (L) into fine resolution over-complete dictionary (D_h).

$$\boldsymbol{P_k^l \approx L D_h s_k} \quad \textbf{(10)}$$

Assumption 4: Another key assumption is that sparse representation within the low-resolution patch is similar to its corresponding high-resolution patch in mild conditions (Yang *et al.,* 2010). It means a local sparse vector is the same for both fine and coarse resolution for location k.

$\boldsymbol{P_k^h = D_h s_k}$ and $\boldsymbol{P_k^l = D_l s_k}$

$P_k^l \approx L D_h s_k$ (10)

$P_k^l = D_l s_k$ (5)

$\Big\}\ L D_h s_k \approx D_l s_k \implies \boldsymbol{D_l \approx L D_h}$ **(11)**

(9) & (11) $\implies$ $\boldsymbol{\|P_k^l - D_l s_k\|_2 \leq \varepsilon}$ **(12)**

Equation (12) explained that recovering the sparse representation for each k locations was computed through its corresponding coarse resolution patch (p_k^l) and the coarse resolution dictionary (D_l). Because it assumed that sparse vector is similar for both coarse and fine resolution, fine-resolution patches can be recovered by multiplying its corresponding sparse vector (s_k) with the fine resolution dictionary (D_h). This is a main core of using sparse representation for super-resolution approach from a single fine resolution image Yang *et al.* (2010).

4.8.4 Sparse representation algorithm

This part explains the structure of the sparse representation algorithm. The sparse representation algorithm used in this study consists of two main phases:

training and reconstruction. Within the training phase, a fine resolution image was down-scaled to computing parameters such as: D_l and D_h. Then, for the reconstruction phase, fine-resolution patches (p_k^h) were recovered through D_h and by sparse representation, which were computed from extracted coarse resolution patches (p_k^l) from the input coarse resolution image. Both the training and reconstruction phases are explained in more detail in the following sections.

4.8.4.1 Training

Training is the first step of performing the sparse representation method. It involves collecting a predefined number of patches from a single fine resolution image. The aim of this stage is to generate two coupled dictionaries D_l and D_h which will be trained from $y_l^{'}$ and $y_h^{'}$ data sets respectively. D_l represents the dictionary for the coarse resolution image, while D_h offers the dictionary for fine resolution image. Selecting the correct choice of training image, is a critical step for the sparse representation method. Since both fine and coarse dictionaries will be used in the construction phase to estimate the lost texture and the edges of coarse resolution image. The Figure 4.6 describes the main steps involved in the training phase.

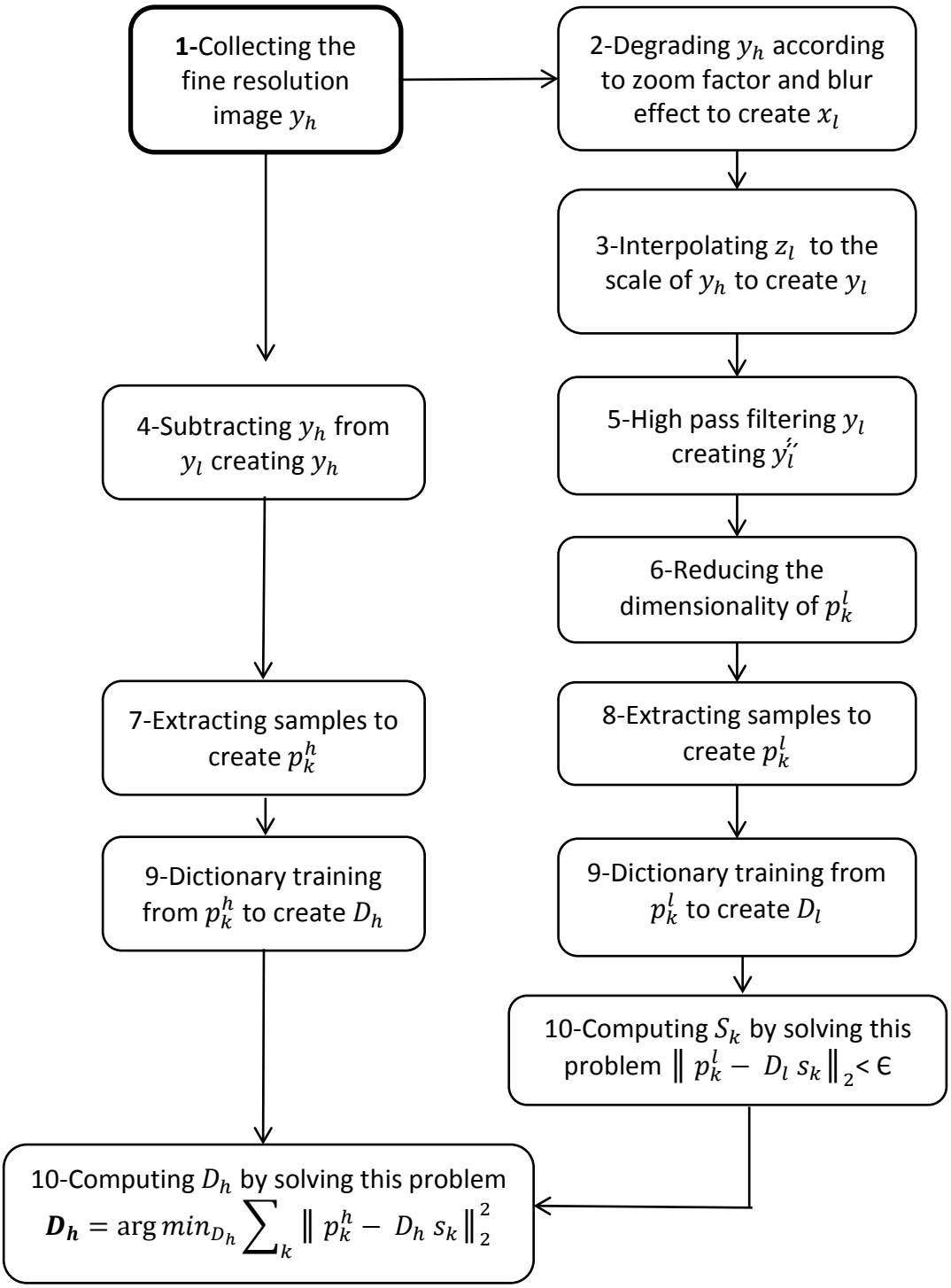


Figure 4.6. The workflow of the training phase.

The key point for selecting appropriate training images is the statistical similarity in the structure between the input coarse resolution image and the training image. In other words, the designated training image should offer

similar texture variety and complexity to that of the coarse resolution image. In this case, with a medium resolution image of a complex urban area that required spatial enhancement, the training image should be from an urban area offering similar land cover types (step 1 in Figure 4.6 producing Figure 4.10).

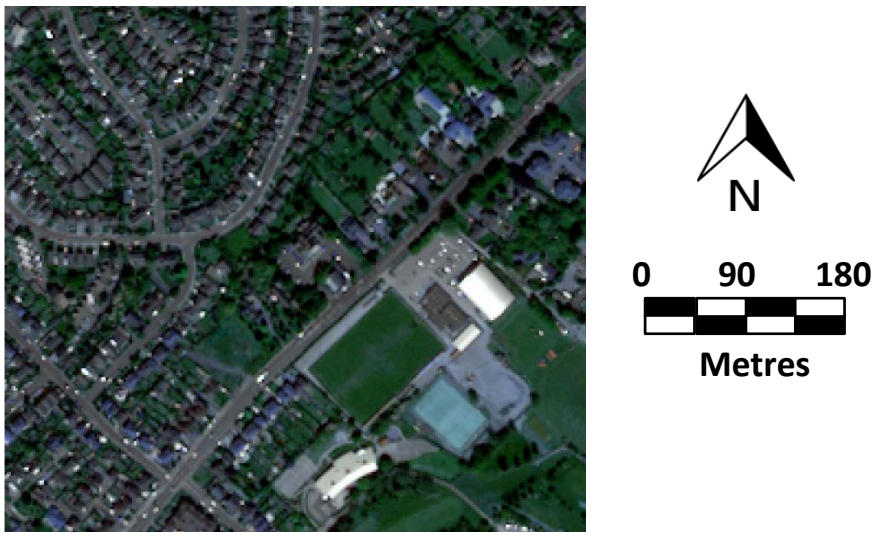


Figure 4.7. The true RGB colour of training image with 2m spatial resolution.

Each step of performing the sparse representation was demonstrated through a figure. The Figure 4.7 represents the true colour of training image (WV-2) with 2m spatial resolution. According to the reconstruct constraint assumption (equation 1), a fine resolution training image is degraded to the same resolution of the input coarse resolution image (in reconstruction phase) to form the corresponding coarse resolution image (x_l) (step 2 in Figure 4.6). The figure 4.8 represents the degraded version of WV-2 imagery with 30m spatial resolution.

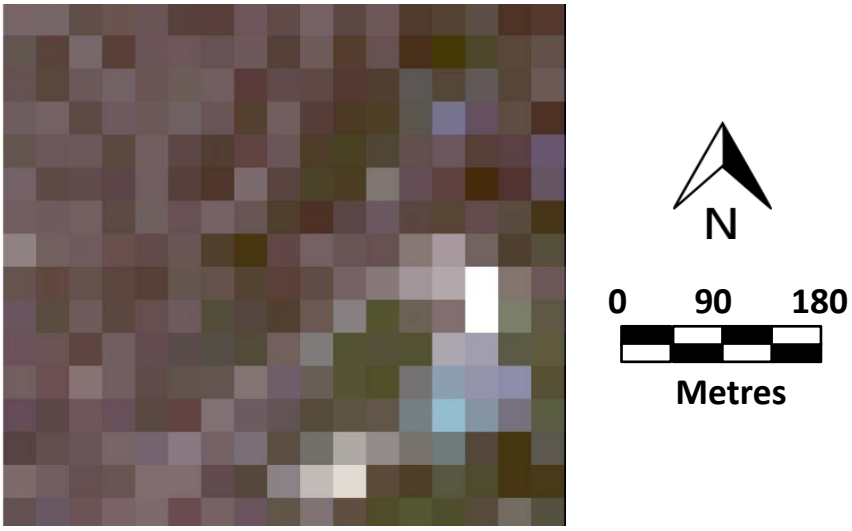


Figure 4.8. The true RGB colour of degraded 2m into 30m.

x_l scaled up into its corresponding fine resolution image (y_h) (equation 2) using Bilinear interpolation technique (Figure 4.9). This would produce y_l offering the x_l information, while it has the same resolution as y_h. This up-scaling procedure makes sure that similar patches size can be extracted as patch pair extractions from both fine (y_h) and coarse resolution (y_l) (step 3 in Figure 4.6).

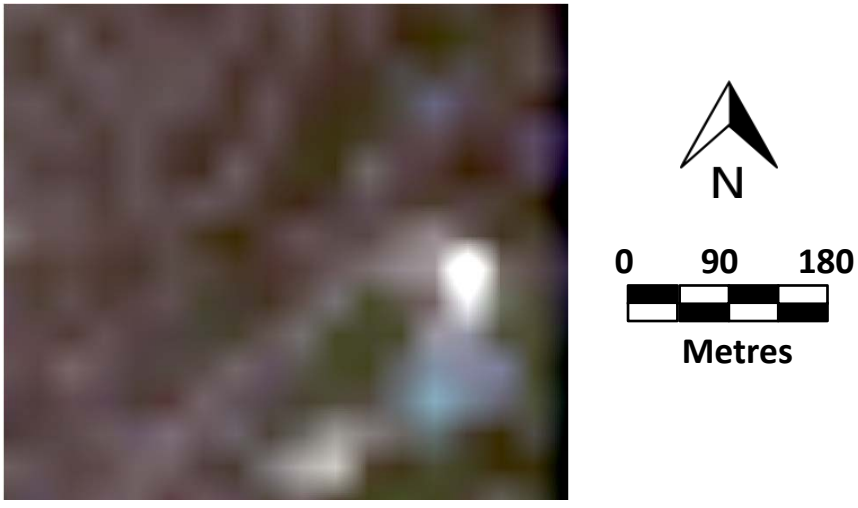


Figure 4.9. The true RGB colour of bilinear interpolation 30m to 2m.

The sparse representation method designed to characterising the relationship between coarse resolution image (y_l), and high frequency information (including the edge and texture) of its corresponding fine resolution image (y_h) (Cheng, 2015). Dictionaries are trained through subtracting digital number value of each pixel from the average value of the relevant patch rather than actual digital number values. So, the next step is to extract these residual values from both y_l and y_h. As a result the fine resolution imagery is subtracted from the interpolated coarse resolution image ($y_h^{'}$ = y_h - y_l) (step 4 in Figure 4.6).

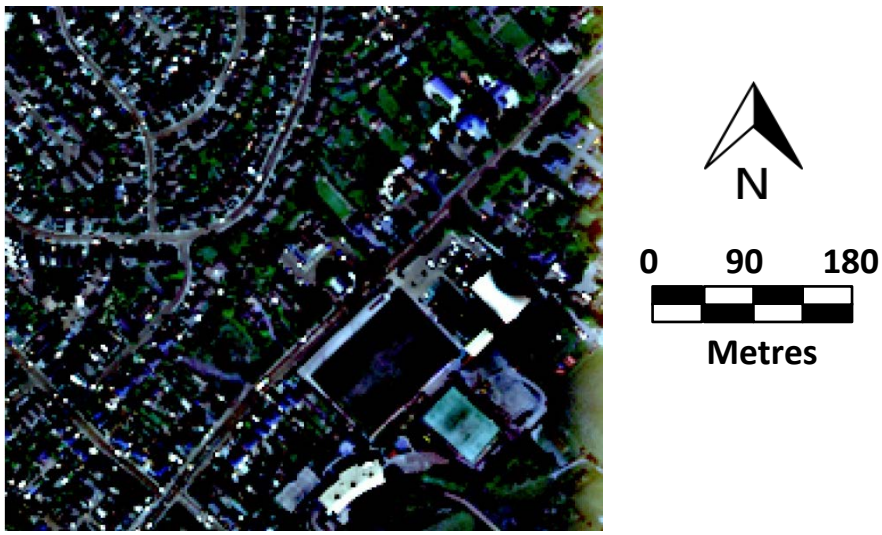


Figure 4.10. The true RGB colour of subtracting $Y_h - Y_l$.

Next, high pass filtering performed on coarse resolution images (y_h) to extract features which related to their high frequency contents. In fact, coarse resolution dictionary D_l, sparse representation vector (S), and fine resolution dictionary (D_h) are computed according to those residual values from the coarse resolution image ($y_h^{'}$). They are feasibly offering the most important information to estimate the lost high-frequency content in the final fine

resolution image (Yang *et al.,* 2010). Therefore, $y_h^{'}$ and $y_l^{'}$ represents the source images for training dictionaries as fine and coarse resolution respectively (step 5 in Figure 4.6).

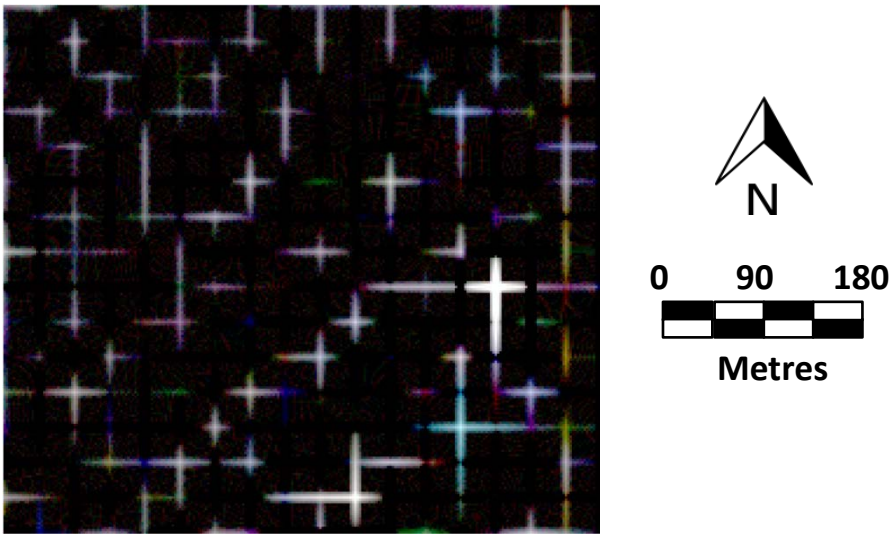


Figure 4.11. The true RGB colour of high pass filtering the y_l.

Just before training dictionary, principle component analysis technique is performed to reduce the dimensionally of $y_l^{'}$ (filtered, interpolated and degraded y_l), while it preserve 99.9% of actual values. $y_l^{'}$ reduced the computational needs of coarse patches for both training and reconstruction phases (step 6 in Figure 4.6).

Then as the next step, patch pair extraction was performed to select a predefined number of patches randomly from both coarse and fine data sets. These patches are extracted from the same location of k for both data sets (step 7 in the Figure 4.6). Although a simple training approach can be selecting perfect patch pairs which preserving the correspondence information between coarse and fine resolution data sets (Yang *et al.,* 2010), this approach generates

large dictionaries requiring expensive computation. The strategy used in this method was to select a small number of patches holding the most valuable information.

The dictionary training is designed in a way which both coarse and fine patches have the same sparse representation. Yang *et al.* (2010) stated that jointly training both coarse and fine dictionaries enforce sparse representation similarity between patched pairs from the coarse and fine resolution images in their dictionaries. Readers can refer to Lee *et al.* (2007) for more detailed explanations on training algorithm of these over-complete dictionaries (step 8 in Figure 4.6).

The next step involves computing the sparse coding. The sparsity of fine resolution patches does not directly compute from the fine resolution patches. According to the equation 14 ($\|P_k^l - D_l\, s_k\|_2 \leq \varepsilon$) sparse vector (s_k) is computed merely according to coarse resolution patches (P_k^l) and an over-complete coarse resolution dictionary (D_l). The result of Donoho's (2006) research in sparse signal representation showed that the linear relationships among H-resolution signals can be accurately recovered from their low-dimensional projections. According to theoretical outcomes of compressed sensing theory, the sparse representation can be correctly recovered from their down sampled signals, under mild conditions (Yang *et al.,* 2010). Moving from signal to image, it is assumed that sparse representation is the same for both coarse and fine resolution. As a result, sparse representation which computed from coarse resolution patches (s_k) directly was used to recover the

corresponding fine resolution patch by considering D_h (Yang *et al.,* 2010) (step 9 in Figure 4.6).

According to the Equation 5 ($P_k^h \approx D_h\ s_k$), the fine resolution patches can be estimated by multiplying fine resolution dictionary D_h into their sparsity vector. Therefore, D_h is defined to be the solver of problem below (Cheng, 2015).

$$D_h^{'} = \arg min_{D_h} \sum_k \| p_k^h - D_h\ s_k \|_2^2$$

p_k^h represents fine resolution patches from training image ($y_h^{'}$), D_h acts as initial trained fine resolution dictionary and s_k is sparse representation measured from coarse resolution (step 10 in Figure 4.6).

4.8.4.2 Reconstruction phase

As the second phase of sparse representation, the reconstruction phase aims to scale-up the input coarse resolution image according to both coarse and fine resolution trained dictionaries. Referring to the first assumption, it assumed that this coarse resolution image is a result of performing blur effect and down-sampling operator. In the training phase, a series of steps is carried out by down-scaling a spatial settlement of an input fine resolution training image. Despite all computed parameters, both coarse and fine resolution dictionaries (D_l and D_h respectively) will be used in the construction. The Figure 4.2 describes the main steps are involved in training reconstruct phase.

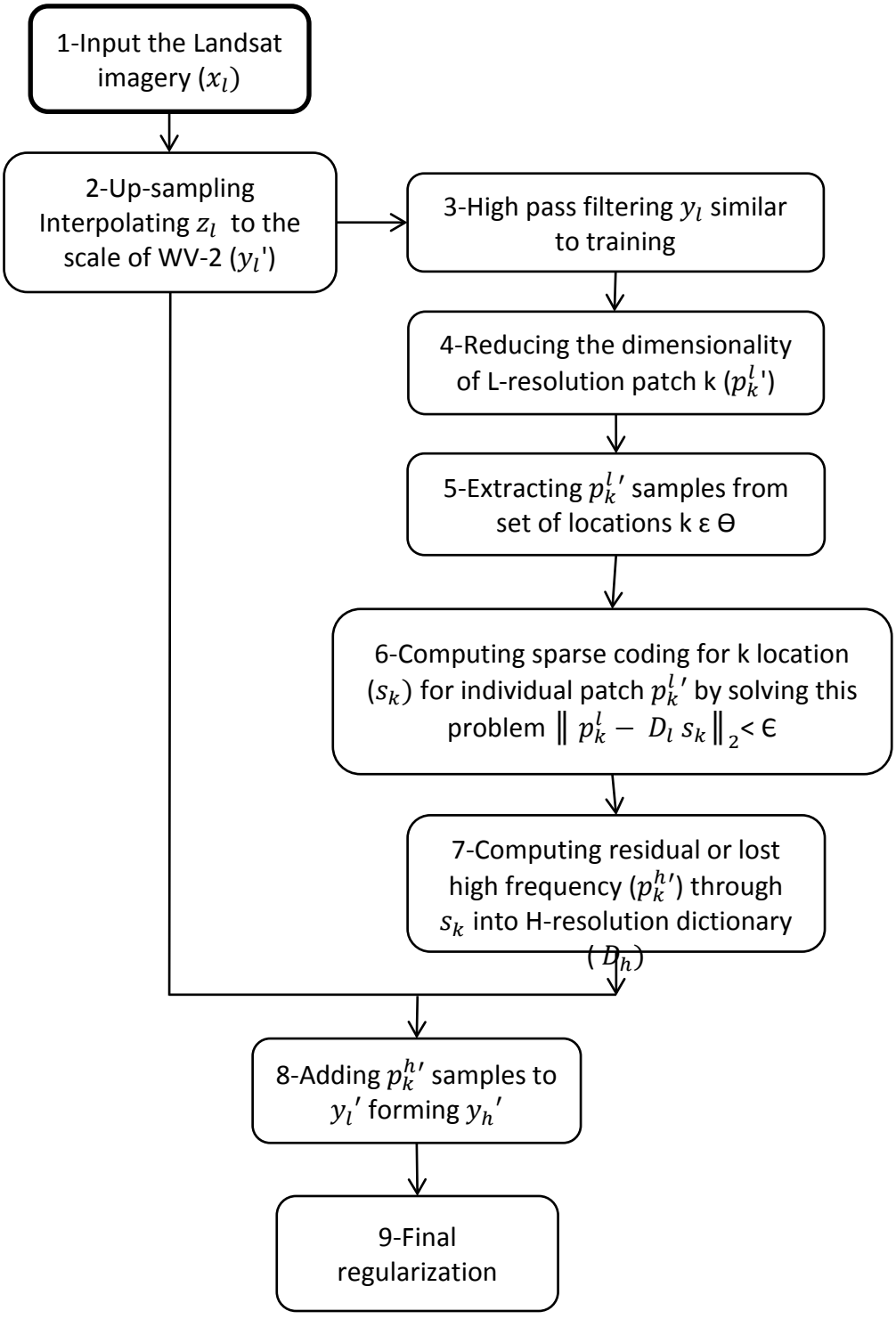


Figure 4.12. The workflow of the reconstruction phase.

As the first step of reconstruction phase, a coarse resolution image (x_l) is collected as an input data source phase (step 1 in Figure 4.12).

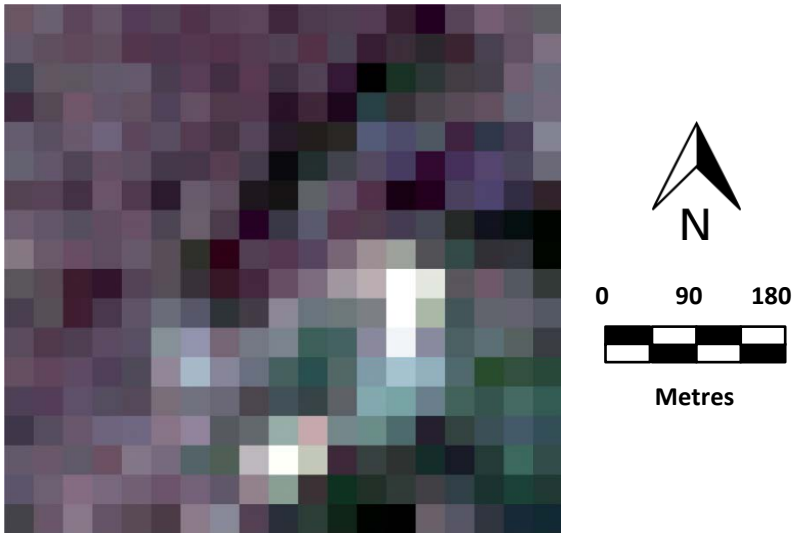


Figure 4.13. The true RGB colour of Landsat OLI with 30m (x_l).

The next step is to up-sample the x_l by performing bicubic interpolation technique. This creates an image (y_l) with finer pixel size that meets final enhanced image scale. Besides estimating lost high frequency, this up sampled image (y_l) would be the base for final enhanced image. In fact, those estimated high frequencies will be added to y_l to form final spatially enhanced image (step 2 in Figure 4.12).

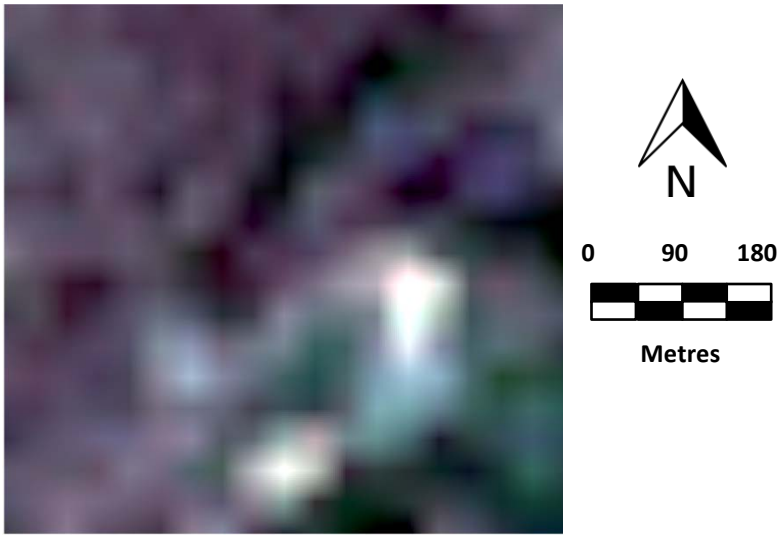


Figure 4.14. The true RGB colour of up-sampled Landsat 30m to 2m (y_l).

This step highlights the texture and edge from y_l by performing a similar high pass filter used in training image (step 3 in Figure 4.12).

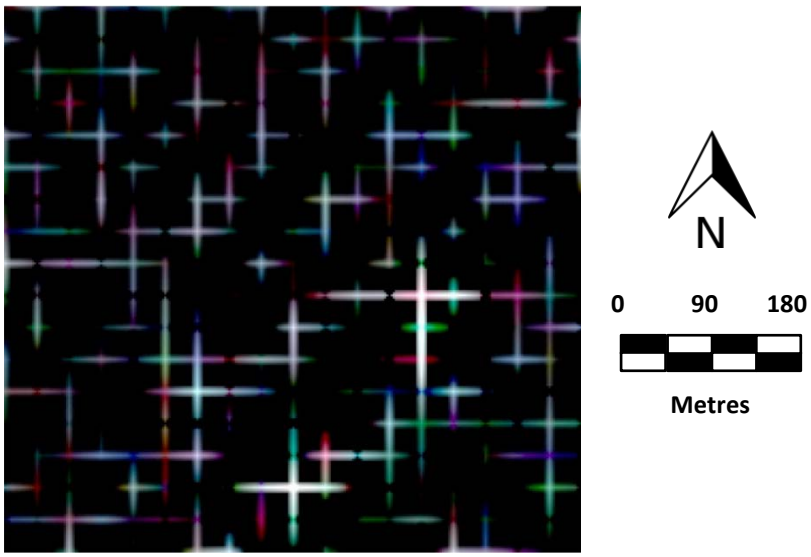


Figure 4.15. The true RGB colour of Landsat OLI high pass filtered with 2m resolution (y_l').

Then principal component analysis is performed to reduce the dimensionality of the extracted patches (step 4 in Figure 4.12).

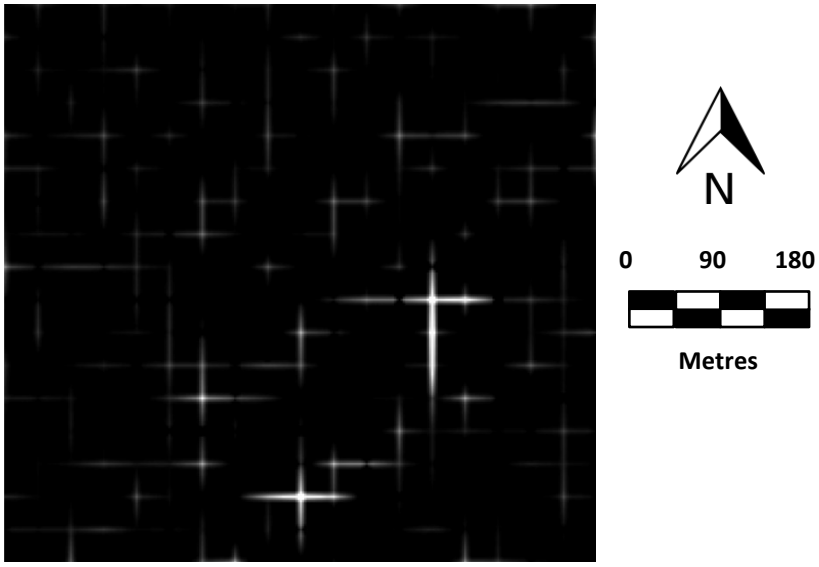


Figure 4.16. The true RGB colour of PCA on filtered Landsat OLI with 2m resolution ($y_l^{\prime}$).

This step extracts the patches (p_k^l) from k locations of filtered resampled coarse resolution images ($y_l^{\prime}$). Each K location (k ε Θ) corresponds to the same location from former K location in training (step 5 in the Flowchart 4.2).

Now, it is time to compute the sparse representation for each extracted patch (p_k^h). Similar to the training phase, sparsity can be calculated by solving equation 14 ($\|P_k^l - D_l\, s_k\|_2 \leq \varepsilon$). Here, similar D_l learnt in the training phase is going to be used (step 6 in Figure 4.12).

Within this step, the initial result of recovered fine resolution is computed. Referring to equation (5) ($P_k^h \approx D_h\, s_k$), fine resolution patch is equal to multiplying the sparse representation (s_k) into fine resolution dictionary (recall, D_h is another of training phase outcome) (step 7 in the Figure 4.2).

As has been mention, this technique performs based on the residual information. So, these recovered high frequencies added to up-sampled image (y_l) in order to construct final spatially enhanced image (step 8 in Figure 4.12).

As a final step, global optimisation is performed to remove the possible reconstruction error. That can reassure that the recovered image is more consistent and natural (Yang *et al.,* 2010) (step 9 in Figure 4.12).

Since the sparse representation works based on a single training image, this novel method can also be used to increase the spatial resolution of remotely sensed data sets. The following Chapter will assess the potential of the sparse representation method to increase the accuracy of urban land cover mapping using Landsat data.

Chapter 5: Using Sparse Representation to Enhance Landsat OLI Imagery Using A Single WorldView-2 Image

5.1 Overview

The main message of Chapter 3 was the advantage of using very fine spatial data to map complex urban areas accurately. It concluded that using VHR imagery (such as WV-2 with 2m spatial resolution) should result in a more accurate classification of the land cover of complex urban areas when compared to classification with coarser spatial resolution data. However VHR data come with their limitations (which have already been explained) and to constrain analyses to these data would be wasting the plethora of information contained in remotely sensed data with a coarser resolution (for example, acquired by the different Landsat series sensors). This chapter, therefore, again using the urban area of Nottingham as a testing set, explores the potential of a sparse representation method to increase the spatial resolution of medium resolution data, such as that provided by Landsat sensors. The key here is to be able to use a single Landsat scene and to apply a novel sparse representation analysis method developed by Yang *et al.* (2010). In this chapter, the sparse representation method was employed to increase the spatial resolution of a coarse resolution image using a single VHR imagery (Section 4.8 described the concept of sparse representation method in detail). It is expected that by performing sparse representation on a Landsat OLI image of the urban area of

Nottingham using a single VHR training image (of WV-2), the accuracy of the subsequent land cover mapping using the spatially enhanced Landsat OLI will be increased over that using the original spatial resolution Landsat data. Although, in this case, the training image is spatially and temporally coincident to the image which is being enhanced spatially, and subsequently being used in the classification, if successful this paves the way for application beyond the coincident data, both in space and time.

In this chapter, the spatial resolution of Landsat OLI imagery was enhanced to 2m, using two different strategies. In the first strategy, the sparse representation method was performed directly on Landsat OLI, which increased the spatial resolution of Landsat OLI from 30 to 2m. In the second strategy, the sparse representation method was performed on the pan-sharpened Landsat OLI. In this strategy Landsat OLI multispectral imagery was pan-sharpened using its panchromatic data as a pre-requisite step before performing sparse representation method.

Another point which distinguishes this research from any other work, is the level of magnification (zoom factor). Previous studies have adopted a zoom factor up of to three (Yang *et al.,* 2010). Enhancing the Landsat OLI imagery from 30 to 2m directly means enhancing the spatial resolution by 15 times, significantly more than in any previous work. However, improving the pan-sharpened Landsat OLI from 15 to 2m means enhancing the spatial resolution by 7.5 times, which seems more feasible for sparse representation method. Due to the advantage of fusing the sparse representation with the pan-

sharpening method, it is expected that the second strategy will provide the most accurate result.

In this Chapter, the spatially enhanced Landsat OLI data will be used in a land cover classification of the Nottingham area (in the same way as Chapter 3) with consideration of both non-parametric pixel-based (SVM) and object-based (OBIA/SVM) classifiers.

This chapter is organised as follows: Section two explains the research material including the study area and the data sets which were used. Section three explains the two strategies which will be performed to enhance Landsat OLI spatial resolution. Section four explores the pan-sharpening techniques and the assessment approaches implemented. Section five explains the operation of the sparse representation in four steps including pre-processing, training, reconstruction, and post processing. Section six defines the classification scheme and details of the performed SVM and OBIA/SVM classification procedures. Section seven describes the accuracy assessment and statistical test procedure carried out on the resultant land cover maps. Section eight discusses the results and finally, section nine summarises this chapter.

5.2 Research Materials

5.2.1 Study area

Following chapter three, Nottingham was selected as the study area to assess the capability of sparse representation method for spatial resolution enhancement. Figure 5.1 shows the Landsat OLI data for Nottingham as the area of interest (please refer to Figure 3.1 for WV-2 of Nottingham). Both

Landsat OLI and WV-2 multispectral data sets cover the same geographic area of Nottingham at 121km^2 (11 x 11 km).

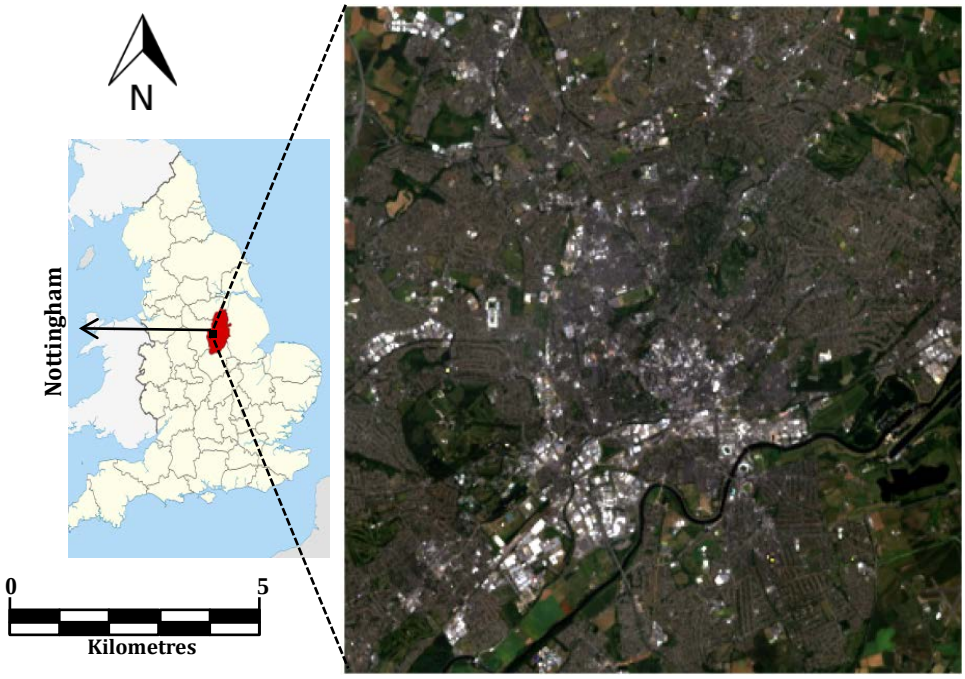


Figure 5.1. The Landsat OLI imagery of Nottingham as the study area,

5.2.2 Data

Landsat OLI offers eight multi-spectral bands with 30m spatial resolution and a panchromatic band with 15m spatial resolution (Table 5.1).

The panchromatic band of Landsat OLI with 15m resolution was used as finer resolution band in the pan-sharpening procedure. The eight multispectral bands of Landsat OLI with 30m spatial resolution used as the input coarse resolution data to be spatially enhanced within the reconstruction phase of the sparse representation method.

Band number	Band name	Spectral range	Spatial resolution
Band 1	Coastal aerosol	430 – 450 nm	30m
Band 2	Blue	450 – 510 nm	30m
Band 3	Green	530 – 590 nm	30m
Band 4	Red	640 – 670 nm	30m
Band 5	Near Infrared (NIR)	850 – 880 nm	30m
Band 6	Short-wave Infrared (SWIR) 1	1570 – 1650 nm	30m
Band 7	Short-wave Infrared (SWIR) 2	2110 – 2290 nm	30m
Band 8	Panchromatic	500 – 680 nm	15m
Band 9	Cirrus	1360 – 1380 nm	30m

Table 5.1. Spectral properties of Landsat OLI with their corresponding spatial resolution (USGS, 2015).

WV-2 imagery with eight multispectral bands at 2m spatial resolution (see Figure 3.2, of Section 3.2.2) was used as the training data source for the sparse representation method. Additionally, since WV-2 offers a resolution at VHR scale, it acted as an accurate benchmark for subsequent analysis.

One of the key requirements of operating the sparse representation method is spectral bands similarity between images in both the training and reconstruction phases. In other words, spectral bands of WV-2 can act as training for the same corresponding spectral band in Landsat OLI. In this case, though both Landsat OLI and WV-2 offer images with eight multispectral bands, their spectral similarity is limited to five spectral bands (Figure 5.2) including coastal blue, blue, green, red, and near-infrared (NIR). Since the coastal band is designed for studying coastal and aerosol content, it was not deemed suitable for use here. It should be mentioned that three spectral bands including green, red, and NIR are not identical. In fact, these three spectral

bands receive the reflected radiation in the wider spectral channel in WV-2 compared to Landsat OLI. Since sensors do not necessarily follow similar protocols regarding their bands' spectral ranges, this dissimilarity was expected. Therefore, only the blue, green, red, and near-infrared (NIR) bands of Landsat OLI and WV-2 matched.

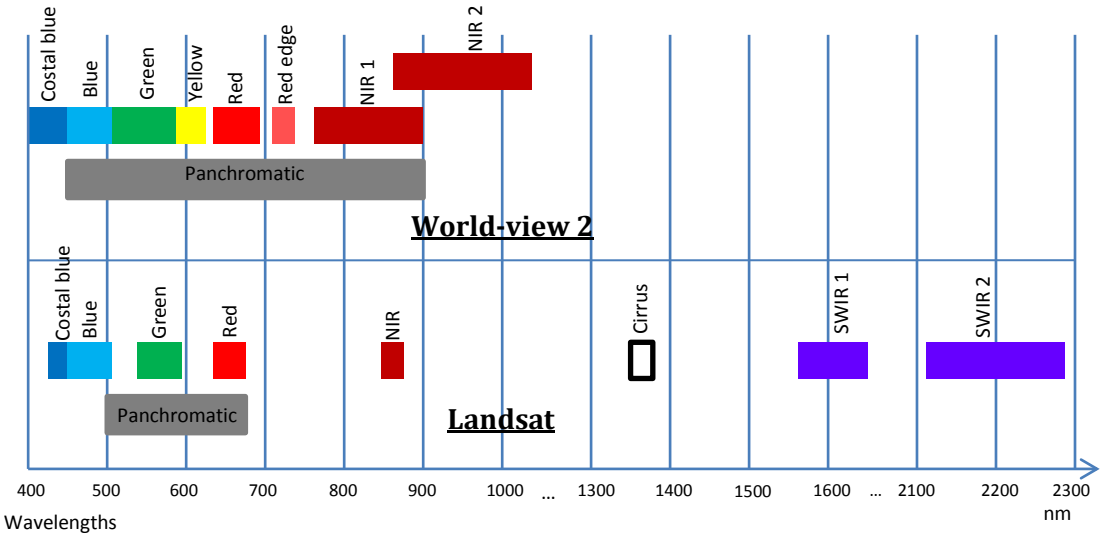


Figure 5.2. Spectral sets visual demonstration for Landsat OLI and WV-2,

Another point worth consideration is the one-year temporal difference in acquisition date of the WV-2 and Landsat OLI imagery. While the WV-2 data were captured on 26th May 2012, the OLI image was acquired on 15th July 2013. However, it is expected that the effect of imperviousness change and phenology are limited since there is only one-year gap between images and also, both images belong to a similar season. As a result, both images should contain similar land cover types.

Since the study area in this chapter is identical to chapter three, the reference data (i.e., for training and testing data) in Chapter 3 was modified according to new classification system with the eight classes.

5.3 Landsat OLI spatial enhancement strategies

The first strategy to increase the spatial resolution of the Landsat data from 30 to 2m relied on using only the sparse representation method, which was directly performed on the 30m Landsat scene. Besides adopting a zoom factor of 15 ($\frac{30\mathrm{m}}{2\mathrm{m}}$), sparse representation was also performed on pan-sharpened Landsat OLI, as an alternative strategy. In this strategy a zoom factor of 7.5 ($\frac{15m}{2m}$) was adopted. This study adopted a zoom factor of 15 and 7.5 for first and second strategy, respectively. Within the first strategy, a pixel with $900\mathrm{m}^2$(30m x 30m) coverage was replaced by 225 pixels with $4\mathrm{m}^2$; while in the second strategy a pan-sharpened pixel with $225\mathrm{m}^2$ (15m x 15m) coverage was replaced by 56 pixels with $4\mathrm{m}^2$. This comparison shows the difficulty of sparse representation to spatially enhance Landsat OLI compared to pan-sharpened Landsat OLI.

Pan-sharpening is a conventional image fusion technique synthesising a panchromatic band to other multispectral bands with a coarser resolution (Ghahremani and Ghassemian, 2016). In other words, the pan-sharpened image is comprised of multispectral bands with a spatial resolution of the panchromatic band. However, the common disadvantage of performing pan-sharpening is spectral distortion which smooths the spectral values (Jalan and Sokhi, 2012; Marcello *et al.,* 2013).

5.4 Pan-sharpening Landsat OLI

The strategy to find the optimum pan-sharpening technique in this research was to test several popular pan-sharpening techniques, which are accessible through commercial geospatial analysis software and have been recommended by recent studies.

The five pan-sharpening techniques employed were: (i) wavelet resolution merges (Yocky, 1996); (ii) hyperspectral colour sharpening (HCS) (Ghosh and Joshi, 2013); (iii) Hue-intensity-saturation (HIS) (Samadzadegan *et al.,* 2012; Khan *et al.,* 2008); (iv) brovey (Basaeed *et al.,* 2013); and (v) Gram-Schmidt (Makarau *et al.,* 2012; Palubinskas, 2013). Although many pan-sharpening techniques are designed to improve single spectral band images, these pan-sharpening methods were chosen as they can handle a multispectral image as an input image and produce a multispectral pan-sharpened image.

5.4.1 Assessing the pan-sharpening quality

The most effective pan-sharpening technique should improve the spatial resolution of the input coarse resolution imagery, while preserving spectral properties of that original image. Recent literature on pan-sharpening assessment suggests that the combination of qualitative and quantitative techniques should be used to test the quality of the pan-sharpened image (Vaiopoulos, 2011; Civco and Witharana, 2012; Marcello *et al.,* 2013). As a result, to assess each of the five pan-sharpening techniques, both visual and quantitative metrics were employed.

Quantitative assessment incorporates metrics measuring the quality of the pan-sharpened image's radiometric resolution and how well each pan-sharpened image has preserved its original spectral properties (Marcello *et al.,* 2013). On the other hand, visual assessment as a form of qualitative evaluation is capable of identifying sharpness of the edges in the image. Strait *et al.* (2008) pointed out that visual assessment is suitable for spatial quality assessment, including geometric shape and the size of objects.

Despite several comprehensive studies to develop more effective quality assessment indicators, researchers have not yet agreed on a unique and common quantitative metric (Meenakshisundaram, 2005; Zhang, 2008; Mitchell, 2010). Therefore, five quantitative metrics were chosen to evaluate the quality of five pan-sharpened techniques. These five metrics include the:

(i) Difference in variance (DIV) index which is based on the variance of pan-sharpened image and reference image which represented as σy^2 and σx^2 respectively.

$$\text{DIV} = \frac{\sigma x^2}{\sigma x^2} - \frac{\sigma y^2}{\sigma x^2} = 1 - \frac{\sigma y^2}{\sigma x^2} \qquad (1)$$

Where $\sigma_x{}^2$ is the variance of the original image and $\sigma_y{}^2$ is the variance of the pan-sharpened image. This index presents the decrease or increase of information content during pan-sharpening process. The interesting point for DIV is that it identifies any changes on the pan-sharpened image regarding information loss or gain. A positive value ($\sigma_x > \sigma_y$) depicts an information loss while, additional information is represented by a negative value ($\sigma_x < \sigma_y$)

(Meenakshisundaram, 2005). Zero represents the ideal value for DIV where both image are similar ($\sigma_x = \sigma_y$).

(ii) The correlation coefficient (CC) which is the most commonly used index measuring correlation between pan-sharpened and the reference image. Besides estimating the global spectral quality, CC quantifies the similarity between pan-sharpened and the reference image (Marcello *et al.,* 2013).

$$\text{CC} = \frac{\delta xy}{\delta x\ \delta y} \qquad (2)$$

The CC value range from -1 and 1, while 1 indicates the most accurate value.

(iii) ERGAS is a French acronym for Erreure Relative Globale Adimensionnelle de Synthese (dimensionless global relative error of synthesis). This index measures spectral distortion in the image by considering the spatial resolution of the pan-sharpened image (Strait *et al.,* 2008). It encompasses relative average spectral error's (RASE) formula by involving pan-sharpened and reference spatial resolution ratio. A lower value for ERGAS value means the finer quality of the pan-sharpened image.

$$\text{ERGAS} = 100\,\frac{P}{M}\,\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{\text{RMSE}\,(band_i)^2}{\bar{x}_i{}^2}} \qquad (3)$$

P is the spatial resolution of the panchromatic band while *M* depicts spatial resolution of the multispectral image. $\bar{x}_i{}^2$ is the mean of the number of band and $band_i$ referred as root mean square error (RMSE). The following equation computes RMSE:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \qquad (4)$$

n is the number of pixels.

(iv) Relative average spectral error (RASE) which characterises the average accuracy of the pan-sharpened image (Choi, 2006). ERGAS with lower value represents the finer pan-sharpened image.

$$\text{RASE} = \frac{100}{\bar{x}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \text{RMSE}\,(band_i)^2} \qquad (5)$$

N represents total number of spectral band (in this study *N* = 4); $\bar{x}$ is the mean value of the *N* spectral bands.

(v) Q refers to universal image quality index. This quality index model any distortion as a combination of three different factors: loss of correlation, luminance distortion, and contrast distortion (Wang and Bovik, 2002, Shahdoosti and Ghassemian, 2015).

$$Q = \frac{\partial xy}{\partial x\, \partial y} \cdot \frac{2\, \mathfrak{m}_x\, \mathfrak{m}_y}{\mathfrak{m}_x{}^2 + \mathfrak{m}_y{}^2} \cdot \frac{2\, \partial_x\, \partial_y}{\partial_x{}^2 + \partial_y{}^2} \qquad (6)$$

While x referred the values for true segments, y represents the values of extracted segments. $\mathfrak{m}_x$ and $\mathfrak{m}_y$ are the mean for x and y values, respectively. The value for Q index ranges from -1 to 1 and 1 represent the ideal situation in which pan-sharpened image is similar to the reference image (x = y) (Marcello *et al.,* 2013). These five metrics were performed using a MATLAB programme developed by Vaiopoulos, (2011).

The reference data source was chosen differently for quantitative and visual assessment. Five pan-sharpened images were compared against original Landsat OLI for quantitative assessment to measure spectral dissimilarity to its original image. However, for visual assessment, the sharpening of these five pan-sharpened images was compared to Ordnance Survey GB (OSGB) MasterMap Topography Layer mapping.

Recent studies recommended using the original multispectral image as reference data by degrading the pan-sharpened image into a similar spatial resolution of the original multispectral image (Du *et al.,* 2007; Javan *et al.,* 2013). Since, the quantitative assessment aimed to assess which pan-sharpening techniques preserve the spectral and radiometric resolution, using original Landsat OLI as reference image enables direct comparison of spectral and radiometric distortion of pan-sharpened Landsat OLI compared to its raw image.

Regarding quantitative assessment, the main condition of the reference image was similar in spectral, radiometric, and spatial resolutions of the original multispectral image (which is Landsat OLI here). Both pan-sharpened Landsat OLI and its original Landsat OLI image were similar in the number of the spectral bands set (four bands) and radiometric resolution (11 bit) however, their spatial resolution differed. As a result, those five pan-sharpened Landsat OLI images were degraded to 30m resolution to match the spatial resolution of Landsat OLI. Table 5.3 depicts the result of the quantitative assessment. The finest accuracies are highlighted for each metric.

Pan-sharpening algorithms	RASE	Q	ERGAS	CC	DIV
HCS	6.32	0.74	1.63	0.84	8.65
IHC	10.28	0.73	5.8	0.77	7.97
Brovey	8.18	0.79	4.61	0.82	7.21
Gram-Schmidt	8.18	0.79	4.61	0.82	7.21
Wavelet Resolution Merge	9.9	0.75	1.81	0.72	2.66
Ideal value	lower	1	lower	1	0

Table 5.2. The result of five metrics, highlighted values represent the most accurate.

In addition to a quantitative approach, the visual assessment was also performed to assess the sharpness of the pan-sharpened image. OSGB MasterMap Topography Layer mapping acted as visual reference data set. It offers the most detailed digital map of the UK with 1m positional accuracy in urban areas (1:1250 scale) at the 99% confidence level scale (Ordnance Survey, 2009). Because the smallest ground feature that can be identified on Pan-sharpened Landsat OLI image is 15m, features with area less than $225\mathrm{m}^2$ (15m × 15m) were excluded from the MasterMap data. Then, 100 random features in which represent polygons in OS MasterMap Topography (i.e., buildings, roads, water bodies, and other impervious structures) were considered as a visual reference samples. Comparing the objects in MasterMap to the pan-sharpened image, each object is ranked as 1, 2 or 3. For example, value one was allocated to the object samples which are in MasterMap, but can not be identified in the image. The value of 3 is allocated to those objects which are clearly identifiable in the image according to MasterMap. Those objects which can be seen on the image but their edges are not sharp as three are allocated the number 2.

Regarding the quantitative measurements, the overall quality of each method was normalised metric to 0 - 1 range. This normalisation enables direct comparison among each these five quantitative measurements. For each of the individual metric, the minimum and maximum values were identified. Then through this formula $(\frac{value-maximum}{maximum-minimum})$ each metric was normalised (Becker and Salber, 2010, Podvezko and Podviezko, 2010). For metrics zero represents the ideal, value, whereas the higher number in visual assessment depicts the sharpest algorithm. Table 5.4 summarises the results was obtained for both the quantitative and qualitative assessments.

Pan-sharpening algorithms	RASE	Q	ERGAS	CC	DIV	Total	Visual assessment
HCS	0.00	0.88	0.00	0.00	1.00	1.88	2.88
IHC	1.00	1.00	1.00	0.59	0.89	4.48	2.66
Brovey	0.47	0.06	0.71	0.20	0.76	2.20	2.16
Gram-Schmidt	0.47	0.00	0.71	0.17	0.76	2.11	2.10
Wavelet	0.90	0.61	0.04	1.00	0.00	2.55	1.86

Table 5.3. Normalised quantitative and metric pan-sharpening quality assessment.

According to the result, HCS pan-sharpening method obtained the most accurate result which is 1.88 and 2.88 for the qualitative and quantitative assessment respectively. This means while HCS method enhanced the Landsat OLI best, it also maintains the most of its original spectral values. Therefore, the HCS pan-sharpening technique was performed to improve Landsat OLI spatial resolution from 30 to 15m. Figure 5.3 provides a visual comparison of

each five pan-sharpening methods. It is evident that HCS pan-sharpening creates the sharpest imagery.

The reason for the success of HCS over other four pan-sharpening can be referred to as the meddling concept. Padwick *et al.* (2010) pointed out that the HCS algorithm models the data statistically instead of employing a physical model involving the exact spectral band passes of the sensor. Since HCS models the data through a statistical method, it is particularly useful when there is no strong spectral overlap between multi-spectral bands (blue, green, red, and NIR) and panchromatic band. Here, the spectral range of panchromatic covers only green and red band and blue and there is no overlapping between panchromatic and blue, NIR bands.

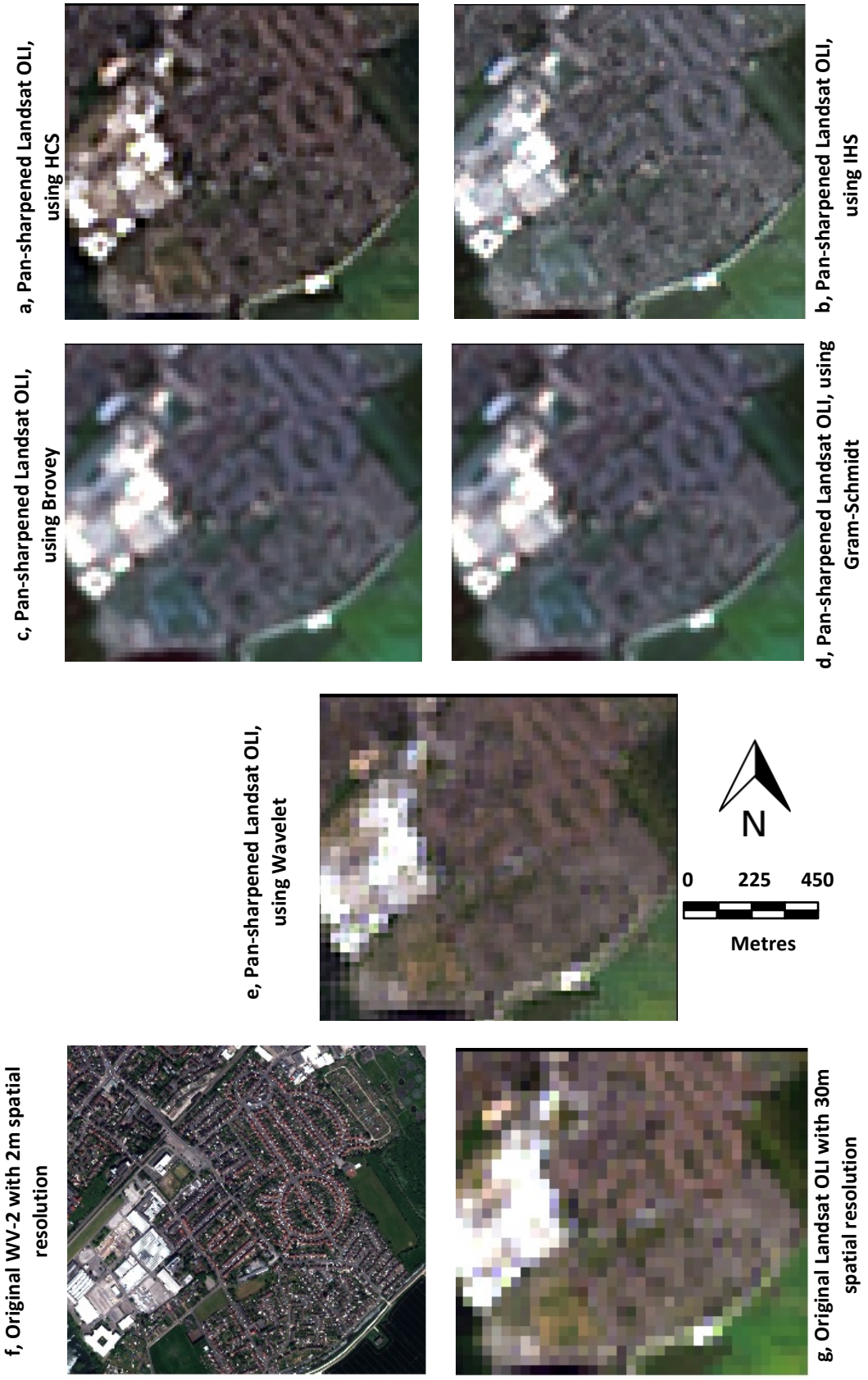


Figure 5.3. Visual comparison of pan-sharpening of a. HCS, b. IHS, c. Brovery, d. Gram-Schmidt, e. Wavelet merge resolution, f.WV-2 and g. Landsat OLI using Erdas Imagine software.

5.5 Sparse representation to enhance Landsat OLI imagery

5.5.1 Introduction

As stated in Section 4.8, two strategies were adopted to increase the spatial resolution of Landsat OLI imagery. The first strategy was performing sparse representation to enhance the spatial resolution of the Landsat OLI with 30m and four spectral bands (blue, green, red, and NIR) to 2m resolution. The second strategy was performing sparse representation to enhance the spatial resolution of the pan-sharpened Landsat OLI with 15m pixel and four spectral bands (blue, green, red, and NIR) to 2m resolution.

The sparse representation method operates in two phases including training and reconstruction (as detailed in Section 4.8). In short, within the training phase, both fine and coarse resolution dictionaries are generated based on training image (WV-2). Moving to reconstruction phase, first, the sparse vector for each patch is computed through the L-resolution dictionary and input coarse resolution image (Landsat OLI imagery) and then the spatially enhanced Landsat OLI was estimated by adding recovered lost texture (which determine through H-resolution dictionary and sparse vector) into the up-sampled input coarse resolution image.

5.5.2 Pre-processing

Before applying the sparse representation method to the Landsat scene, two pre-processing steps were considered including radiometric conversion and spectral subsetting.

Sparse representation computation is based on the digital number (pixel values). Sparse representation enhances each band individually according to the spectral reflectance. In fact, instead of raw spectral reflectance, this method focuses training process on characterising the relation between low-resolution patches, the edges and texture content within the corresponding high-resolution patches (Elad, 2010, P.347). Please refer to section 4.8.4 of Chapter 4 for step by step explanation on sparse representation performance including the images' spectral process.

Landsat OLI records the radiation reflectance in 16 bit (ranging from 0 to 65535), however radiometric resolution of WV-2 is 11 bit (ranging from 0 to 2047). Since the spatial resolution of Landsat OLI enhances by computing coarse and fine resolution dictionaries which were trained by the WV-2 image, it is an essential that both images have the same radiometric scale. Because WV-2 is a training resource to increase Landsat OLI's spatial resolution, its radiometric resolution was converted to Landsat OLI's radiometric resolution. A simple formula was used to perform the radiometric conversion. Firstly, the minimum value of each WV-2 band was subtracted from entire pixels for that specific band; then each digital pixel number was multiplied into 32 ($\frac{2^{16}}{2^{11}}$).

The MATLAB code associated with the sparse representation method was developed by Yang *et al.* (2010). It was initially programmed for imagery with three spectral bands (RGB) with "BMP" format. The MATLAB code was modified to operate according to the individual band with the "Tif" file format. This modification ensures that sparse representation is fit for remotely sensed applications, where the spectral bands normally exceeds three spectral bands. Each of the three multi-spectral datasets (Landsat OLI, pan-sharpened Landsat OLI, and WV-2) were saved as four images with the single band as "tif" format. This means the entire sparse representation process will be performed on single band images.

It should be noted that there is no unique protocol for performing sparse representation (including parameters in the training and construction phases) and using sparse representation is mainly limited to portraying imagery rather than complex remotely sensed imagery from large urban areas (Yang *et al.,* 2010). Therefore, a trial-and-error approach was adopted to find the optimum values within both training and construction phases.

5.5.3 Training phase

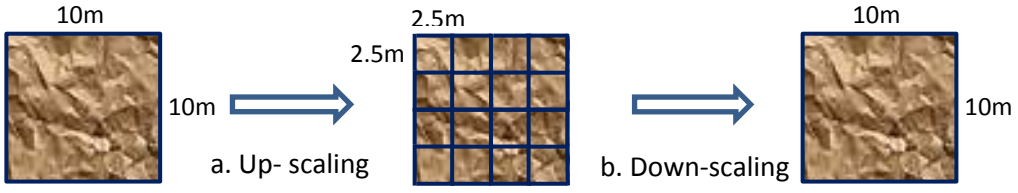
Training is the first phase of implementing sparse representation algorithm. In short, it involves acquiring fine resolution imagery to train two H-resolution and L-resolution dictionaries (i.e., D_h and D_l, respectively).

The first part of training phase is to feed the algorithm with training data sets. WV-2 imagery with the 2m resolution was used as a single training image. The

next task was to allocate the most optimum values for four parameters including dictionary size, a number of samples, patch size, and zoom factor.

A number of samples determine a total number of randomly selected patches. However, the dictionary was designed in the way to select the patches which hold distinctive information and it excludes the patches with similar (repetitive) information.

Figure 5.4 distinguishes the role of zoom factor within the training and reconstruction phases. Within the training phase, zoom factor acts as a decimation factor by discarding a relevant number of rows and columns to form a new cell of a larger size. For example, input training imagery with the 2.5m spatial resolution was degraded to imagery with 10m spatial resolution (Figure 5.4.b). However, the zoom factor acts as a magnification factor in the reconstruction phase. As Figure 5.4.a shows a pixel with the 10m spatial resolution was replaced by 16 pixels with 2.5m spatial resolution (Figure 5.4.a).



Zoom factor: $4 = \frac{10}{2.5}$

Figure 5.4. a) zoom factor role for reconstruction phase by increasing the spatial resolution from 2.5 to 10m, b) zoom factor role for training phase by degrading spatial resolution from 10 to 2.5m.

Since the sparsity vector (which then was used to estimate the spatially enhanced Landsat OLI imagery) is computed according to the trained L-resolution dictionary (D_l) (from degraded WV-2 imagery) and input coarse resolution image (Landsat OLI), it is important that training image (WV-2 imagery) is degraded to match the similar spatial resolution of the input coarse resolution image (Landsat OLI) to ensure the optimum result. For example, in the case of the first strategy (increasing the spatial resolution of Landsat OLI directly from 30 to 2m), the zoom factor selected as 15 means that WV-2 imagery with 2m spatial resolution is degraded to the resolution of Landsat OLI (30m). However, the zoom factor 7.5 ($\frac{15m}{2m}$) was assigned for the second strategy using the pan-sharpened Landsat OLI as input coarse resolution image. Next, patch size defines the size of grouping cells forming a sample. There is a direct relationship between zoom factor and patch size. According to Peleg and Elad (2014), the minimum patch size should be no less than the zoom factor. Other studies emphasised that selecting larger patch size increases the overlap between the estimated fine resolution patches leading to the most accurate image (Elad, 2010)**.** Therefore, the patch size 30 x 30 was chosen for the first strategy (when zoom factor was selected as 15), whereas the patch size 15 x 15 was chosen for the second strategy (when zoom factor was selected as 7.5).

Dictionary size number varies from small to large size (like 64, 128, 256, 512, 1024, 2048). As it explained in Section 4.7, a dictionary represents the main structure of the entire imagery, and its size is the key to successful enhancement. Cheng (2015) recommended adopting a larger dictionary size (like 1024) for the more optimum result. On the other hand, dealing with a

larger dictionary requires more computational power and time. Since there was only one image providing the information for a dictionary, it is important to select an appropriate dictionary size to reassure the dictionary is true representative of a single training image (WV-2 with $121\mathrm{km}^2$). The training samles were chosen in the way that to be the true representative of the entire the urban area. As a result, the training areas acquired from the city centre were mostly covered by dense imprevious surfaces and the suburbs were covered by mostly vegetated features. According to the result of the trial-and-error process, 2048 was allocated as dictionary size for both strategies.

5.5.4 Reconstruction phase

As the second part of the sparse representation, the reconstruction phase aims to estimate the lost texture and edges of both Landsat OLI with 30m spatial resolution and pan-sharpened Landsat OLI with 15m spatial resolution.

Zoom factors in the reconstruction phase were chosen the same way as zoom factors in the training phase. This can ensure that an input coarse resolution (i.e., Landsat OLI with 30m spatial resolution) was enhanced to the target spatial resolution (i.e., 2m spatial resolution). Similar to training, zoom factor 15 was assigned to improve the spatial resolution of Landsat OLI from 30 to 2m. Next, zoom factor 7.5 was allocated to enhance pan-sharpened Landsat OLI from 15 to 2m. Table 5.5 depicts the values that were allocated for parameters in both training and reconstruction phases.

Phase	Image	Case1: Upscaling Landsat OLI from 30 to 2m	Case2: Upscaling pan-sharpened Landsat OLI from 15 to 2m
Training	Spectral bands	Blue, Green, Red, and NIR	Blue, Green, Red, and NIR
	Dictionary size	2048	2048
	Number of samples	100,000	100,000
	Patch size	30	15
	Downscale factor	15	7.5
Reconstruction	Image	Original Landsat OLI with 30m resolution	Pan-sharpened Landsat OLI with 15m resolution
	Spectral bands	Blue, Green, Red, NIR	Blue, Green, Red, NIR
	Up-scale factor	15	7.5

Table 5.4. The parameters were allocated for both training and reconstruction phase.

5.5.5 Post-processing

After completing the sparse representation, sixteen enhanced spectral bands sets were produced as individual spectral bands (such as blue, green, red, and NIR). Several post-processing steps were performed to prepare the multispectral images for classification.

Firstly, these single band images did not have any spatial referencing and projection system. Hale and Rock (2003) suggested performing geometric and topographic corrections using topographic maps. As the first pre-processing task, these sixteen single band images were georeferenced according to the OS MasterMap.

Before merging spectral bands to form a multi-spectral image, red and NIR spectral bands of each case were used to generate a normalised difference vegetation index (NDVI = $\frac{nir - red}{nir + red}$). This is because the inclusion of the NDVI can

potentially increase the accuracy of the classifications result (Hamedianfar and Shafri, 2015).

Finally, within both strategies, those five spectral bands were merged to generate a multispectral imagery with five spectral band sets (blue, green, red, NIR, and NDVI). Therefore, referring to two different strategies, two set of the multispectral images with 2m resolution were generated.

5.6 Classification

5.6.1 Classification scheme

Despite the classification scheme used in Chapter 3 with eleven land-cover classes, in this Chapter, a classification scheme with eight land-cover classes was used. Firstly, the shadow was eliminated from the classification scheme because no shaded feature was identified on both enhanced Lansat OLI with 2m resolution. Secondly, the TD measured at the finest level revealed significant spectral similarity among clay versus concrete roof tiles (1 – 1.2) and broad-leaved versus needle-leaved trees (0.7 – 0.8). Richards and Jia (1999) recommended the spectral separation value less than 1.3 as a poor spectral separability. Therefore, needle-leaved and broad-leaved trees merged forming a tree class. Also, roof tile class defined by merging clay and concrete roofs classes. As a result, eight classes were selected water, soil, roof tiles, slate roofs, asphalt, metal roofs, trees, and grass. Table 5.6, explains the land cover classification scheme employed in this chapter.

No	Class	Description
1	Asphalt	Urban ground surfaces covered in asphalt such as roads and car parks
2	Roof tiles	Predominantly residential buildings covered in red and dark grey concrete tiles
3	Slate roofs	Predominantly residential buildings covered in light grey slate tiles
4	Metal roofs	Predominantly industrial buildings covered in white metal panels
5	Grass	Areas of grassland such as urban parks and lawns, plus surrounding rural agriculture
6	Tress	Patches of broad-leaved and needle-leaved trees
7	Bare soil	Open areas covered by bare soil
8	Water	Water bodies including lakes, rivers, ponds and canals

Table 5.5. The land cover classification scheme used for Nottingham.

5.6.2 Classification training

In addition to the two spatially enhanced Landsat OLI images with 2m spatial resolution, three others data sets were used to map the land cover of the urban area of Nottingham including the original Landsat OLI image, the pan-sharpened Landsat OLI image, and WV-2 image with 30, 15, and 2m spatial resolution, respectively. The purpose of including these three data sets was to demonstrate the potential of two spatially enhanced images in the land cover classification of urban areas. Landsat OLI and pan-sharpened Landsat OLI images can be used as benchmarks to demonstrate the degree of enhancing overall accuracy of sparse representation method. However, WV-2 imagery can explain the effectiveness of sparse representation method for mapping enhancement.

Mather and Koch (2011) recommend a minimum of 10 ρ to 30 ρ training samples per land cover class. ρ stands for a number of spectral bands which is five here. Since a classification scheme with eight land cover classes was adapted here, it required a minimum of 50 to 150 per class or 400 to 1200 per the entire classification. In total, 2895 pixels were selected randomly as 325

blocks of nine (3x3) pixels for WV-2 imagery and both spatially enhanced Landsat OLI images. To ensure direct comparability between different results, training samples for the five data sets were selected from similar locations. However, some slight modifications were necessary for the pan-sharpened Landsat OLI image, and original Landsat OLI of the coarser spatial resolutions of 15 and 30m respectively. Similar to Chapter three, two different training pattern strategies were employed. As a training pattern, a block of 3 x 3 pixels was adopted for WV-2 and two spatially enhanced Landsat OLI with 2m spatial resolution, while a conventional individual pixel was used for both original and pan-sharpened Landsat OLI with 30 and 15m spatial resolution. The detailed explanation of the benefits of using a block of 3×3 pixels is provided in Section 3.4.1.

Adopting a block of 3 x 3 pixels for pan-sharpened OLI with 15m resolution means each training sample covers an area with $2025\mathrm{m}^2$ (45m x 45m) or even near 1 hectare for Landsat OLI with 30 resolution. The basic rule of the hard classification training process is to select pure or as pure as possible pixels for training samples which are unlikely to be found in this size in an urban environment. This condition causes the number of training samples for pan-sharpened Landsat OLI, and Landsat OLI was less than other finer images. In fact, when the spatial resolution becomes coarser, each pixel covers a larger area, and it became more difficult to identify pure pixels leading more pixels to mix generally. To avoid mixed pixels in training, only similar locations of WV-2 which are represented as a pure pixel in coarser resolution data sets were selected as a training sample. As a result, these two modifications decreased

successively the total number of training samples at each coarser spatial resolution (275 and 255 training samples for pan-sharpened Landsat OLI and Landsat OLI respectively).

The spectral dependency of training samples was examined through the TD technique (Richards and Jia, 1999). This assessment was implemented after amending training data set several times to ensure training samples spectrally were separated and minimise mixed pixel of training samples.

5.6.3 Classification methods

5.6.3.1 Pixel-based classification

In Chapter 3, the potential of three classifiers (ML, SVM, and OBIA/NN) were assessed, and it was found that the non-parametric pixel-based and object-based classifiers (SVM and OBIA/NN, respectively) classifiers achieved the highest accuracies in the land cover classifications (Please refer to Section of 3.5.3). As such, the five data sets were classified using SVM and OBIA/SVM classifiers. Section 3.4.2 explained the details on the concepts and parameters setting recommendations for the classifiers.

The accuracy of the SVM classifier is directly related to the types of the kernel and its corresponding parameters (Li *et al.,* 2014). Similarly to Chapter 3, radial basis function was used as a kernel types because, the radial basis function kernel is a non-linear and ideal for the complex environment (Marjanovic *et al.,* 2011; Jebur *et al.,* 2014, Hamedianfar *et al.,* 2014). Gamma and penalty were two parameters of radial basis function kernel which were chosen through trial-and-error as 500 and 0.5, respectively.

5.6.3.2 OBIA/SVM

The OBIA strategy adopted in this chapter is slightly different to that of Chapter 3. This is due to the reduction in the number of land cover classes being mapped (from eleven classes to eight land cover classes) to enable direct comparison between the object-based approach to the equivalent per-pixel method. As a result, there was no need to distinguish between clay versus concrete roof tile and needle-leaved versus broad-leaved trees. Within the first stage of the OBIA/SVM approach, each image was segmented and then classified into vegetated and non-vegetated areas. Then, as a second step, vegetated areas of the image were segmented and classified into trees and grass (Stage 2a), while non-vegetated zones were divided into their constituent classes including roof tiles, water, bare soil, metal roofs, slate roofs, and asphalt (Stage 2b). Figure 5.5 demonstrates the diagram of the OBIA/SVM procedure. Segmentation was conducted through the eCognition software (Trimble, 2016). Performing segmentation in eCognition required three main parameters of interest including scale, shape, and compactness. Detailed information on the explanation of these parameters is provided in the previous chapter. The optimum values for shape, and compactness was identified 0.3 and 0.8 respectively through the trial-and-error.

To mitigate the subjectivity (user interfering) level within trial-and-error segmentation, the minimum and maximum possible scale value for individual image data sets was identified. The minimum value represents the lowest possible value for scale leading to complete over-segmentation, while the maximum scale value depicts the situation which all objects were under-

segmented. Then, an arbitrary value of 15 was used as an interval between the minimum and maximum values. For example, for both pan-sharpening and original Landsat OLI, minimum and maximum scale values were defined as 5 and 140 respectively, which by considering 15 as an interval value, a range of scale was examined including 5, 20, 35, 50, 65, 80, 95, 110, and 140.

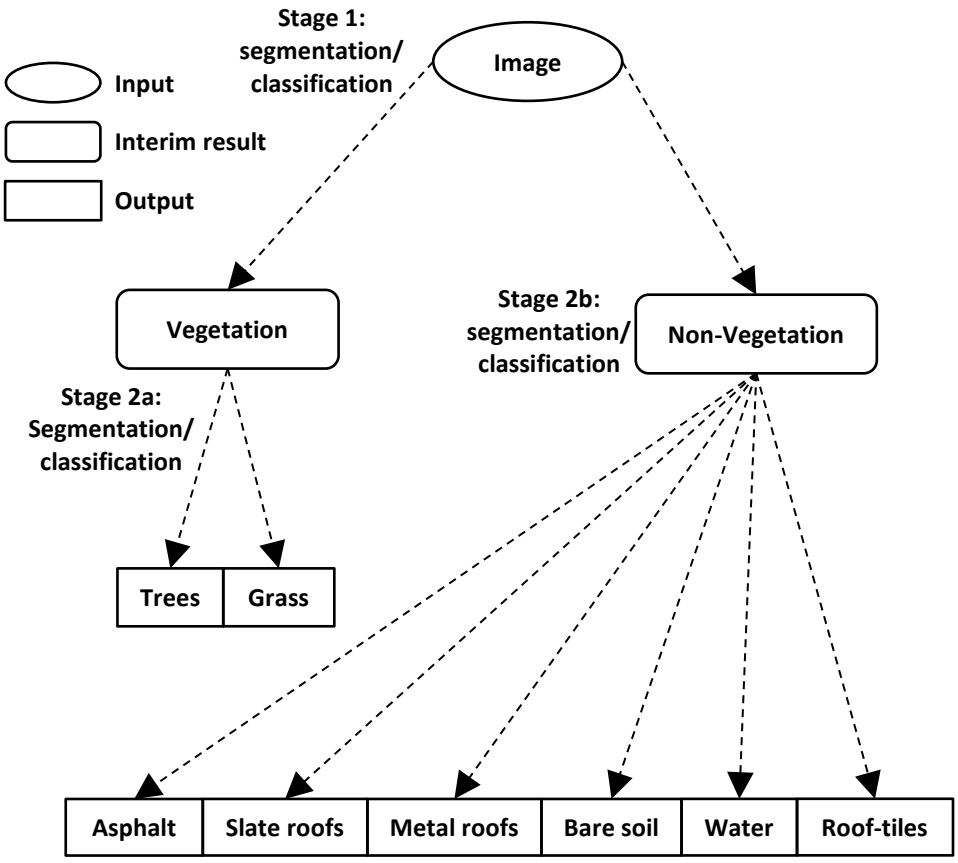


Figure 5.5. OBIA/SVM two-stage segmentation and classification diagram.

A mixture of quantitative and qualitative assessment was performed to determine the most optimum value for scale because scale value is key in this multi-resolution segmentation. Firstly, the quality of all segmentation was tested according to five metrics including area fit index (AFI), relative area sub, relative area super, quality rate (QR), and D. A summarised explanation of these

five metrics is provided in Chapter three, while for further details, please refer to (Clinton *et al.,* 2010). These metrics were tested against the reference objects extracted from OS MasterMap. Besides quantitative indices, the visual assessment was performed to find the optimum value for scale. This quality assessment process of extracted segments is the same as the process used in Section 3.4.3.2. While here a brief summary of this assessment process is provided, please refer to Section 3.4.3.2 for the more detailed explanation. Table 5.7 describes the details of segmentation parameters allocated during two stages for both classification and segmentation.

Data sets		Stage 1. Separation of Vegetation and Non-Vegetation		Stage 2a. Identification of Individual Vegetation Classes		Stage 2b. Identification of Individual Non-Vegetation Classes	
Name	Spatial Resolution	Segmentation Parameters (Scale, Shape, Compactness)	Spectral Merging Threshold	Segmentation Parameters (Scale, Shape, Compactness)	Spectral Merging Threshold	Segmentation Parameters (Scale, Shape, Compactness)	Spectral Merging Threshold
Landsat OLI	30m	6, 0.3, 0.8	NA	7, 0.3, 0.8	NA	4, 0.3, 0.8	NA
Pan-sharpened Landsat OLI	15m	15, 0.3, 0.8	NA	18, 0.3, 0.8	NA	10, 0.3, 0.8	NA
Spatially enhanced Landsat OLI using sparse representation	2m	30, 0.3, 0.8	NA	40, 0.3, 0.8	NA	20, 0.3, 0.8	NA
Spatially enhanced pan-sharpening Landsat OLI using sparse representation	2m	30, 0.3, 0.8	NA	40, 0.3, 0.8	NA	20, 0.3, 0.8	NA
WV-2	2m	25, 0.3, 0.8	20	35, 0.3, 0.8	20	17, 0.3, 0.8	10

Table 5.6. The details on values were allocated for scale, shape, and compactness.

After segmentation extraction, the classification procedure was implemented. While in Chapter Three the nearest neighbour method was used as a classifier, here the SVM classifier was employed. Using SVM as a classifier can ensure direct comparison between the pixel and object-based classification by eliminating the influence of considering different classification algorithm. Because performing SVM on objects does not offer a straightforward tool in OBIA software, a combination of eCognition, ArcGIS, and Envi was employed to classify the most accurately extracted segments.

5.7 Accuracy assessment and statistical test

Following classification completion, an accuracy assessment was implemented to measure the degree of correctness in the class allocations of those ten thematic land cover maps (i.e., the original Landsat OLI, pan-sharpening Landsat OLI, spatially increased Landsat OLI using sparse representation, spatially increased pan-sharpened Landsat OLI using sparse representation, and original WV-2 using SVM and OBIA/SVM classifiers). Since those ten land covers maps are represented as pixel and object spatial units, the similar point-based strategy was employed for evaluation of both pixel and object based thematic maps. Adopting similar testing samples as a point-based approach offers a direct comparison between the pixel and object-based land cover maps.

In total, 334 testing samples were selected randomly for the entire image area. Testing samples range from 30 to 49 per class, and 42 was the average of testing sample allocated per class. It is expected that the testing samples evenly

distribute among the entire land cover map. A unique testing data set was considered for the entire 10 land cover maps ensuring direct comparability between results. A combination of field work, online very fine resolution image, and OS MasterMap were used to allocate the class of the testing samples.

The result of accuracy assessment was represented through confusion matrices (error matrices) by comparing correspondence between classified land cover maps and their relevant testing samples. Besides overall accuracy, each confusion matrix depicts the class-level accuracies (including users and producers accuracy) and inter-class error (Foody, 2002). Janssen and van der Wel (1994) state that accuracy statements should be compared in a statistically rigorous manner and the results expressed with confidence limits. Congalton, Janssen and Vanderwel (1994) have identified the confusion matrix as the core of accuracy assessment. However, they have stated that there is much scope to extend the analysis beyond the confusion matrix. As the second phase of classification evaluation, the McNemar statistical test was computed to determine whether the differences between the accuracy of classification and different spatially enhanced approaches are statistically significant or not. Details on the statistical test concept are provided in Chapter three.

5.8 Results

In total, ten land cover maps were generated, using a combination of two classifiers (SVM, and OBIA/SVM) and five image data sets (Landsat OLI, pan-sharpened Landsat OLI, spatially increased Landsat OLI using sparse

representation, spatially increased pan-sharpened Landsat OLI using sparse representation, and WV-2).

The main aim of this chapter was to demonstrate the capability of the sparse representation method to increase the spatial resolution of Landsat data for mapping a complex urban environment. The overall accuracies of these two spatially enhanced Landsat OLI images techniques were compared against both original Landsat OLI and WV-2, which acted as the coarsest and finest benchmarks respectively. Another objective of this chapter is to compare the accuracy of SVM, and OBIA/SVM classifiers on both spatially enhanced images. Figure 5.6 demonstrates 15 small subsets of five different images including Landsat OLI 30m, pan-sharpened Landsat OLI 15m, spatially enhanced Landsat OLI 2m, spatially enhanced pan-sharpened Landsat OLI 2m, and WV-2 2m. These 15 images provide the visual comparison of Landsat dataset and its three up-scale versions through the three scales (1:20,000, 1:10,000, and 1:5000). It is evident that spatially enhanced pan-sharpened Landsat data offer a greater level of spatial information compared to pan-sharpened Landsat and spatially enhanced Landsat data.

Figure 5.7 represents a small subset (700m × 700m) of full extent (21km × 21km) 10 land cover maps. It offers a visual comparison of the accuracy of both SVM and OBIA/SVM classifiers over five dataset images. It should be noted that because these figures cover a small subset of the full extent of the study area, they should be interpreted with some caution.

Though WV-2 imagery offers more detailed maps compared to both spatially increased Landsat data using sparse representation and spatially increased

pan-sharpened Landsat data, using sparse representation, produced maps which offer more details and sharper edges compared to pan-sharpened and original Landsat.

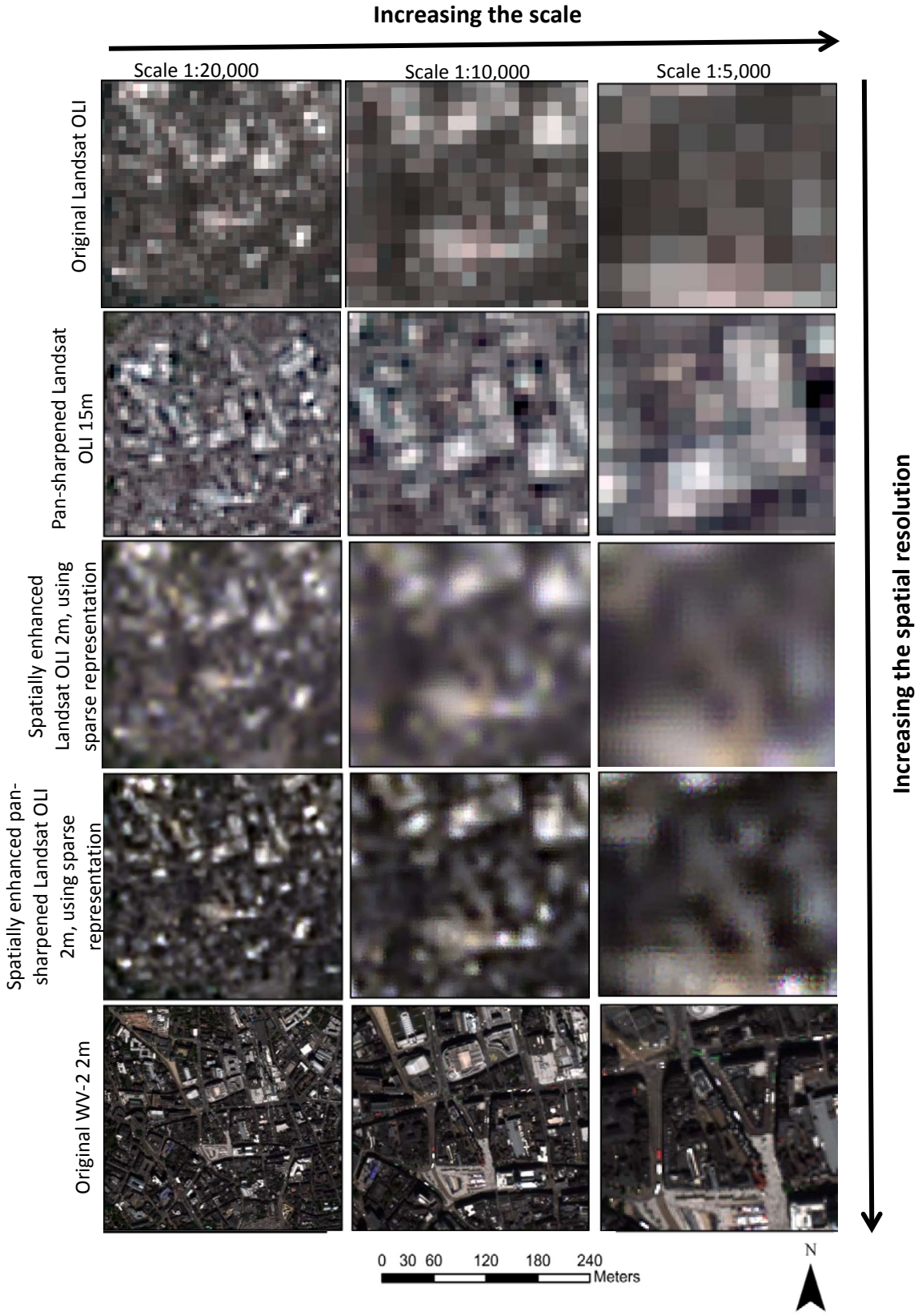


Figure 5.6. 15 small subsets of five different images (RGB) of Nottingham: i) Landsat OLI 30m, ii) pan-sharpened Landsat OLI 15m, iii) spatially enhanced Landsat OLI 2m using sparse representation, iv) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation, and v) WV-2 2m) according to three scales (1:20,000, 1:10,000, and 1:5000).

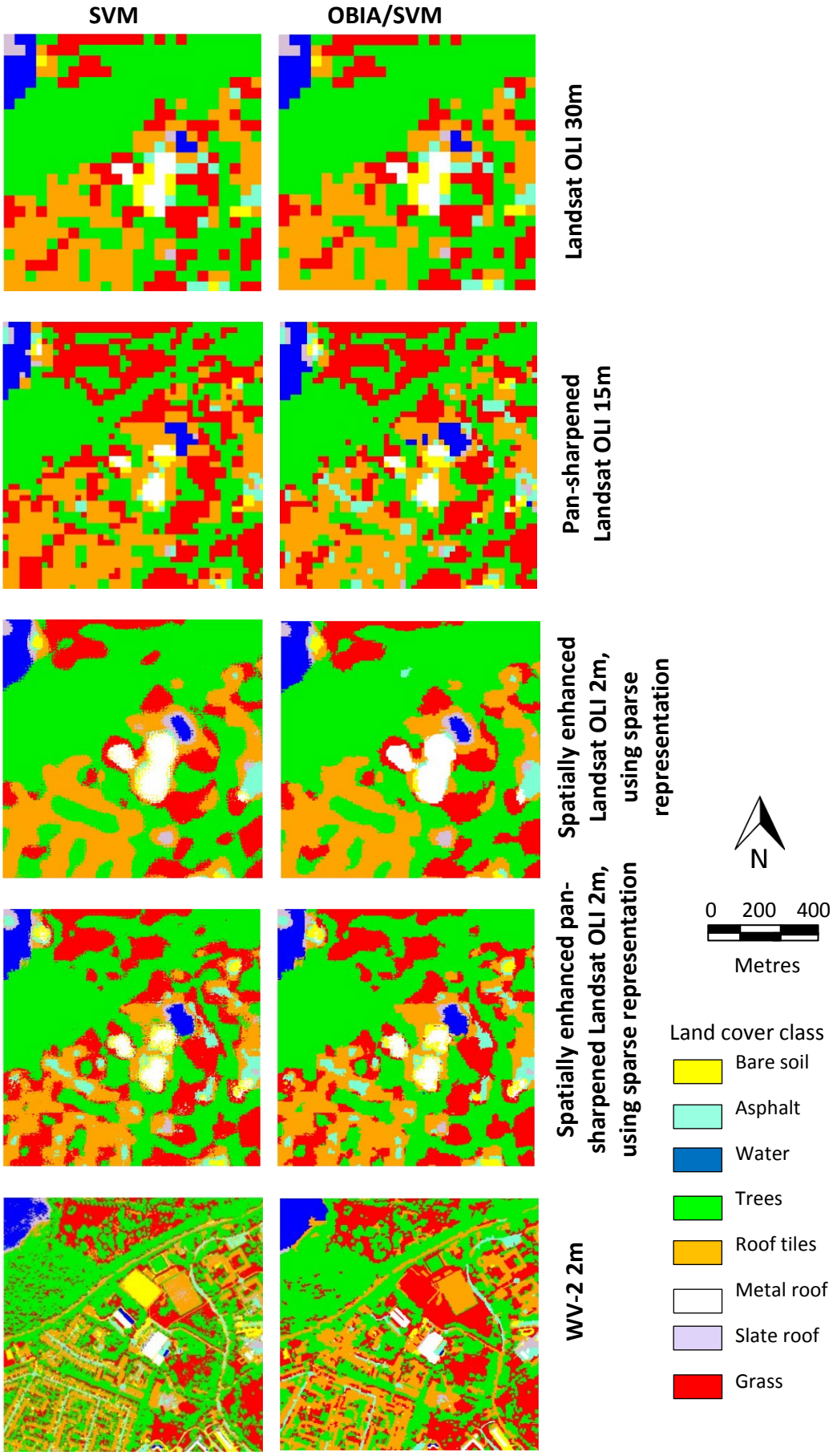


Figure 5.7. 10 small subsets (750 × 750m) of land cover maps which were produced by performing SVM and OBIA/SVM over five data sets: i) Landsat OLI 30m, ii) pan-sharpened Landsat OLI 15m, iii) spatially enhanced Landsat OLI 2m using sparse **representation, iv) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation, and v) WV-2 2m.**

Figure 5.7 offers the visual comparison among 10 classified maps by performing both SVM and OBIA/SVM classifiers over five data sets: i) Landsat OLI 30m, ii) pan-sharpened Landsat OLI 15m, iii) spatially enhanced Landsat OLI 2m using sparse representation, iv) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation, and v) WV-2 2m. It is clear in both Figures 5.6 and 5.7 that sparse representation method significantly increases the spatial resolution of Landsat OLI imagery and consequently increases the thematic details (from a visual perspective) of classified Landsat OLI imagery. According to both Figures 5.6 and 5.7, performing sparse representation over the pan-sharpened Landsat OLI achieved more accurate result compared to performing sparse representation over the Landsat OLI.

Figure 5.8 depicts a summary of the overall classification accuracy. This can enable direct comparison of the accuracy of the all ten land covers maps. As it was expected, the WV-2 with the finest spatial resolution achieved the most accurate classifications, and Landsat OLI with the coarsest spatial resolution obtained the least accurate classification. While this result demonstrates the capability of WV-2 imagery for accurately mapping detail urban areas, at the same time it also demonstrates the inability of Landsat OLI with 30m to successfully map a complex urban area. However, performing sparse representation on pan-sharpened Landsat OLI increased the overall accuracy of Landsat OLI by 22% using the SVM classifier.

OBIA/SVM achieved the most accurate classification with 92% followed by SVM with 89% for WV-2 with 2m resolution. Here, the accuracy of SVM and OBIA/SVM on Landsat OLI obtained overall accuracy by 54% and 54%, respectively.

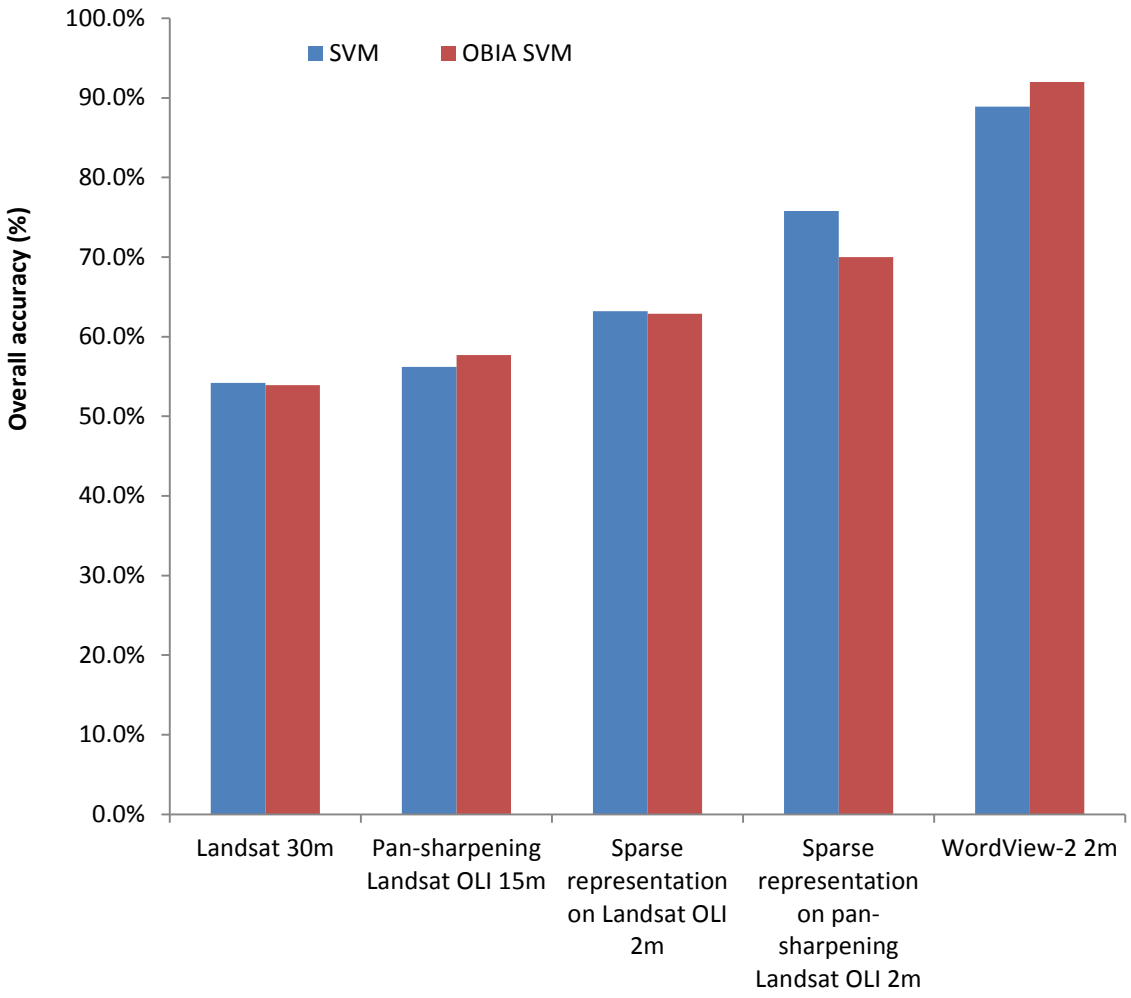


Figure 5.8. The overall classification accuracies for the ten land cover maps.

While the full set of statistical classification results (based on overall accuracy) is summarised in figure 5.8, the six tables below depict full error details with their overall accuracy (remaining error matrix provided at Appendix Section Page 284).

OBIA\SVM on spatially increased Landsat OLI 2m using sparse representation

Overall Accuracy: 63%	Reference data								
Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**14**	0	0	0	0	5	1	1	**67**
Asphalt	4	**11**	0	0	8	0	4	1	**39**
Water	0	2	**35**	0	0	0	0	0	**95**
Trees	0	2	1	**32**	3	0	2	4	**73**
Roof tiles	5	23	5	7	**34**	0	11	7	**37**
Metal roofs	1	1	0	0	0	**35**	0	0	**95**
Slate roofs	0	5	1	0	1	0	**22**	0	**76**
Grass	3	1	0	9	3	0	0	**27**	**59**
Producer accuracy %	**47**	**24**	**83**	**67**	**69**	**88**	**55**	**68**	

Classification result

Table 5.7. The error matrix for the OBIA/SVM classification on spatially increased Landsat OLI with 2m spatial resolution using sparse representation.

SVM on spatially increased Landsat OLI 2m using sparse representation

Overall Accuracy: 63%	Reference data								
Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roofs	Slate roofs	Grass	User accuracy %
Bare soil	**16**	2	0	0	0	2	0	1	**76**
Asphalt	1	**10**	0	0	7	0	5	1	**42**
Water	0	2	**34**	0	0	0	0	0	**94**
Trees	0	1	0	**31**	2	0	0	6	**78**
Roof tiles	7	24	6	7	**35**	0	13	7	**35**
Metal roofs	1	0	0	0	0	**38**	0	0	**97**
Slate roofs	0	5	2	0	1	0	**22**	0	**73**
Grass	5	1	0	10	4	0	0	**25**	**56**
Producer accuracy %	**53**	**22**	**81**	**65**	**71**	**95**	**55**	**63**	

Classification result

Table 5.8. The error matrix for the SVM classification on spatially increased Landsat OLI with 2m spatial resolution using sparse representation.

OBIA\SVM on spatially increased pan-sharpening Landsat OLI using sparse representation

Overall Accuracy: 70%	Reference data								
Classification result / Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**16**	1	0	0	0	5	0	1	**70**
Asphalt	6	**19**	1	1	8	1	3	1	**48**
Water	0	0	**38**	0	0	0	0	0	**100**
Trees	0	2	0	**32**	2	0	1	1	**84**
Roof tiles	3	19	3	7	**35**	0	13	5	**41**
Metal roofs	1	0	0	0	0	**34**	0	0	**97**
Slate roofs	0	2	0	0	1	0	**23**	0	**88**
Grass	4	0	0	8	3	0	0	**32**	**68**
Producer accuracy %	**53**	**43**	**90**	**67**	**71**	**85**	**58**	**80**	

Table 5.9. The error matrix for the OBIA/SVM classification on spatially increased pan-sharpened Landsat OLI with 2m spatial resolution using sparse representation.

SVM on spatially increased pan-sharpening Landsat OLI using sparse representation

Overall Accuracy: 76%	Reference data								
Classification result / Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**21**	1	0	0	1	3	0	0	**81**
Asphalt	2	**27**	1	1	5	1	3	1	**66**
Water	0	2	**38**	0	0	0	0	0	**95**
Trees	0	2	0	**33**	0	0	1	1	**89**
Roof tiles	3	12	3	5	**38**	0	10	4	**51**
Metal roofs	0	0	0	0	0	**36**	0	0	**100**
Slate roofs	0	1	0	0	0	0	**26**	0	**96**
Grass	4	0	0	9	5	0	0	**34**	**65**
Producer accuracy %	**70**	**60**	**90**	**69**	**78**	**90**	**65**	**85**	

Table 5.10. The error matrix for the SVM classification on spatially increased pan-sharpened Landsat OLI with 2m spatial resolution using sparse representation.

SVM on WV-2

Overall Accuracy: 89% / Class	Reference data: Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**29**	0	0	0	1	2	0	0	**91**
Asphalt	1	**39**	5	0	1	0	4	0	**78**
Water	0	1	**33**	0	0	0	0	0	**97**
Trees	0	0	0	**45**	0	0	0	4	**92**
Roof tiles	0	3	1	2	**45**	0	5	0	**80**
Metal roofs	0	0	0	0	0	**38**	0	0	**100**
Slate roofs	0	0	4	0	2	0	**31**	0	**84**
Grass	0	0	0	1	0	0	0	**37**	**97**
Producer accuracy %	**97**	**91**	**77**	**94**	**92**	**95**	**78**	**90**	

Classification result

Table 5.11. The error matrix for the SVM classification on WV-2 imagery with 2m spatial resolution.

SVM on Landsat OLI

Overall Accuracy: 89% / Class	Reference data: Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**16**	2	2	0	0	9	1	4	**47**
Asphalt	2	**10**	0	1	3	4	3	0	**43**
Water	0	0	**31**	0	0	0	0	0	**100**
Trees	1	1	4	**22**	2	0	0	9	**56**
Roof tiles	2	25	2	12	**38**	0	16	6	**38**
Metal roofs	2	0	0	0	0	**25**	0	0	**93**
Slate roofs	1	5	2	0	2	0	**18**	1	**62**
Grass	5	1	0	13	5	2	2	**20**	**42**
Producer accuracy %	**55**	**23**	**76**	**46**	**76**	**63**	**45**	**50**	

Classification result

Table 5.12. The error matrix for the SVM classification on Landsat OLI imagery with 30m spatial resolution,

		SVM					OBIA/SVM				
		Landsat OLI 30m	Pan Landsat OLI 15m	Sparse representation on Landsat OLI 2m	Sparse representation on Pan Landsat OLI 2m	WV-2 2m	Landsat OLI 30m	Pan-sharpening Landsat OLI 15m	Sparse representation on Landsat OLI 2m	Sparse representation on Pan Landsat OLI 2m	WV-2 2m
SVM	Landsat OLI 30m										
	Pan Landsat OLI 15m	0.8									
	Sparse representation Landsat OLI 2m	**3.8**	**3.3**								
	Sparse representation on Pan Landsat OLI 2m	**7.1**	**6.9**	**5.1**							
	WV-2	**9.4**	**8.8**	**7.2**	**4**						
OBIA/SVM	Landsat OLI 30m	1	0.9	**3.8**	**7.1**	**9.4**					
	Pan Landsat OLI 15m	1.2	0.8	2.7	**6.5**	**8.4**	1.3				
	Sparse representation on Landsat OLI 2m	**3.6**	3.1	0.5	**5.3**	**7.4**	**3.7**	2.5			
	Sparse representation on Pan Landsat OLI 2m	**5**	**4.9**	2.4	**4.3**	**5.7**	**5.1**	**4.6**	2.7		
	WV-2	**8.9**	**8.6**	**6.9**	3.2	1.4	**9**	**8.3**	**7**	**5.2**	

Table 5.13. The result of McNemar statistical test (statistically significant at 99% level of confidence shows bolded and underlined).

Table 5.14 shows the statistical differences between the 10 land cover maps. The significant differences with 99% level of confidence highlighted by bold and underline effects. Although the overall accuracy of performing sparse representation on Landsat OLI achieved 63% (the average between two classifiers), which is 27% less accurate than the original WV-2 imagery with 2m spatial resolution, performing the sparse representation increased the overall accuracy of Landsat OLI by 9%, for both SVM and OBIA/SVM classifiers. Additionally, according to the statistical test this difference was statistically significant at the 99.9% level of confidence, for both SVM and OBIA/SVM classifiers. Compared to pan-sharpening, performing sparse representation method on Landsat OLI increased the overall accuracy by 9% and 5% (for both SVM and OBIA/SVM classifiers, respectively) and this difference was statistically significant at 99% level of confidence. Nonetheless, the additional level of processing for such small improvement in the overall accuracy may not be beneficial and practical for such land cover classification.

Increasing the Landsat OLI's spatial resolution directly from 30 to 2m estimates 225 pixels with a 2m resolution from a Landsat OLI pixel with 30m resolution. Likewise, increasing the spatial resolution of pan-sharpened Landsat OLI from 15 to 2m means to estimate 56 pixels with a 2m resolution from a pan-sharpened Landsat OLI pixel with 15m resolution. This comparison shows the using sparse representation method was more challenging on Landsat OLI compared to pan-sharpened Landsat OLI.

Performing sparse representation on pan-sharpened Landsat OLI increased the overall accuracy of performing sparse representation on Landsat OLI by 10%

(average for both classifiers). This difference was statistically significant at the 99% level of confidence, for both SVM and OBIA/SVM classifiers. This mapping difference between performing sparse representation on pan-sharpened Landsat OLI and Landsat OLI increased (by 13%) when SVM classifier was used. Performing sparse representation on pan-sharpened Landsat OLI increased the overall accuracy of Landsat OLI by 19% (averaging between both SVM and OBIA/SVM classifiers). Although the overall accuracy of spatially increased pan-sharpened Landsat OLI was 18% less accurate than the overall accuracy of WV-2 imagery, it increased the overall accuracy of Landsat OLI by 22%, when SVM classifier was used.

The impact of using different classifiers is more complex compared to the variation associated with the different spatial enhancement techniques. Both SVM and OBIA/SVM classifiers achieved similar overall accuracy when Landsat OLI and its pan-sharpened version were used. The McNemar test showed that there was no significant statistical difference between these two classifiers at 30m resolution at 99% level of confidence. These results highlight that the effect of using different classifiers was marginalised due to an inadequate level of spatial details provided with Landsat OLI and its pan-sharpened version.

The SVM was the superior classifier compared to OBIA/SVM classifier for classifying both spatially increased Landsat OLI images with 2m resolution. While SVM advantage was negligible over OBIA/SVM when sparse representation performed on Landsat OLI imagery, SVM advantage was more evident when sparse representation performed on pan-sharpened Landsat OLI imagery in which it outperformed the overall accuracy of using OBIA/SVM by

6%. The result of the statistical test showed that there is no statistical difference between classifiers when sparse representation was performed on Landsat OLI imagery. However, there was statistical significant between SVM and OBIA/SVM, when sparse representation was performed on pan-sharpened Landsat OLI imagery.

OBIA/SVM was the worst classifier for both spatially enhanced Landsat OLI images. This can be potentially be attributed to the quality of extracted segments, which may not be accurately defined on spatially increased Landsat OLI images (Kavzogl and Yildiz, 2014). Perhaps, the lower quality of extracted segments can be due to the nature of sparse representation method, which changes the spectral values of the pixels.

The sparse representation method allocated the new spectral values for pixels in a way that a homogeneous feature in real world represents by several objects in which it should be only one object (in the best case scenario). The spectral values of many homogenous features have been individually checked to understand how sparse the representation method allocates new spectral values. It was that after identifying the edge of a feature, sparse representation allocated new spectral values of pixels within a homogenous feature (i.e., a roof tile feature) in the way that the spectral values gradually change to match the spectral values of edge pixels. However, for WV-2 imagery the spectral values of the pixel within the same feature were more and less similar. This is particularly problematic when the spectral values allocated to either side of a feature are significantly different. In this case, instead of one texture a ground feature is represented by several textures. Since segmentation algorithms use

texture change to identify the boundary of an object, this condition causes an over-segmentation problem, with each object formed by several smaller polygons. Over-segmentation is one of the reasons which reduces the accuracy of OBIA (Karl and Maurer, 2010). This significant over segmentation was identified during quality segmentation assessment.

5.8.1 Class level details

Since the main aim of this chapter is to assess the capability of the sparse representation method, the explanation of class details is represented by averaging both SVM and OBIA/SVM classifiers. However, the classifier advantage would be highlighted if there is a significant difference between classifier's accuracy in an individual class. Figures 5.9 and 5.10 represent the producer and user accuracy of eight individual land cover classes. These figures were generated by averaging the accuracy statistics of both SVM and OBIA/SVM classifiers.

5.8.1.1 Soil

Performing sparse representation on pan-sharpened Landsat OLI increased both user and producer accuracies of the bare soil of Landsat OLI imagery by 7% (Figure 5.9 and 5.10). In fact, using sparse representation whether on Landsat OLI or its pan-sharpened version increased the reliability of mapping bare soil consistently from 50% to 72% and 76% respectively. The major confusion of bare soil occurred towards roof tile class. This can be explained by the spectral similarity between these two classes since one of the materials used for roof tile (especially clay roof tile) is soil.

5.8.1.2 Asphalt

Performing sparse representation on pan-sharpened Landsat OLI increased the accuracy of Landsat OLI by 20%. Furthermore, performing sparse representation on pan-sharpened Landsat OLI increased the reliability of Landsat OLI for mapping asphalt by 15% and 30% for user and producer accuracies, respectively (Figure 5.9 and 5.10). The largest confusion occurred between asphalt and roof tiles classes. This confusion can be addressed by the image's lack of spatial details to distinguish between asphalt and roof tiles classes since buildings and roads as narrow linear urban features are spatially correlated. Additionally, the result of the last chapter highlighted the needs for a very fine resolution data set to accurately classify asphalt and roof tiles classes. Referring to the result of Chapter 3, increasing the spatial resolution from 30 to 10 to 4, and 2m with four spectral band sets images increased the accuracy of classifying asphalt from 47% to 62% to 86% and 88% respectively. Regarding the classifier influence, SVM was more successful in improving the accuracy of mapping asphalt class. For example, SVM improves the accuracy of asphalt of Landsat OLI to twice that of the OBIA/SVM classifier (SVM by 37% comparing to OBIA/SVM by 18%), for spatially increased Landsat OLI using sparse representation.

5.8.1.3 Water

Furthermore, performing sparse representation on Landsat OLI and its pan-sharpened version increased the accuracy of water class of Landsat OLI imagery by 5% and 14%, respectively (Figure 5.9 and 5.10). Since the majority of the water class consisted of large features including ponds, lake, and river, there

was not a significant difference among Landsat OLI, and it's both spatially increased versions to classify features of water classes.

5.8.1.4 Trees

Both spatially increased versions of Landsat OLI increased the classification accuracy of the tree class for the original Landsat OLI imagery by around 20%. In particular, spatially enhanced pan-sharpened Landsat OLI increased both producer and user accuracies of Landsat OLI imagery by 22% and 31%, respectively (Figure 5.9 and 5.10). The tree class was classified with confusion against grass and roof tiles classes.

Perhaps, the confusion of tree to roof tiles occurred due to the interaction of trees with neighbouring buildings (Corr *et al.,* 2003) which may lead to overlapping between the boundaries of tree canopies and buildings. Furthermore, the difficulty in distinguishing between trees and grass classes can be referred to the spectral similarity between the natures of these two vegetation types (Rokitnicki-Wojcik *et al.,* 2011).

5.8.1.5 Roof tile

Performing the sparse representation on pan-sharpened Landsat OLI increased the accuracy of mapping the roof tiles class of Landsat OLI by 9%. Despite the 76% producer accuracy of Landsat OLI to for classifying the roof tiles class, the poor user accuracy of Landsat OLI imagery provides an indication of the probability that only 38% of testing samples allocated to roof tiles class represent the roof tiles on the ground. There was significant confusion between roof tiles and other classes (i.e., bare soil, asphalt, trees, slate roofs,

and grass) and this confusion more likely suggests overestimation for classifying the roof tiles class. The reason for this confusion can be attributed to the highly dense distribution of roof tiles (which represent most residential urban buildings) with its nearby spatially correlated classes. This is particularly problematic for using coarse resolution imagery leading to a mixed pixel problem. A mixed pixel contains several land cover classes, and the class has with the largest proportion would be allocated to that pixel. This condition possibly addresses roof tile fine accuracy and poor reliability.

5.8.1.6 Metal roof

Metal roof class mainly represents large building use in the industrial sector. Due to the size of metal features which are greater than Landsat OLI pixel size, it obtained good producer and acceptable user accuracy for Landsat OLI imagery. Despite slight enhancement in reliability, the accuracy of metal roof class increased significantly by 28% and 25%, using spatially increased Landsat OLI and pan-sharpened Landsat OLI, respectively (Figure 5.9 and 5.10). The most significant confusion occurred between the metal roofs and bare soil classes using the Landsat OLI similarly for both classifiers. In this case, nine metal roof testing samples (23%) were classified as bare soil. Reviewing these ground features of metal roofs class showed that a number of testing samples are located at the boundaries of metal roof class and these geographical structures are mostly surrounded by bare soil features. As a result, this confusion can be addressed by spatial correlation and insufficient level of

spatial information to accurately map between the metal roof and bare soil classes.

5.8.1.7 Slate roof

Performing sparse representation on pan-sharpened Landsat OLI increased the mapping producer and user accuracies of slate roofs class for Landsat OLI imagery by 30% and 16%, respectively (Figure 5.9 and 5.10). The slate roof is the second major building roof type, and it mostly represents residential buildings. There was confusion between slate and both roof tile and asphalt classes. The confusion between the slate roof and roof tile classes can be attributed to spectral similar between these two classes due to containing similar materials. The confusion between the slate roof and asphalt classes can be addressed by the spatial correlation between buildings and road, particularly in residential areas.

5.8.1.8 Grass

Both spatially enhanced Landsat OLI increased the producer and user accuracies of the classification of grass class using the Landsat OLI imagery. Particularly, performing sparse representation on pan-sharpened Landsat OLI increased the Landsat OLI accuracy and reliability by 33% and 25%, respectively (Figure 5.9 and 5.10). Tree and roof tile are two classes which were confused mostly with the grass class. The source of confusion between grass and roof tile can be related to the spatial correlation between residential buildings and their lawn gardens. Further, the confusion between grass and trees classes can be

attributed to the spectral similarity between the nature of these two vegetation types (Dillabaugh and King, 2008).

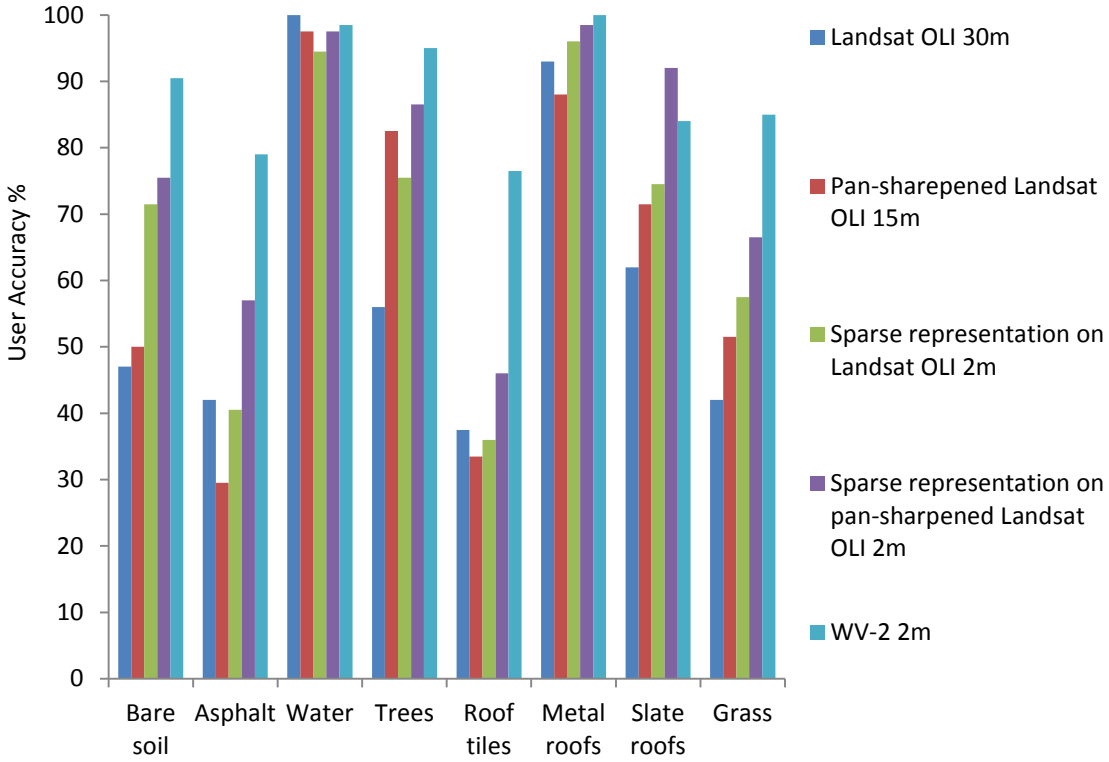


Figure 5.9. The user accuracy of each class for five classified land cover maps (by averaging two classifiers).

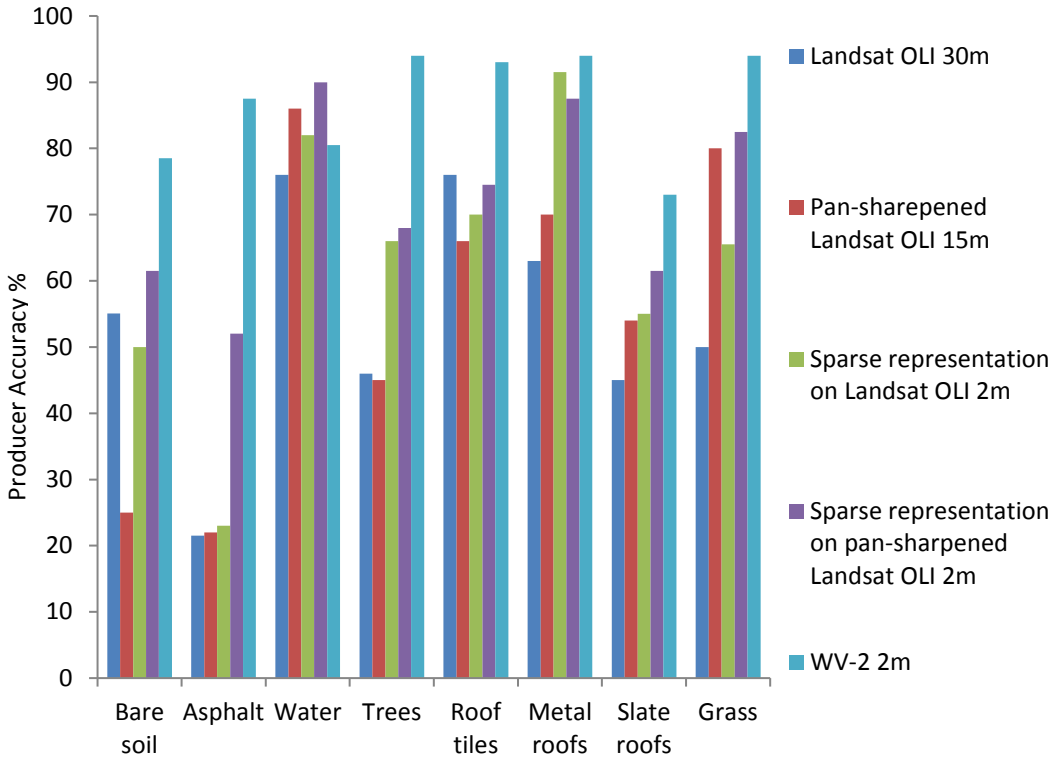


Figure 5.10. The producer accuracy of each class for five classified land cover maps (by averaging two classifiers).

5.9 Summary

The coarse spatial resolution of freely available medium resolution remotely sensed imagery is one of the main factors for inaccurately mapping complex urban areas. This chapter tried to overcome Landsat OLI's inadequate spatial resolution to accurately map complex urban areas by proposing new sparse representation techniques using a single training image. Besides increasing the spatial resolution of the Landsat OLI directly from 30 to 2m, the second approach was to perform sparse representation to increase the spatial resolution of the pan-sharpened Landsat OLI from 15 to 2m.

The results demonstrated the capability of using sparse representation to increase the spatial resolution of the Landsat OLI for mapping urban areas in

details. While performing the sparse representation on Landsat OLI increased the overall accuracy of Landsat OLI by 9% (using SVM classifier), performing the sparse representation on pan-sharpened Landsat OLI increased the overall accuracy of Landsat OLI by 22% (using SVM classifier).

Regarding class level analysis, both spatially increased Landsat OLI images increased Landsat OLI's accuracy for eight classes, while this improvement was insignificant for water and metal roof classes.

Furthermore, both SVM and OBIA/SVM similarly increased the overall accuracy of Landsat OLI by performing the sparse representation on Landsat OLI. However, SVM was more successful classifier (increased the overall accuracy 6% more than OBIA/SVM) when sparse representation was performed on pan-sharpened Landsat OLI.

This Chapter demonstrated the advantage of sparse representation to increase the spatial resolution of Landsat OLI for mapping urban areas. However, both input coarse resolution imagery (Landsat OLI) and VHR training imagery (WV-2) represent similar locations and time period, so there is no real need to use spatially enhanced methods when the very fine resolution imagery is available. In fact, the capability of the sparse representation method can be fully demonstrated in wider ranges of applications where there is a spatial and temporal disparity between Landsat data and WV-2. Regarding the spatial disparity between Landsat data and WV-2 images, a single WV-2 imagery of Nottingham can be used as a training sample to increase the spatial resolution of Landsat data from the different city. Additionally, regarding the temporal separation between Landsat data and WV-2 images, a single WV-2 imagery of

Nottingham can be used as a training sample to increase the spatial resolution of historical Landsat image of Nottingham for land cover mapping. This can open up new areas for considering the Landsat archive which has been marginalised limited due to their coarse spatial resolution.

Chapter Six: Assessing the Potential of Sparse Representation Method for Transferability Over Space and Time

6.1 Introduction

The previous chapter demonstrated the effectiveness of sparse representation to enhance the spatial capability of Landsat OLI data for land cover mapping of the urban area. This spatially increased method was successful at improving the overall accuracy of Landsat OLI by 22%. However, because the training sample (WV-2 imagery) was spatially and temporally coincident, this result may not fully demonstrate the capability of the sparse representation method for wider urban mapping applications (i.e., historical), when there is a spatial and temporal separation between Landsat data and VHR training image. Therefore, this Chapter focuses on assessing the potential of the sparse representation method to enhance the spatial capability of Landsat data by considering the spatial and temporal transferability.

This chapter is organised as follows: Section Two explains the research material including the study area, the data sets and classification scheme which were used. Section Three describes the pan-sharpening techniques and implemented assessment approaches. Section Four explains the accuracy of the sparse representation of four steps including pre-processing, training phase, reconstruction phase, and post processing. Section Five discusses the classification procedure in two parts: training and pixel\object labelling.

Section Six describes the accuracy assessment and McNemar statistical test procedure performed on the land cover maps. Section Seven presents the results. Finally, Section Eight summarises this chapter.

6.2 Research Materials

6.2.1 Study area

Nottingham was again the focus of study, but in addition to this, the city of Derby was an area of interest. Section 5.2.1 has the details relating to Nottingham, so here Derby is described. The city of Derby is located at 52.92 degrees latitude and -1.48 degrees longitude within Derbyshire County, as part of the East Midlands. It has a flat topography with an average elevation of 55 meters above sea level. The population of Derby was 248,752 in 2011, an increase of 7.8% since 2001 (Council, 2016). The City of Derby is surrounded by the River Trent valley in the south, the Nottinghamshire, Derbyshire and Yorkshire Coalfields in the east, the South Derbyshire Clay-lands in the west, and the Derbyshire Peak Fringe in the north. Figure 6.1 shows the urban areas of Derby and its suburbs through the Landsat OLI sensor with 30m resolution.

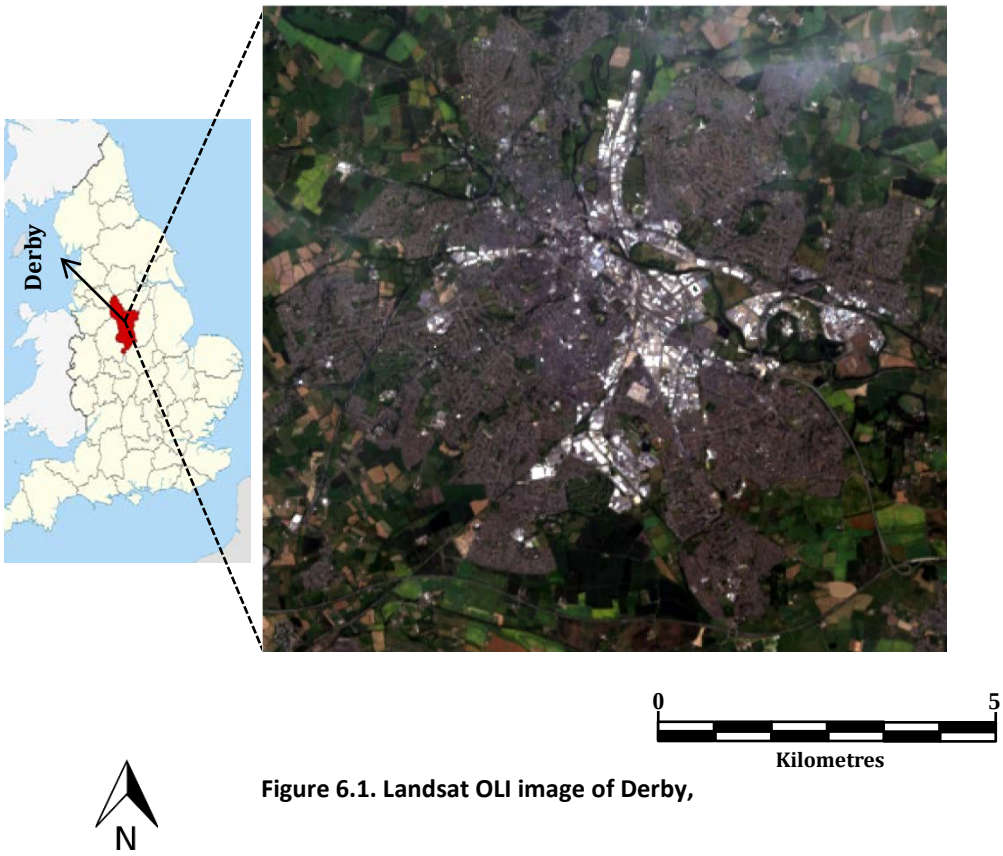


Figure 6.1. Landsat OLI image of Derby,

6.2.2 Data and classification scheme

In total, three multi-spectral data sets images were used in this Chapter. In addition to the previously used Landsat OLI data, a Landsat Enhanced Thematic Mapper (ETM+) image was used. The third image was that of the WV-2 data used in the sparse representation approach to spatially enhance the aforementioned medium resolution Landsat scenes. As well as capturing the city of Nottingham the Landsat OLI data had also captured the city of Derby. This enabled the training statistics that had been developed for the area of Nottingham using the spatially coincident WV-2 data of Nottingham to be applied to Derby (affording spatial transferability), an area of 139km^2 (12.2 × 11.4 km). The Landsat ETM+ data was also spatially enhanced using WV-2 data. In this case, the two datasets are temporally different (affording temporal transferability); the WV-2 was captured on the 26th June 2012 and the Landsat

ETM+ on 16th May 2000. Figure 6.2 depicts the imagery of Nottingham captured by Landsat ETM+ with 30m resolution. For both Landsat OLI and WV-2 sensors' spectral properties, please refer to Table 5.1 in Chapter 5 and Figure 3.2 in Chapter 1 respectively.

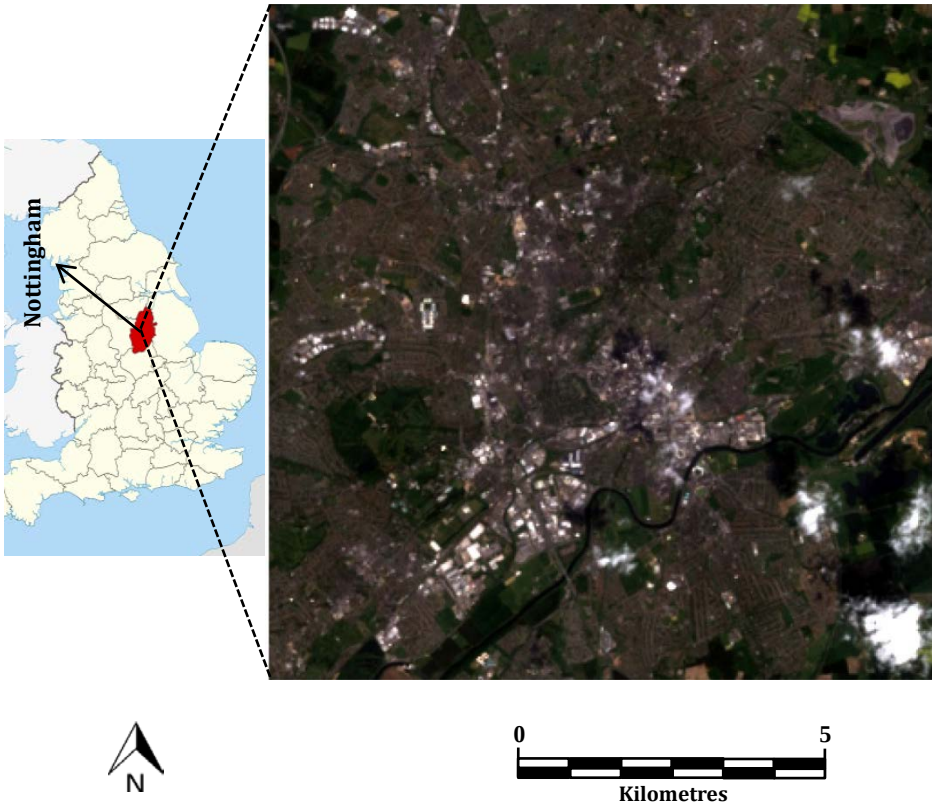


Figure 6.2. Landsat ETM+ image captured Nottingham on 16th of May 2000.

Landsat ETM+ become operational earlier at 15th April 1999, after the Landsat ETM failure to reach the planned orbit, (Dunbar, 2016). As a seventh generation Landsat programme, it captures Earth's features in eight spectral bands including panchromatic band (Table 6.1). Since Landsat ETM+ was the first sensor offering a panchromatic band with 15m resolution, it was selected to assess the accuracy of both spatially increased Landsat ETM+ using sparse representation and spatially increased pan-sharpened Landsat ETM+ using sparse representation on historical medium resolution applications.

Despite having eight spectral bands, only four of the WV-2 spectral bands can be considered in the sparse representation process. These were chosen through the spectral similarity requirement between the training image (WV-2) and input coarse resolution images (Landsat ETM+ and Landsat OLI) used in the sparse representation method.

Bands number	Band name	Spectral range	Spatial resolution
Band 1	Blue	450 – 515 nm	30m
Band 2	Green	525 – 605 nm	30m
Band 3	Red	630 – 690 nm	30m
Band 4	Near Infrared (NIR)	775 – 900 nm	30m
Band 5	Short-wave Infrared (SWIR) 1	1550– 1750 nm	30m
Band 6	Thermal	10400 – 12500 nm	60m
Band 7	Short-wave Infrared (SWIR) 2	2080 – 2350 nm	30m
Band 8	Panchromatic	520 – 900 nm	15m

Table 6.1. The Landsat ETM+ spectral properties and their corresponding spatial characteristics (source: USGS, 2015).

Figure 5.2 in Chapter Five demonstrates the full spectral range of both WV-2 and Landsat OLI images. Only coastal blue, blue, green, red, and near-infrared bands are similar between WV-2 and Landsat OLI. However, coastal blue was found to be ineffective for mapping urban areas in the first research chapter. According to Figure 6.3, blue, green, red, and NIR spectral bands are similar between WV-2 and Landsat ETM+. Therefore, the spectral range of WV-2, Landsat OLI and Landsat ETM+ were limited to four spectral bands range including blue, green, red, and NIR.

Similar to the previous chapter, once the Landsat scenes had been spatially enhanced, the cities of Nottingham (Landsat ETM+) and Derby (Landsat OLI)

were classified into classes in the former case and seven in the latter. The same classes were considered in both cases (Table 6.2), however, because of the insignificant number of buildings featuring a slate roof tile in Derby, this class was removed from the classification scheme in Derby. The table below explains the classification scheme adopted for Derby.

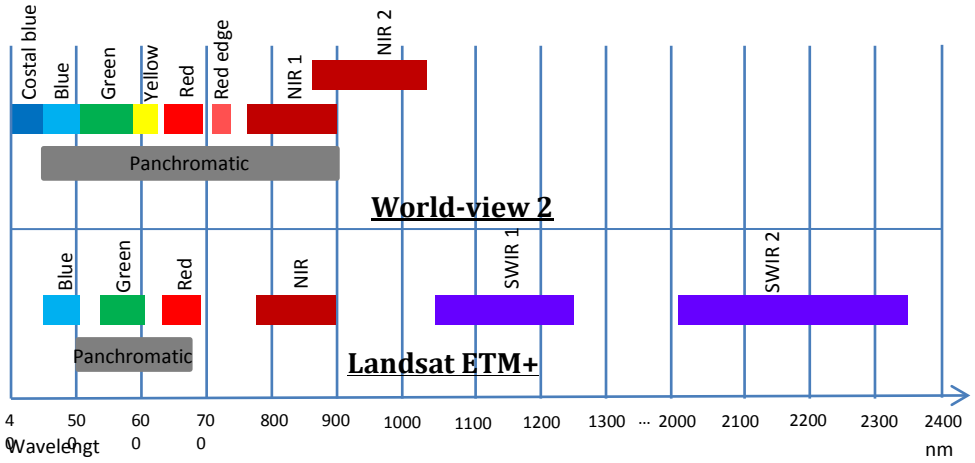


Figure 6.3. Comparing the distribution of spectral properties of Landsat ETM+ (without thermal band) and WV-2 images.

No	Class	Description
1	Asphalt	Urban ground surfaces covered in asphalt such as roads and car parks
2	Roof tiles	Predominantly residential buildings covered in red and dark grey concrete tiles
3	Metal roofs	Predominantly industrial buildings covered in white metal panels
4	Grass	Areas of grassland such as urban parks and lawns, plus surrounding rural agriculture
5	Tress	Patches of deciduous broad-leaved and evergreen needle-leaved trees
6	Bare soil	Open areas covered by bare soil
9	Water	Water bodies including lakes, rivers, ponds and canals

Table 6.2. The land cover classification scheme, which for land cover mapping of Derby.

6.3 Pan-sharpening of the Landsat OLI

As done in Chapter 5, in addition to applying the sparse representation method to the 30m spatial resolution Landsat scenes, sparse representation was also

applied to the Landsat scenes after they had undergone pan-sharpening using the accompanying 15m panchromatic waveband.

In the aforementioned chapter, the accuracy of five pan-sharpening approaches for the sharpening of the Landsat OLI scene had been evaluated using both qualitative and quantitative assessment and the HCS technique was deemed the most successful to be used going forward. Due to the similarity in the sensors between the Landsat OLI and the ETM+ scene it is assumed that HCS would be the most suitable pan-sharpening technique for this chapter too. Therefore, the HCS pan-sharpening technique was chosen to enhance the spatial resolution of both the Landsat OLI data of Derby and historical Landsat ETM+ data of Nottingham from 30 to 15m using their panchromatic bands with a 15m spatial resolution.

6.4 Sparse representation analysis

6.4.1 Introduction

The sparse representation method was applied to the following four scenes using the WV-2 data for Nottingham:

1. Landsat OLI of Derby with 30m spatial resolution,
2. Pan-sharpened Landsat OLI with 15m spatial resolution,
3. Landsat ETM+ of Nottingham with 30m spatial resolution,
4. Pan-sharpened Landsat ETM+ with 15m spatial resolution,

To reiterate, due to the need for spectral similarity between the WV-2 data and the Landsat data, similarly to the previous Chapter only these four spectral bands (blue, green, red, and near-infrared) of both data sets were considered.

6.4.2 Pre-processing

Pre-processing is limited to radiometric conversion and spectral separation. Both radiometric conversion and spectral separation were performed on four medium resolution images including original and pan-sharpening Landsat OLI and Landsat ETM+. Despite WV-2 as a training image with 11 bits radiometric (ranging from 0 to 2047), Landsat OLI and Landsat ETM+ were obtained in 16 and 8 bit respectively. So, WV-2's digital number was converted from 11-bit radiometric resolution into 16 and 8 bits radiometric resolution , respectively. Initially, the lowest value was subtracted from the digital numbers for the entire image in each spectral band, and then, individual digital numbers were multiplied into 32 ($\frac{2^{16}}{2^{11}}$) and $\frac{1}{8}$ ($\frac{2^{8}}{2^{11}}$) for both original and pan-sharpened Landsat OLI and both Landsat ETM+ respectively. Next, since the MATLAB code was re-programmed based on imagery with the single band as "tif" format, each of four images data sets were saved as four single band images with "tif" format, creating sixteen single band image data sets.

6.4.3 Training phase

The WV-2 imagery of Nottingham with 2m resolution was the only fine resolution image used as a training image to increase the spatial details of the four Landsat multi-spectral data sets. The training parameters to be employed in the sparse representation process were set according to values in the previous chapter. As a result, 2048 and 100,000 were allocated as dictionary sizes and the number of patch samples respectively. The zoom factor and patch size for both original Landsat OLI and Landsat ETM+ were chosen as 15 and 30

respectively, while 7.5 and 15 were allocated to the zoom factor and patch size of both pan-sharpened Landsat OLI and Landsat ETM+. A detailed explanation of the possible impacts of parameters on output image and the reasons for choosing these values described in the previous chapter in Section 5.5.3.

6.4.4 Reconstruction phase

The reconstruction phase aims to estimate the spatially increased versions of two input coarse resolution images. Using both H-resolution and L-resolution resolution dictionaries from the training phase, the lost texture and edges of those four images were recovered within the reconstruction phase. The key parameter in the reconstruction phase is the zoom factor.

The four Landsat images (including original and pan-sharpened versions with 30 and 15m spatial resolutions, respectively) spatially increased to images with 2m spatial resolution. As a result, the zoom factor 15 was allocated to both Landsat OLI and Landsat ETM+, and zoom factor 7.5 was chosen for their corresponding pan-sharpened images.

6.4.5 Post-processing

A series of post-processing steps were performed to prepare images for the next step. Since the output images do not contain any spatial referencing, each single band image was georeferenced and was defined according to the OS MasterMap Topography map. Then, an NDVI ($\frac{nir - red}{nir + red}$) image was generated for each of those four image groups. As a last step, four single band images, plus NDVI for each of four different image groups were merged together. This

resulted in four spatially enhanced sets of the multispectral images with 2m resolution.

6.5 Classification

6.5.1 Training procedure

Landsat OLI of Derby and Landsat ETM+ of Nottingham and their spatially increased versions were considered for classification. So in total, three land cover maps were produced for each of these urban areas. In a similar fashion to the previous research chapters, two different pattern strategies were adopted. A block of 3×3 training pixels was used for those spatially increased Landsat OLI/ETM+ images with 2m resolution. However, a single pixel pattern was used for training strategy of the original Landsat OLI/ETM+ with 30m spatial resolution. The advantages of using a block of 3×3 pixels are explained previously in Section 3.4.1.

Training samples of Derby were collected based on online mapping resources including Google Map and Microsoft Bing, whereas very fine resolution images from Historical Google Earth were used as the main resource to create training database for the historic Nottingham images (Gu *et al*., 2015, Tsutsumida and Comber, 2015). An average of 270 and 340 training pixels were selected per class for both spatially enhanced images from historical Nottingham and Derby respectively. Sample sizes are larger than normally (Mather and Koch, 2011) recommended, which ranges from 50 to 150 as a minimum requirement per class. Training samples were spatially similar among three land cover maps for

Derby and Nottingham to offer direct comparability between the results of three images for each of the cities.

The TD technique was used to assess the spectral separability between the training samples of each class. As a part of this process, the possible training samples which hold the mixed spectral signature were removed from training database. This process was carried out until an adequate number of training samples was selected, and the training samples mixing rate among all classes does not exceed the recommended value, according to (Richards and Jia, 1999). Because of coarser resolution of Landsat OLI/ETM+ compared to their spatially increased versions, some of those training samples were used on both spatially increased Landsat (OLI/ETM+) images did not represent that specific class (as result of a mixed pixel) on the original Landsat (OLI/ETM+). As a result, those mixed training samples were removed from the training database. This caused a reduction in the number of training pixels in both Landsat OLI and Landsat ETM+ to 38 and 30, respectively; meaning it was not always possible to achieve the recommended sample size.

6.5.2 Classification approaches

Since the advantages of pixel-based SVM was demonstrated against object-based SVM, SVM was the only classifier used for performing classification. The radial basis function was chosen as the kernel type since this non-linear kernel is suitable for classifying a complex environment (Marjanovic *et al.,* 2011). Based on a trial-and-error strategy, 500 and 0.5 values were allocated for penalty and gamma parameters of the SVM.

6.6 Accuracy assessment and statistical testing

Following the classification process, a classification accuracy assessment was performed to validate the correctness of those six thematic land cover maps. Since there are two different study areas, two different testing sample groups were created using a random sampling design strategy. Also, testing samples were collected as a point-based system. The total number of testing samples for historical Nottingham was greater than for Derby because the classification scheme is different between these areas of interest. The total testing samples selected for historical Nottingham was 288 points, ranging from 19 to 47 for soil and roof tiles classes respectively. On the other hand, 233 testing points were selected in total for Derby ranging 29 to 43 for soil and asphalt classes respectively. These testing sets were used on all the six resultant land cover classifications as appropriate.

Similar to the previous research chapter, a confusion matrix was the core of the accuracy assessment. A confusion matrix offers the fundamental description of classification accuracy by characterising its accuracy, including producer and user accuracy. In addition to the confusion matrix, the McNemar statistical test was performed to assess the degree of statistical difference between each pair of classification land cover maps.

6.7 Results

In total, six land cover maps were produced by performing SVM classifier on three images for both cities: Landsat OLI, spatially increased Landsat OLI, and spatially increased pan-sharpened Landsat OLI were classified for Derby, while

Landsat ETM+, spatially increased Landsat ETM+, spatially increased pan-sharpened Landsat ETM+ were classified for Nottingham.

Previously, the effectiveness of the two proposed remotely sensed images using spatially enhancement methods on Landsat OLI imagery has been demonstrated for mapping urban environments. The main aim of this chapter was to demonstrate the capability of sparse representation by consideration of the spatial and temporal transferability elements to spatially enhanced Landsat data for urban land cover mapping. The accuracies of both spatially increased images were compared against their corresponding original Landsat images (i.e., Landsat OLI for Derby and Landsat ETM+ for Nottingham).

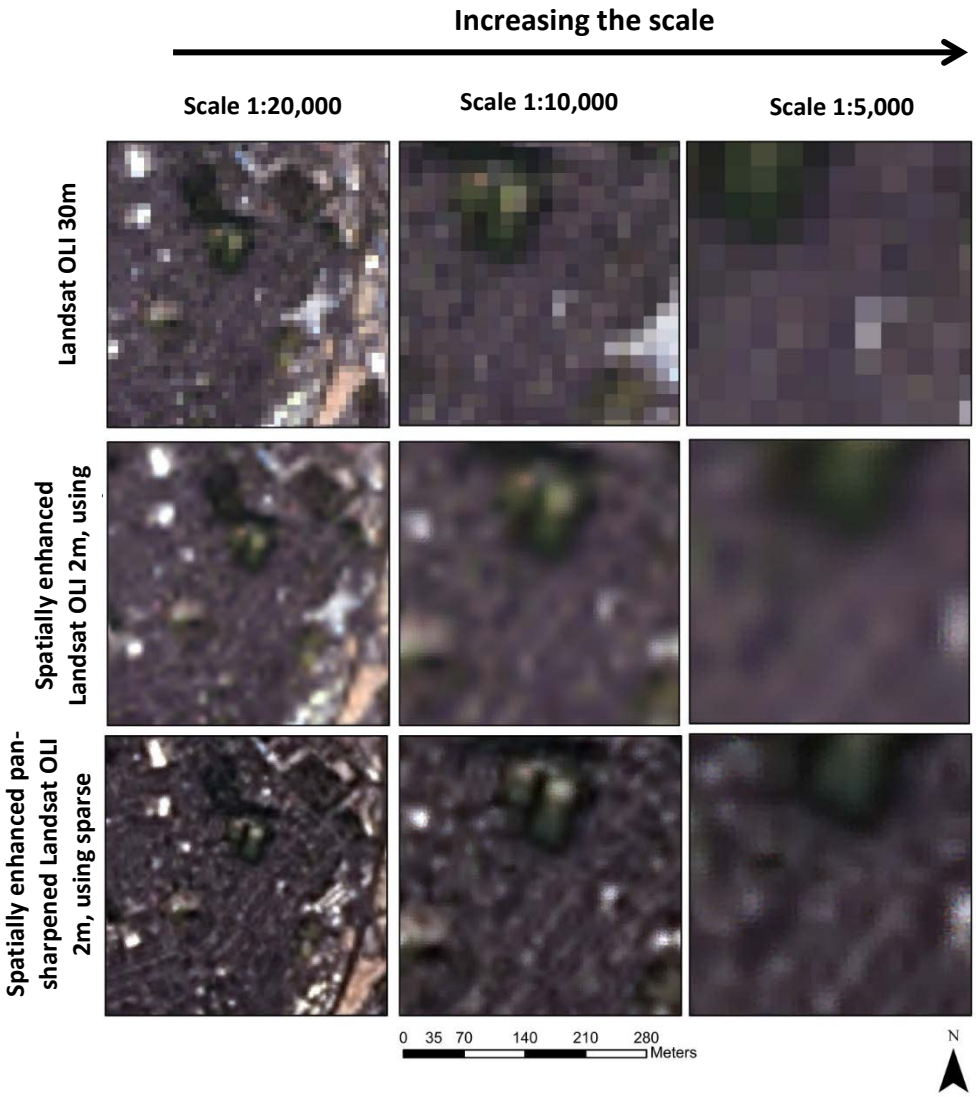


Figure 6.4. 9 small subsets of three different images of Derby (i) Landsat OLI 30m, ii) spatially enhanced Landsat OLI 2m using sparse representation, iii) spatially enhanced pan-sharpened Landsat OLI 2m using sparse representation) according to three scales (1:20,000, 1:10,000, and 1:5000).

Figure 6.4 shows the visual comparison between Landsat OLI imagery of Derby and its two spatially enhanced images versions through the three scales (1:20,000, 1:10,000, and 1:50,000). According to this Figure, performing the sparse representation method on both original and pan-sharpened Landsat OLI successfully increased the spatial information of the Landsat OLI of Derby using WV-2 from Nottingham. In particular, the

spatially increased pan-sharpened Landsat OLI offers the finest spatial information.

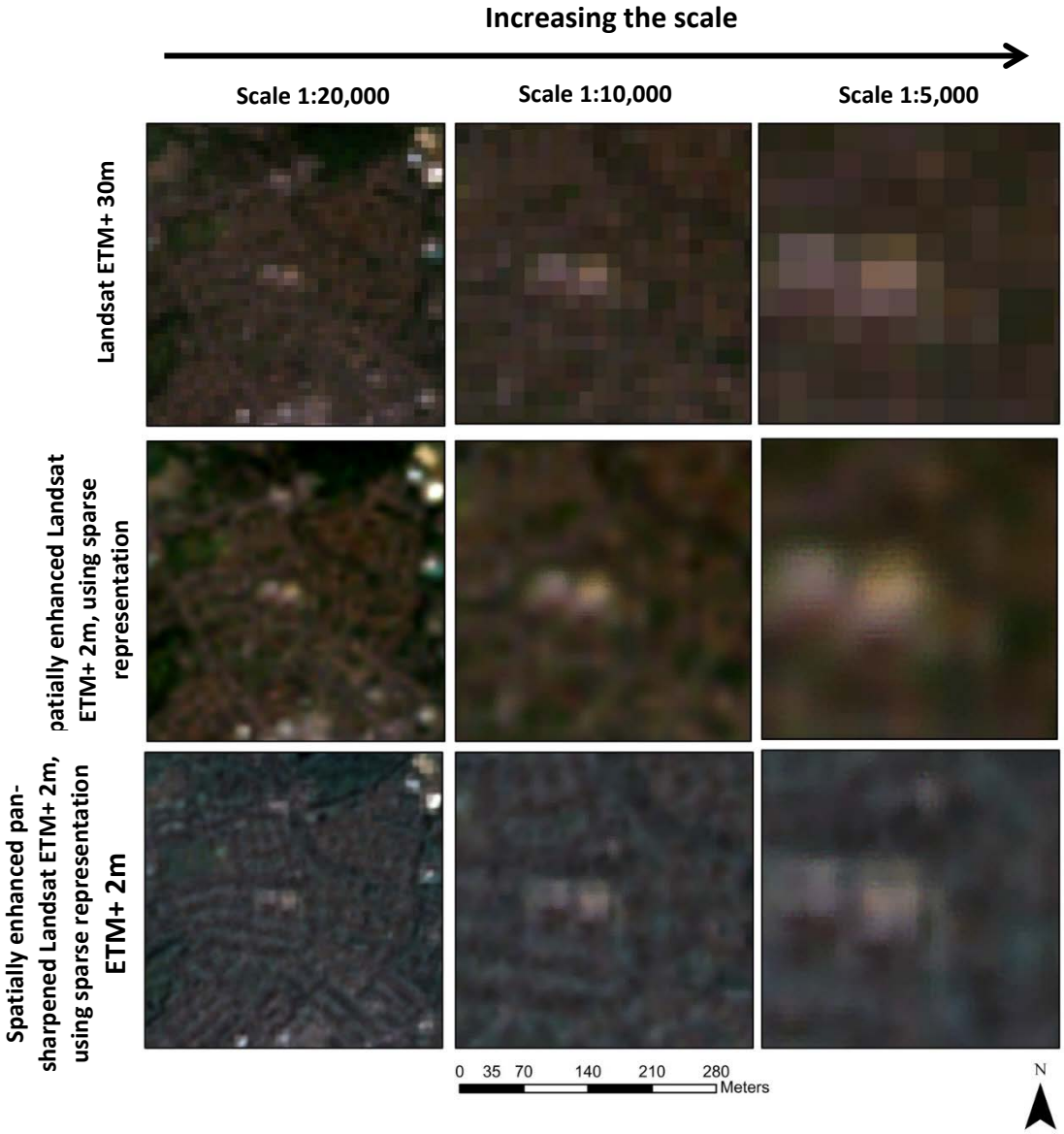


Figure 6.5. 9 small subsets of three different images of historical Nottingham (i) Landsat ETM+ 30m, ii) spatially enhanced Landsat ETM+ 2m using sparse representation, iii) spatially enhanced pan-sharpened Landsat ETM+ 2m using sparse representation) according to three scales (1:20,000, 1:10,000, and 1:5000).

Figure 6.5 offers the visual comparison between Landsat ETM+ imagery of Nottingham and its two spatially increased images versions through the three

scales (1:20,000, 1:10,000, and 1:50,000). A single WV-2 image covered Nottingham on May 2012 was used as the training image to increase the spatial resolution of the Landsat ETM+ image. This Figure illustrates the success of sparse representation to increase of the Landsat ETM+ for historical applications. Regarding spatial transferability of sparse representation methods, the sparse representation was also successful increasing the spatial resolution of the Landsat ETM+ for historical applications.

Moving to the classification result, a series of a small subset of both Derby and Nottingham are presented in Figure 6.6. These six small subset images provide a visual assessment of the potential of sparse representation method, while because these figures cover a small subset (700m × 700m) of the full extent of study, they should be interpreted with some caution. As Figure 6.6 shows, although the subsets images for both Nottingham and Derby represent dense residential areas; four spatially increased Landsat OLI and ETM+ offer a greater level of thematic information. Performing the sparse representation on pan-sharpened Landsat OLI/ETM+ was more successful in offering a greater level of spatial information compared to performing the sparse representation on Landsat OLI/ETM+. According to figure 6.6 classifying the spatially enhanced both Landsat OLI and Landsat ETM+ created sharper boundaries among boundaries buildings, roads, and vegetated areas while classifying the spatially enhanced both pan-sharpened Landsat OLI and Landsat ETM+ helped to distinguish between grass and tree (red and green colour respectively).

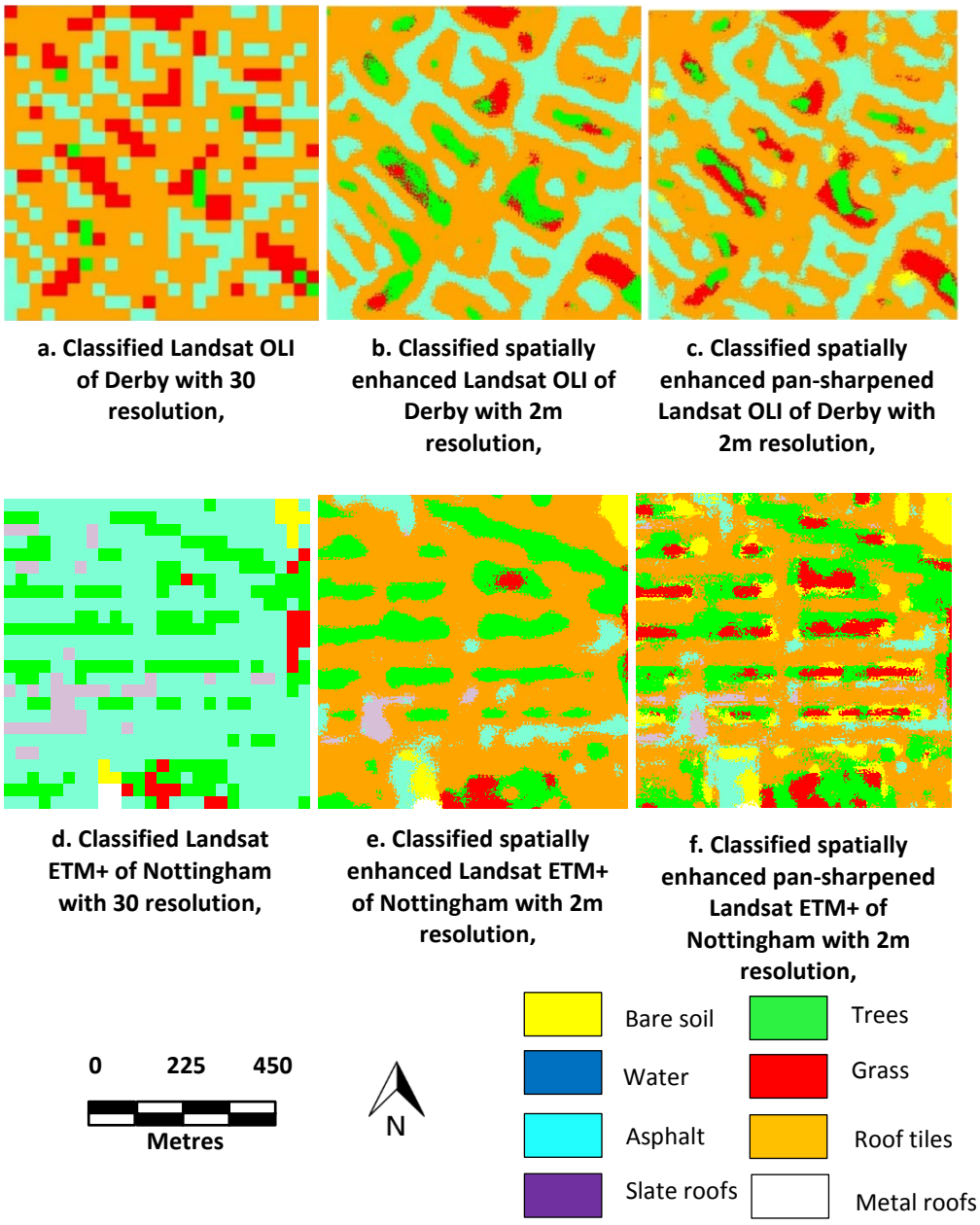


Figure 6.6, 6 small subsets of classified Land cover maps (a. Landsat OLI of derby, b. spatially enhanced Landsat OLI of Derby, c. spatially enhanced pan-sharpened Landsat OLI of Derby, d. Landsat ETM+ of Nottingham, e. spatially enhanced Landsat ETM+ of Nottingham, f. spatially enhanced pan-sharpened Landsat ETM+ of Nottingham,

Figure 6.7 presents a summary of the overall classification accuracy of sparse representation for Derby and Nottingham. Both spatially increased images improved their original overall accuracies of Landsat OLI and Landsat ETM+. Performing the sparse representation on pan-sharpened Landsat OLI and

Landsat EM+ achieved the highest accuracy with 79% and 71% for Derby and Nottingham, respectively. Performing the sparse representation increased the overall accuracy of Landsat OLI and Landsat ETM+ by 10% and 7% respectively. However, performing the sparse representation on pan-sharpened Landsat OLI and Landsat ETM+ increased the overall accuracy of their corresponding Landsat OLI and Landsat ETM+ by 20% and 21%, respectively.

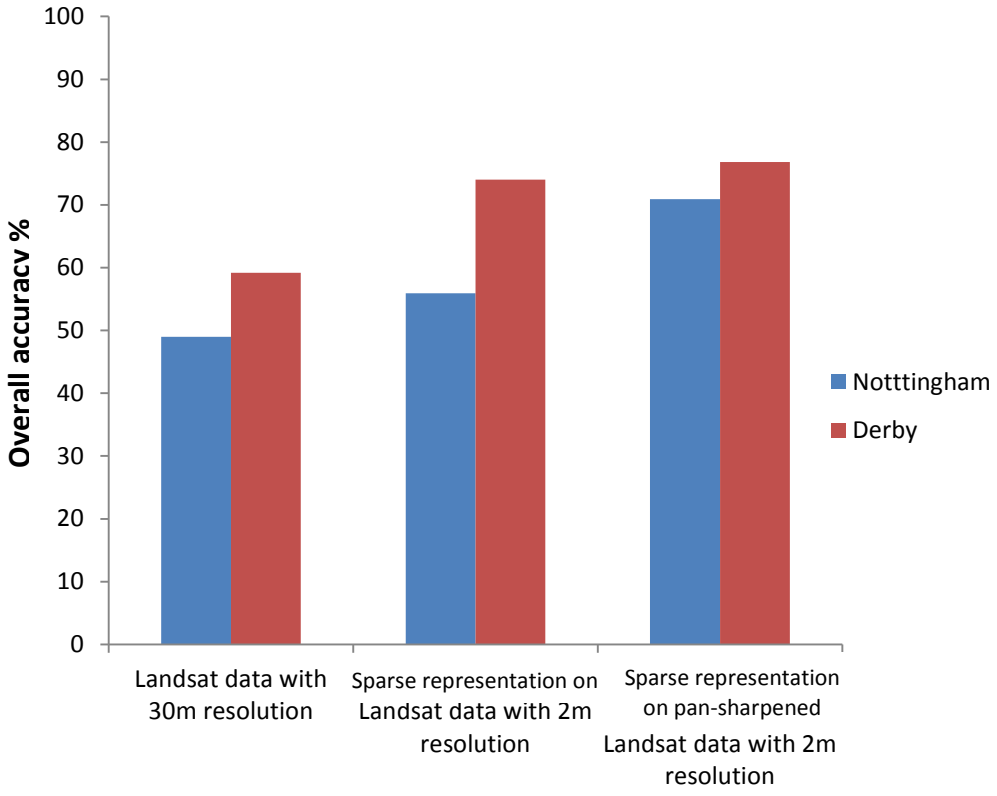


Figure 6.7. The overall accuracies for performing the sparse representation on original and pan-sharpened Landsat data (i.e., OLI for Derby and ETM+ for Nottingham).

The results showed the capability of sparse representation to successfully enhanced the spatial details of Landsat data sets for urban land cover mapping, even when there is a spatial and temporal separation between a single training image and both Landsat OLI and Landsat ETM+ images. Regarding the spatial

transferability, performing the sparse representation method increased the overall accuracy of Landsat OLI of Derby by 10%. Regarding the temporal transferability, performing the sparse representation method increased the overall accuracy of Landsat ETM+ of Nottingham by 8%.

Below, the four error matrices of six land cover maps (three for Landsat OLI of Derby and 3 for Landsat ETM+ of Nottingham) are provided in six tables. The remaining error matrices are provided in the Appendix (Page 285).

Overall Accuracy: 49%	**Reference data**								
Classification result / **Class**	**Bare soil**	**Asphalt**	**Water**	**Trees**	**Roof tiles**	**Metal roofs**	**Slate roofs**	**Grass**	**User accuracy %**
Bare soil	**4**	0	0	0	1	5	0	0	**40**
Asphalt	10	**25**	3	7	39	6	8	8	**24**
Water	0	0	**22**	0	0	0	1	0	**96**
Trees	1	3	4	**22**	2	1	2	6	**54**
Roof tiles	0	0	0	1	**0**	0	0	0	**0**
Metal roofs	5	0	1	0	1	**15**	0	0	**68**
Slate roofs	2	9	1	0	3	1	**27**	0	**63**
Grass	2	0	0	9	1	2	0	**22**	**61**
Producer accuracy %	**17**	**68**	**71**	**56**	**0**	**50**	**71**	**61**	

Table 6.3. The error matrix of performing SVM on Landsat ETM+ of Nottingham with 30m spatial resolution.

Classification result	Overall Accuracy: 59% / Class	Reference data: Bare soil	Asphalt	Water	Trees	Roof tiles	Grass	Metal roofs	User accuracy %
	Bare soil	**25**	1	0	1	0	0	0	**93**
	Asphalt	2	**23**	1	0	10	6	0	**55**
	Water	0	2	**19**	1	3	0	0	**76**
	Trees	0	1	9	**13**	0	0	9	**41**
	Roof tiles	8	15	2	4	**18**	1	4	**35**
	Grass	1	1	0	0	0	**19**	0	**90**
	Metal roofs	3	0	0	7	3	0	**21**	**62**
	Producer accuracy %	**64**	**53**	**61**	**50**	**53**	**73**	**62**	

Table 6.4. The error matrix of performing SVM on Landsat OLI of Derby with 30m spatial resolution.

Classification result	Overall Accuracy: 74% / Class	Reference data: Bare soil	Asphalt	Water	Trees	Roof tiles	Grass	Metal roofs	User accuracy %
	Bare soil	**31**	1	0	0	0	0	1	**94**
	Asphalt	4	**29**	1	0	8	0	2	**66**
	Water	0	0	**28**	1	0	0	0	**97**
	Trees	0	0	2	**19**	2	9	0	**59**
	Roof tiles	3	12	1	2	**23**	4	0	**51**
	Grass	0	0	0	7	0	**21**	0	**75**
	Metal roofs	0	1	0	0	0	0	**25**	**96**
	Producer accuracy %	**82**	**67**	**88**	**66**	**70**	**62**	**89**	

Table 6.5. The error matrix of performing SVM on spatially increased Landsat OLI of Derby with 2m spatial resolution, using sparse representation.

Overall Accuracy: 77%	Reference data							
Classification result / Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Grass	Metal roofs	User accuracy %
Bare soil	**33**	4	0	0	1	3	0	**80**
Asphalt	2	**26**	0	0	5	0	0	**79**
Water	0	3	**29**	0	0	0	0	**91**
Trees	0	0	2	**22**	1	0	9	**65**
Roof tiles	3	10	1	5	**26**	0	4	**53**
Grass	0	0	0	0	0	**25**	0	**100**
Metal roofs	0	0	0	2	0	0	**21**	**91**
Producer accuracy %	**87**	**60**	**91**	**76**	**79**	**89**	**62**	

Table 6.6. The error matrix of performing SVM on spatially increased pan-sharpened Landsat OLI of Derby with 2m spatial resolution, using sparse representation.

Both Figures 6.8 and 6.9 represent the result of McNemar statistical test at the 99.9% level of confidence for land cover mapping of Landsat OLI of Derby and Landsat ETM+ of Nottingham, respectively.

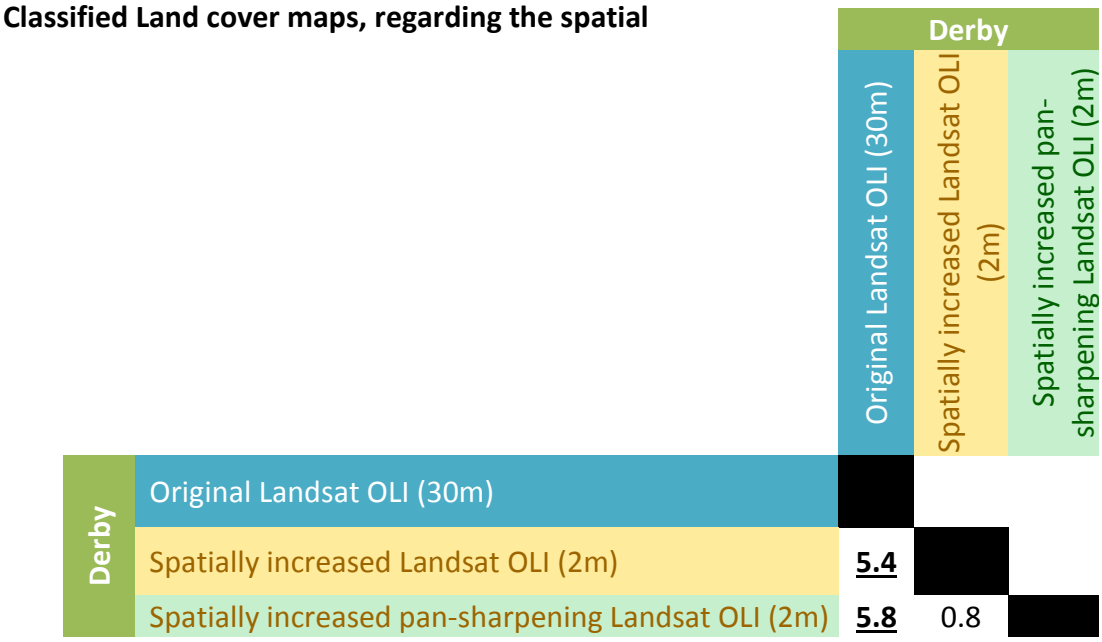


Figure 6.8. The McNemar statistical test on the capability of sparse representation method with regard to spatial transferability factor, to spatially enhance Landsat OLI for land cover mapping (values are statistically significant at 99% level of confidence shows bolded and underlined).

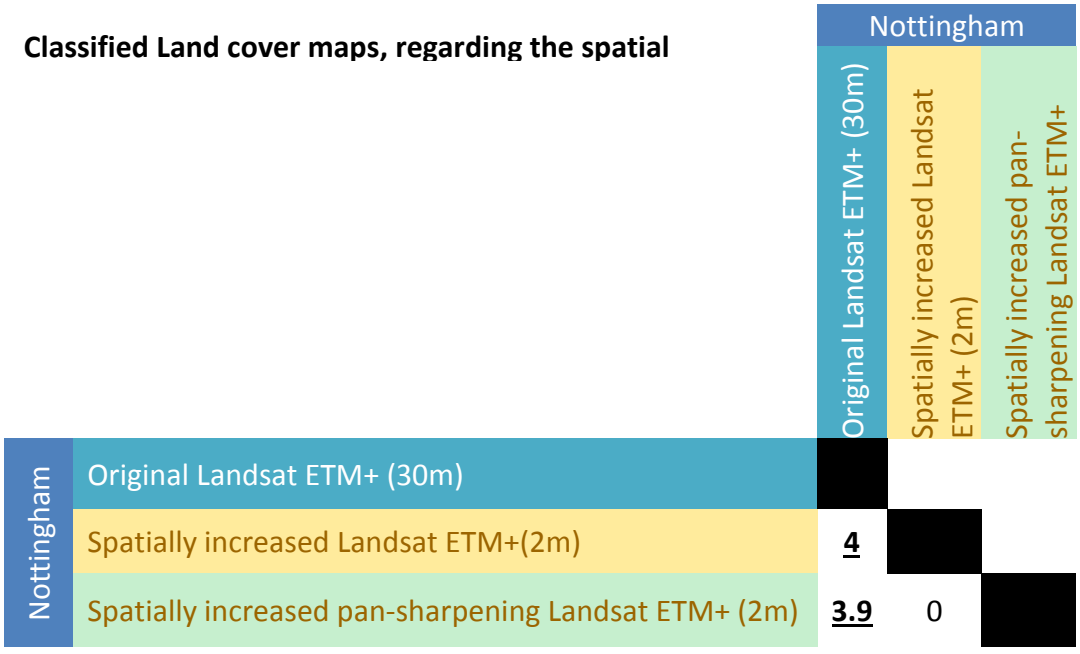


Figure 6.9. The McNemar statistical test on the capability of sparse representation method with regard to temporal transferability factor, to spatially enhance Landsat ETM+ for land cover mapping (values are statistically significant at 99% level of confidence shows bolded and underlined).

According to 6.8, the difference between overall accuracies of Landsat OLI and its two spatially increased versions are statistically significant. Similarly, as Figure 6.9 shows, the difference between overall accuracies of Landsat ETM+ and its two spatially increased versions are statistically significant.

Further, performing the sparse representation method on pan-sharpened Landsat OLI and Landsat ETM+ statically significant increased the overall accuracies of mapping Landsat OLI and Landsat ETM+ by 20% and 22% regarding the spatial and spectral separation, respectively.

Similar to the previous chapter, spatially increased pan-sharpened Landsat data produced a more accurate land cover map compared to spatially increased Landsat data, using sparse representation. For example, performing sparse representation on pan-sharpened Landsat OLI and Landsat ETM+ obtained 10% and 14%, respectively finer overall accuracy compared to performing sparse representation on Landsat OLI and Landsat ETM+. However, for both Derby and Nottingham, there was no statistically significant difference between the performing sparse representation on original Landsat data and pan-sharpened Landsat Data.

Therefore, the result shows the potential of the sparse representation to spatially enhance Landsat data for urban land cover mapping. For more detailed analysis, the accuracy of sparse representation will be examined at class level detail. The user and producer accuracies of six land cover maps represented per each class as four Figures: 6.10 (producer accuracy for Derby),

6.11 (user accuracy for Derby), 6.12 (producer accuracy for Nottingham), 6.13 (user accuracy for Nottingham).

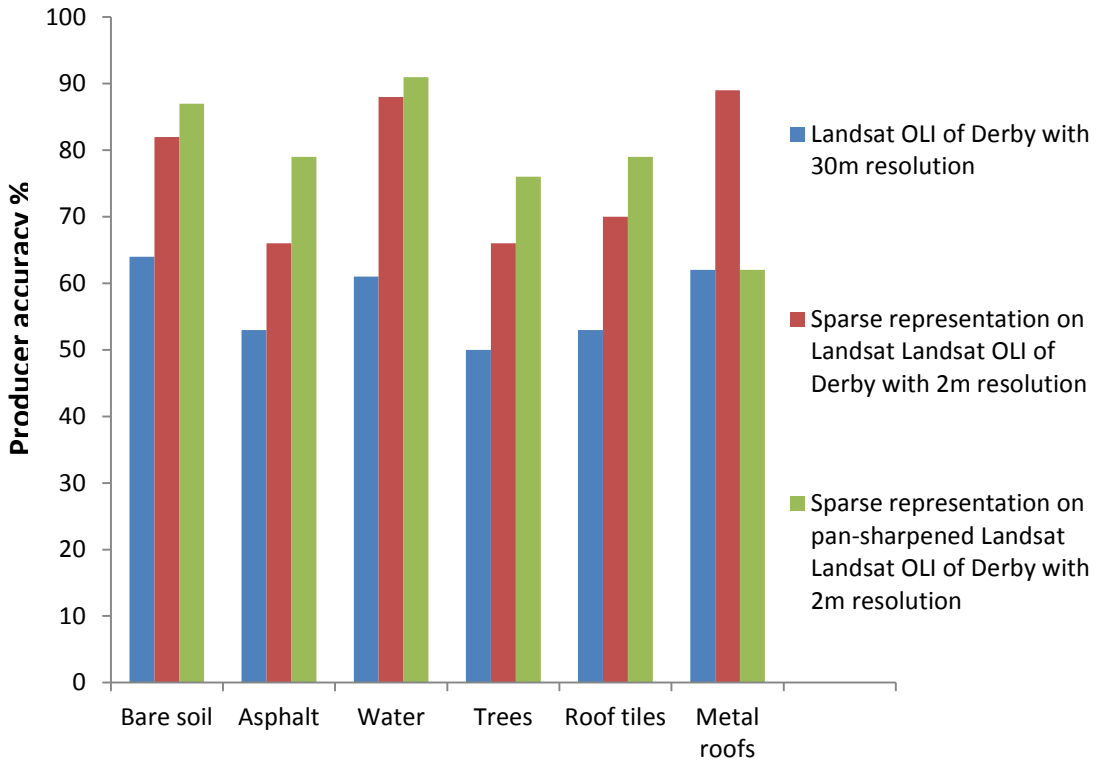


Figure 6.10. The producer accuracies of seven classes for Derby.

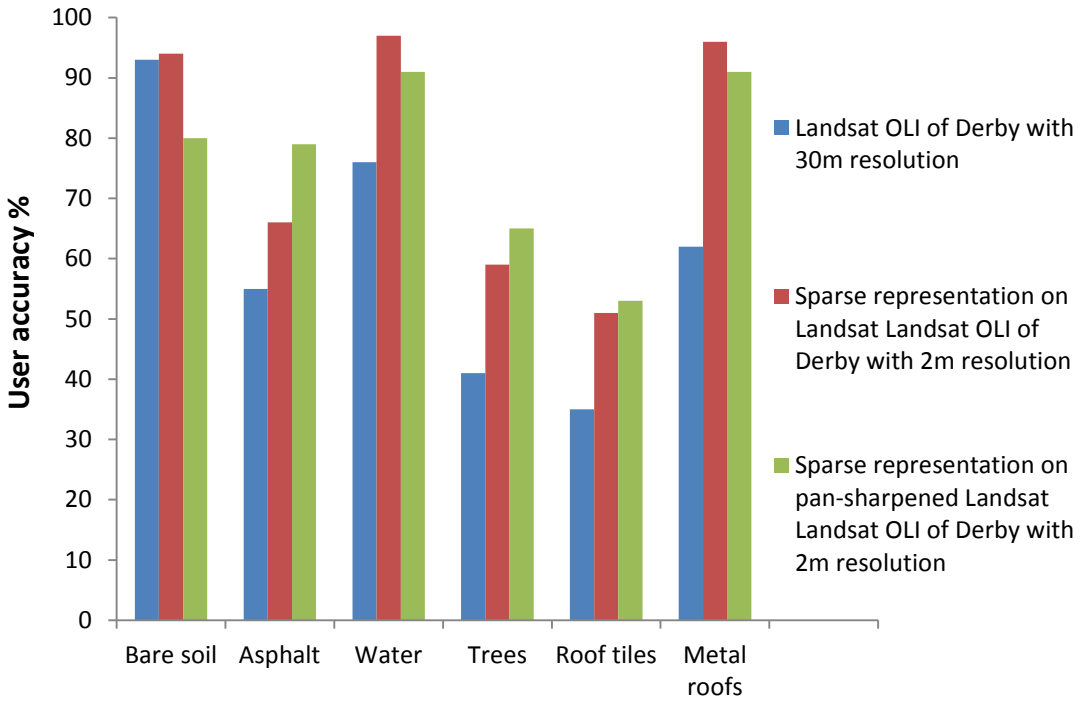


Figure 6.11. The user accuracies of seven classes for Derby.

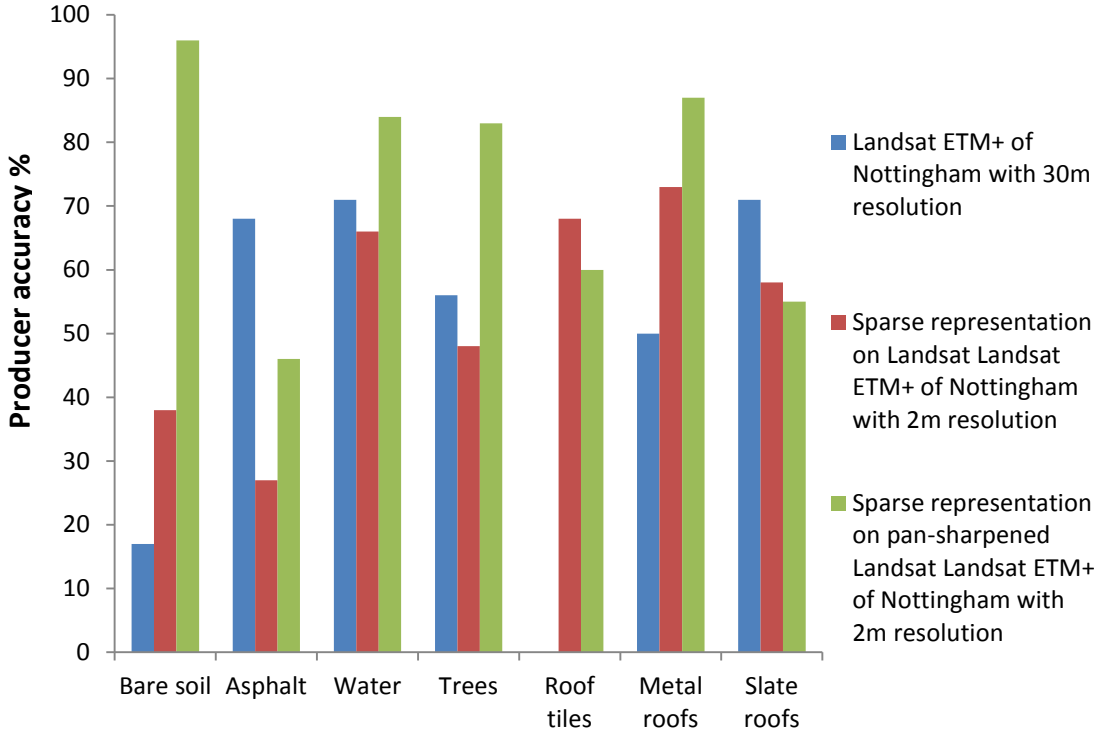


Figure 6.12. The producer accuracies of seven classes of Nottingham.

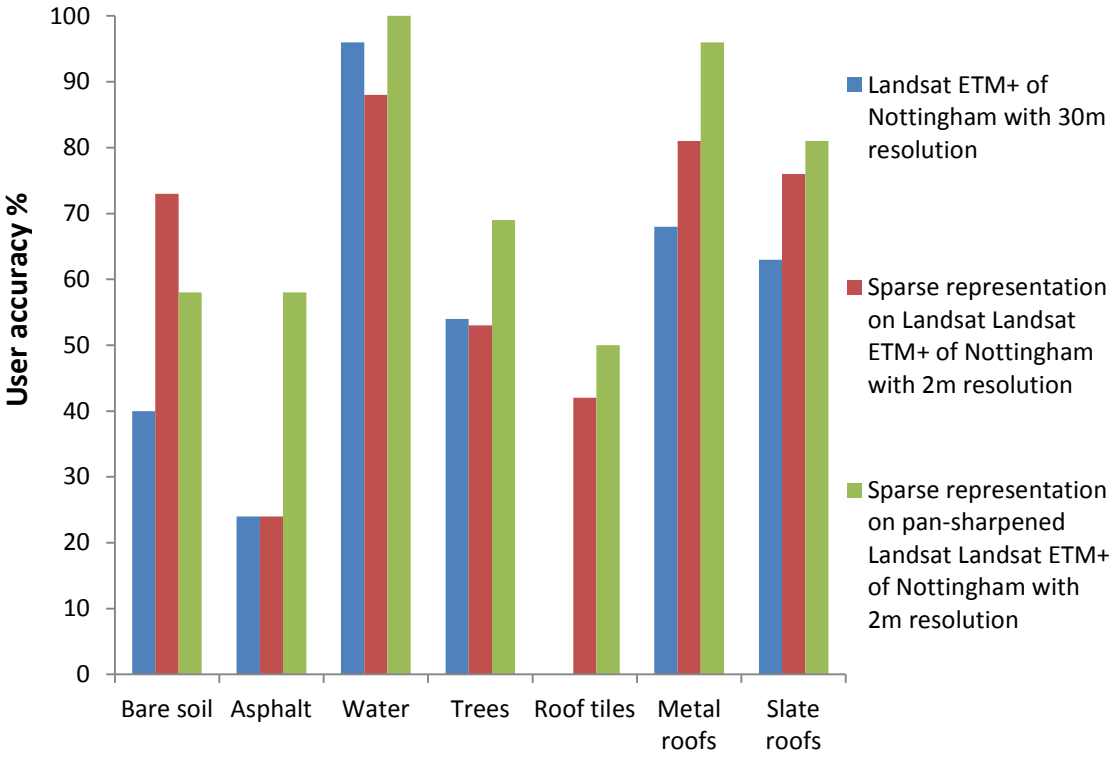


Figure 6.13. The user accuracies of seven classes of Nottingham.

6.7.1 Class level details

Since the main aim of this chapter is to assess the capability of the sparse representation method, the explanation of class details is represented by averaging both SVM classifier. However, the classifier advantage would be highlighted if there is a significant difference between classifiers' accuracy in an individual class.

6.7.1.1 Bare soil

Performing the sparse representation method improved the mapping accuracy of Landsat data within the bare soil class, regardless of spatial and temporal separation between WV-2 and Landsat data sets. In particular, spatially increased pan-sharpened Landsat data increased the producer accuracy of its corresponding Landsat OLI by 23% and Landsat ETM+ by 69% for Derby and Nottingham, respectively. These improvements were less significant for user accuracy. Similarly to the previous Chapter, the major confusion of bare soil class for both Derby and Nottingham occurred with the roof tile class. This can be explained by using soil materials to produce roof tiles (including clay and concrete roof tiles) class.

6.7.1.2 Asphalt

Performing sparse representation improved the producer and user mapping accuracy of Landsat OLI for the asphalt class, while the asphalt class was classified more accurately in Derby compared to Nottingham. For example, performing sparse representation on Landsat OLI of Derby and its pan-sharpened version increased the producer accuracy of asphalt class by 13% and

26%, respectively. However, performing sparse representation on Landsat ETM+ of Nottingham and its pan-sharpened version reduced the producer accuracy of asphalt class by 41% and 11%, respectively. This reduction in the producer accuracy for Landsat ETM+ can be attributed to the size and distribution of asphalt features in the Nottingham. Asphalt is mainly represented by roads, parking spaces, and industrial sites. The industrial structures are distributed mostly in the suburbs of Nottingham, while the major industrial structures are located in the city centre of Derby. Compared to Nottingham, asphalt features formed larger homogenous patches. This can address the reason why the producer accuracy of asphalt for Landsat OLI (30m spatial resolution) of Derby was 41% more accurate.

6.7.1.3 Water

Since the majority of the water class consisted of large features (i.e., ponds, lakes, and river), the classification accuracy of the water class was higher than other classes, using Landsat OLI and ETM+ images with a 30m resolution for both Derby and Nottingham. However, performing sparse representation on pan-sharpened Landsat data even increased the mapping accuracy of water by 13% and 30% for Nottingham and Derby, respectively. In fact, these improvements were not limited to producer accuracy. Performing the sparse representation of pan-sharpened Landsat data increased the user accuracy of water by 4% and 15% for Nottingham and Derby, respectively.

6.7.1.4 Trees

Sparse representation was successful to increase the mapping accuracy of trees. For example, performing the sparse representation of pan-sharpened Landsat data increased the mapping accuracy of the tree by 27% and 26% for Nottingham and Derby, respectively. These improvements were consistent too for user accuracy in which performing the sparse representation of pan-sharpened Landsat data increased the mapping accuracy of the tree by 15% and 16% for Nottingham and Derby, respectively. Similarly to the previous Chapter, significant confusion of tree class occurred toward both grass and roof tile classes. Readers can refer to Section 5.8.1.4 for possible reasons for this confusion.

6.7.1.5 Roof tile

The roof tiles class is a good example of the potential of sparse representation to enhance the spatial capability of Landsat data for mapping the urban areas. The producer accuracies of mapping Landsat ETM+ of Nottingham increased by 42% and 50% using sparse representation on Landsat ETM+ of Nottingham and its pan-sharpened version, respectively. These improvements were consistent by involving the spatial transferability factor in which performing the sparse representation on Landsat OLI and its pan-sharpened version increased the mapping accuracy of roof tiles by 17% and 26%. Besides producer accuracies, the user accuracies of classified maps for both Derby and Nottingham were increased by performing sparse representation.

Similarly to the previous Chapter, there were significant numbers of testing samples including bare soil, asphalt, trees, and grass, which were classified as

the roof tiles class in both Landsat OLI and Landsat ETM+ images. As a result of this, the area of the roof tiles class in the classified images is probably overestimated, which accounts for these poor user accuracies. Roof tiles class represents the residential areas in both cities. The most probable reason for this was the significant dense distribution of roof tiles class (orange colour) over the classified map.

6.7.1.6 Metal roof

Similarly to water features, the metal roofs class -which mainly represents the large industrial buildings- achieved an accuracy for Landsat OLI and Landsat ETM+. Performing sparse representation on Landsat data and their pan-sharpened versions increased the accuracy of metal roof class with 23% and 37% for Landsat OLI of Derby and 19% and 8% for Landsat ETM+ of Nottingham. Regarding reliability, performing sparse representation on Landsat data and their pan-sharpened versions increased the user accuracy of metal roof class with 34% and 29% for Landsat OLI of Derby and 13% and 28% for Landsat ETM+ of Nottingham.

6.7.1.7 Slate roof

Despite the rest of the seven classes which are common to Derby and Nottingham, slate roofs class was only considered for mapping the Nottingham case. While using the sparse representation method increased the user accuracy of mapping of Landsat ETM+ of Nottingham, using the sparse representation method was unsuccessful to increase the accuracy mapping of Landsat ETM+ of Nottingham. For example, though performing the sparse

representation on Landsat ETM+ and its pan-sharpened version increased the user accuracy (reliability) of Landsat ETM+ by 13% and 18%, the finest accuracy achieved 71% using original Landsat ETM+ with 30m spatial resolution.

Reviewing the Landsat ETM+ of Nottingham showed that in the most cases, buildings with slate roofs were normally constructed within a large site where the entire building covered by slate roofs. So, this creates a large homogenous patch of the slate roofs which is more convenient for Landsat ETM+ image to be identified. There was confusion between slate and both roof tiles and asphalt classes. These confusions can be addressed by the spectral similarity between the slate roofs and roof tiles classes and spatial correlation between slate roofs and asphalt classes.

6.7.1.8 Grass

Performing sparse representation on pan-sharpened Landsat data increased the accuracy of grass class with 16% and 13% for Landsat OLI of Derby and Landsat ETM+ of Nottingham, respectively. Regarding the reliability, performing sparse representation on pan-sharpened Landsat data increased the user accuracy of grass class with 10% and 15% for Landsat OLI of Derby and Landsat ETM+ of Nottingham, respectively.

6.8 Summary

In the previous Chapter, the capability of sparse representation method was demonstrated to enhance the spatial resolution of Landsat data for the urban land cover mapping application, when the training sample (WV-2 imagery) and Landsat OLI imagery were spatially and temporally coincident. This Chapter

assessed the capability of sparse representation method when there is a spatial or temporal separation between the training sample (WV-2 imagery) and Landsat data to demonstrate the practicality of the sparse representation for wider urban land cover mapping applications.

Regarding spatial transferability factor of sparse representation, a single WV-2 image of Nottingham was used as the training sample to spatially increase the Landsat OLI for land cover mapping of Derby. Regarding temporal transferability factor of sparse representation, a single WV-2 imagery of Nottingham was used as the training sample to spatially increase the Landsat ETM+ which was acquired Nottingham 12 years earlier for land cover mapping. The result of this Chapter showed that sparse representation method successfully spatially enhanced the mapping capability of Landsat data, regarding the spatial and temporal transferability factors. Using only a single VHR imagery to enhance the spatial resolution of Landsat data offers the wider user community to access the finer spatial resolution data sets for various urban mapping applications with significantly lower cost.

Chapter 7: Discussions and Conclusions

7.1 Overview

One of the main goals of remote sensing is to improve the accuracy of land cover mapping (Aplin *et al.,* 1999; Franklin *et al.,* 2002; Pal and Mather, 2003). From an urban perspective, increasing the accuracy of mapping urban areas offers a more realistic and reliable representation of urban geographical features and thus provides a better understanding of the status of the urban environment and the processes of change. Accurate mapping can help authorities and citizens to mitigate the undesirable urbanisation impacts on the environment at regional and global scales including worsening water quality, increasing flood volumes, exceeding air pollution, noise, and creating urban heat islands.

The research reported in this thesis aimed to enhance the land cover mapping of urban areas using free cost and accessible medium resolution remotely sensed imagery. To achieve this goal, the effects of varying image spatial and spectral properties on the detailed urban land cover map were firstly investigated, using OBIA/SVM and parametric and non-parametric pixel-based (ML and SVM, respectively) classification approaches (Chapter 3). The main message of the first part of this research was that using data sets with finer spatial resolution increased the accuracy of the final land cover map. However, due to the high cost and inherent challenges of VHR sensors (i.e., poor temporal resolution and short swath width), the widespread use of these data sets is

limited. Conversely, medium resolution sensors (i.e., Landsat sensor products) address the inherent challenges of VHR sensors and are available to the public free of charge. However, the main limitation of these medium spatial resolution sensors is their coarse spatial resolution. To address this issue, the second part of this research tested the potential of sparse representation to enhance the medium resolution imagery acquired by multispectral sensors, such as Landsat OLI, using WV-2 imagery with similar spatial coverage. Initially, the spatial resolution of the Landsat OLI for Nottingham was enhanced using a single WV-2 imagery of Nottingham. Then, the capability of sparse representation method was assessed to enhance the spatial resolution of the Landsat products (ETM+ and OLI), by consideration of spatial and temporal transferability. As a result, a single WV-2 image of Nottingham was used to enhance the spatial resolution of a Landsat OLI for Derby and a Landsat ETM+ of Nottingham which had been captured more than a decade earlier.

This chapter begins by discussing the result of the influence of the image's properties (i.e., spatial and spectral resolution) and the choice of non-parametric object-based (OBIA/NN) and parametric and non-parametric pixel-based (i.e., ML and SVM) classifiers. Then, the accuracy of the sparse representation method was studied to increase the spatial resolution of the Landsat OLI to map detailed urban areas. Finally, the capability of sparse representation is assessed for historical applications and applications which require spatial transferability to be performed on the Landsat products (i.e., ETM+ and OLI).

7.2 The influence of spatial resolution

The first part of this research evaluated the influence of spatial resolution, the spectral band set, and classification approaches for detailed mapping of urban land cover. A WV-2 image provided the basis for a set of 12 images with varying spatial and spectral characteristics, and these were classified using three different approaches (ML, SVM, and OBIA/SVM) to yield 36 output land cover maps.

It was expected that increasing the spatial resolution of imagery can significantly improve the accuracy of the final land cover map. The influence of increasing the spatial resolution on the overall classification accuracy can be addressed through two factors: i) increasing within-class variability and b) reducing boundary effect (Latty *et al*., 1985; Toll, 1985). The within-class variability has been considered by examining the effect of class covariance of pure pixels on the classification error (Hsieh *et al.,* 2001). As the spatial resolution becomes finer, the spectral covariance increases, which reduces the class spectral separability, leading to an increase in classification errors. The increase in classification errors is sharp if the classes have high within-class variability (Cushnie, 1987; Treitz *et al.,* 1992). Secondly, the boundary effect, which also referred as mixed pixels which often occur from the field boundaries between two or more classes within a pixel. The general problem of coarse resolution imagery is the mixed pixel problem over heterogeneous landscapes, where individual features are often smaller in size than individual pixels (Low and Duveiller, 2014; Sun *et al.,* 2016). Latty *et al.* (1985) suggested that

increasing the spatial resolution reduced the boundary effect among the smaller sized features, which potentially reduces the misclassification.

According to the results, increasing the spatial resolution from 30m to 10m to 4m and 2m, consistently enhanced the accuracy of classified maps with statistical significance at a 99% level of confidence. For example, the statistical differences between data sets with 30m versus 10m, 30m versus 4m, and 30m versus 2m spatial resolution increased by 3.3, 8.6, and 11.2 (these are the averaged values obtained via the McNemar tests when the influence of spectral resolution and classifier types were considered), respectively.

Figure 7.1 demonstrates the improvements of the user and producer accuracies for 11 classes, using finer spatial resolution. Regarding class level details, shadow obtained the most significant improvement by using finer spatial resolution, while the water class achieved the least significant improvement. This can be attributed to the size of features for both water and shadow. Since the size of features defined as water (i.e., pool and lake) are larger than the size of features that were defined as shadow (i.e., small shaded areas), it can be concluded that using the finer resolution enhanced the mapping accuracies of features with the smaller size, because data sets with finer spatial resolutions contain richer spatial information. This provides greater potential to extract more detailed thematic information (i.e., different land cover), cartographic features (i.e., buildings and roads), and metric information with *stereo-images* (i.e., height and area) (Weng, 2014).

7.3 The influence of spectral resolution

Another objective in the first part of this research was to evaluate the potential for a new generation of sensors, such as WV-2 and WV-3, offering new spectral bands (i.e., coastal blue, yellow, red edge, and NIR2) compared to the traditional four spectral bands of products such as IKONOS and QuickBird, to improve detailed mapping of the urban areas. As a result, three spectral resolution groups were compared including four conventional (blue, green, red, and NIR), six (adding red edge and NIR2 to four bands) and eight (full WV-2 range) spectral bands.

Since each spectral band is designated to detect the electromagnetic radiation within a particular region of the spectrum, using a wider range of spectral bands can identify a bigger range of different land cover types potentially leading to more accurate mapping accuracy. As a result, it is expected that having more spectral information could result in a higher accuracy of mapping urban areas.

According to the overall accuracy of 36 land cover maps (Figure 3.8), improving the spectral resolution in general terms can lead to modest increases in classification accuracy. Notably, the greater spectral capability of WV-2 can increase urban mapping accuracy compared to conventional four-band (i.e., IKONOS) imagery (Figure 3.5 in Section 5.1). It should be noted that the new bands provided by WV-2 and WV-3 (i.e., coastal, yellow, red edge and NIR2) were not necessarily developed with urban environments in mind. Instead, the main stated intentions were to enhance capabilities for bathymetry (coastal

blue) and vegetation (yellow, red edge, and NIR2) analysis (Marchisio *et al.,* 2010).

In fact, a consistent and direct trend has been identified by using a wider range of spectral band sets, and finer spatial resolution. In other words, the benefit of using more spectral bands becomes more evident in data sets with finer spatial resolution. For example, the overall accuracy improved by 1.2%, 1.4%, 2.4% and 3.9% (by averaging the differences between overall accuracy of data sets with four versus 6 and six versus eight spectral band sets) when spatial resolution increase from 30m to 10m to 4m to 2m, respectively. Overall there were no statistical significances at 99% level of confidence among classification accuracies of data sets with more spectral band sets, while the advantage of using greater spectral bands is statistically significant only at the finest spatial resolutions and especially when using OBIA/NN. According to Price (1997), the spectral response from ground features is strongly related to the sensors spatial resolution.

Figure 7.1 demonstrates the influence of using data sets with three different spectral bands set on the user and producer accuracies for eleven land cover classes. In this image 36 users and producer, accuracies were averaged based on spatial resolution and classifier types to emphasise the variation of a number of spectral bands. It is evident that the user and producer accuracies for the entire 11 classes stayed static or increased modestly as the number of spectral bands increased.

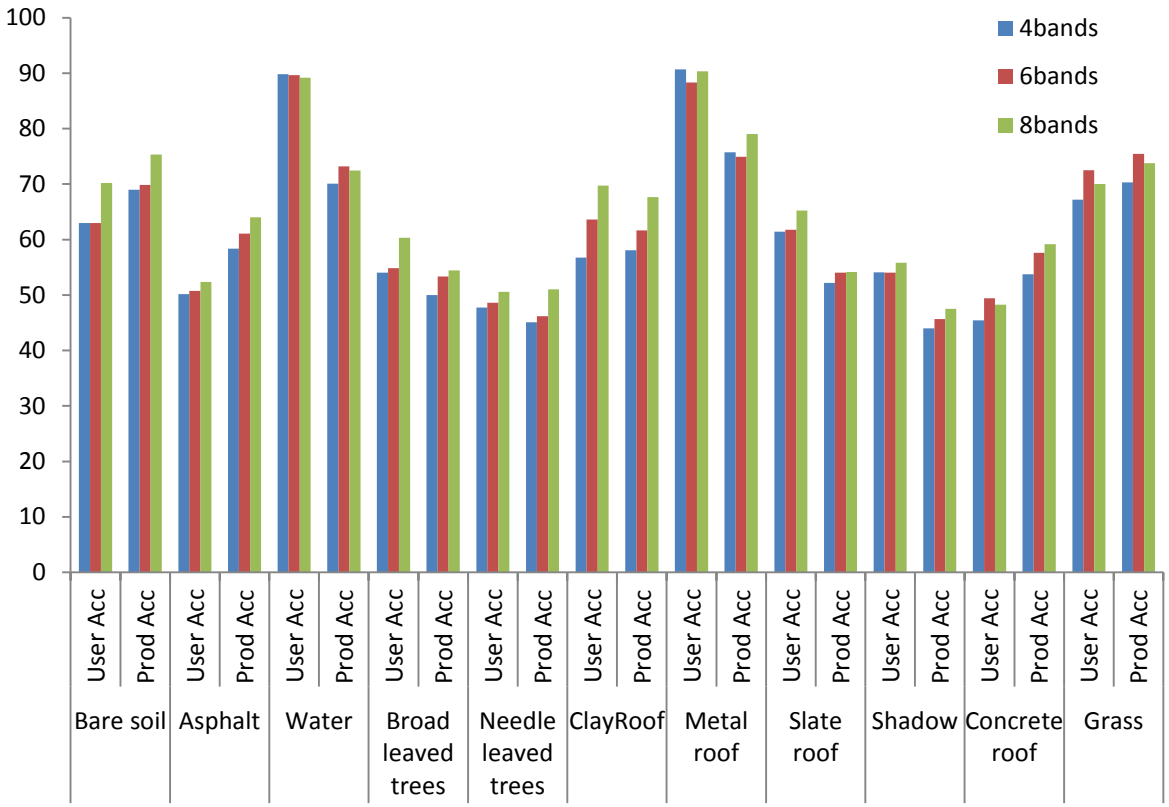


Figure 7.1. User and producer accuracies of 11 land cover classes highlighting the influence of spectral resolution (averaging the spatial resolution and classifier types).

Another possible point which may help to increase the accuracy associated with data sets with eight spectral bands is integrating three additional indices including Normalized Difference Vegetation Index (NDVI), Normalized Difference Bare Soil Index (NDBSI), and Normalized Difference Brick Roof Index (NDBRI). However, for data sets with four and six spectral bands, only NDVI was considered. The preliminary experiment demonstrates the capability of NDBSI and NDBRI indices to distinguish between urban impervious surfaces (including different roofs and road). According to Figure 7.2, the producer accuracies of those classes with impermeable property using eight spectral band sets (including NDBSI and NDBRI), performed marginally better than using six and four band sets (excluding NDBSI and NDBRI).

While designing this experiment, the effect of the Hughes phenomenon was noticed as an influential controlling factor on classification accuracy. This effect refers to the "curse of dimensionality" where adding spectral bands can, in fact, reduce classification accuracy, because more statistical demands are being made of inherently limited or sparse training data. Clearly, any such effect would counteract the intuitive expectation that added spectral detail should increase class separation. Overall, there was no noticeable Hughes effect. The Hughes effect is perhaps more of a concern with higher dimension data, i.e., hyperspectral, where it may be necessary to perform data reduction on a data set with 10s or 100s (referred as a number of spectral bands) (Bruzzone and Carlin, 2006). It seems WV2's eight-band data set is sufficiently small not to invoke any Hughes effect. This is useful and addresses why there is no need to consider data reduction at the outset of any project, at least from an accuracy perspective, although there may be other considerations, such as computer processing time.

7.4 The influence of the choice of classifier

Significant advances in classifier algorithms within recent decades, especially shifts from pixel to segment, opened up new interest towards enhanced urban mapping. The next objective of the first part of this research was to demonstrate the capabilities of three classifiers from ML (a conventional parametric pixel based classifier) to SVM (a non-parametric pixel-based classifier), and OBIA/NN on the accuracy of urban mapping. While the choice of the classifier is important in determining the success of classifying complex

urban environments, the optimum choice will vary depending on image data characteristics.

Referring to the results in Section 3.5.4, within the medium spatial scale (i.e., 10 meters and 30meters), ML performed slightly better followed by SVM and OBIA/NN, while these improvements were not statistically significant. In particular, ML algorithm classifies data sets with four spectral bands more accurately compared to data sets with six and eight spectral bands. Perhaps the advantage of ML over SVM and OBIA/NN can be addressed through the possible normal statistical distribution of the medium resolution data sets and a careful section of training data. This response would be more likely to address ML advantage because ML achieved the poorest accuracy for more complex data sets with six and eight spectral band sets at 4m and 2m spatial resolutions. Because data sets with 4m and 2m spatial resolutions potentially included noise and a high degree of spectral variability within and between classes, the data may not have been perfectly distributed normally, and as a result ML classification may have been impaired (Thomas *et al.,* 2003; Myint *et al.,* 2011). Another reason can be addressed through the data redundancy, in which additional spectral bands simply repeat existing spectral information. This repetitive additional spectral information may add complexity, leading to more confusion using ML, while more sophisticated classification methods like SVM and OBIA/NN may be able to deal with this complexity efficiently.

Focusing on medium spatial resolution (i.e., ranging from 30m to 10m), all classifiers obtained their poorest mapping accuracies, while both ML and SVM (both pixel-based classifiers) outperformed the OBIA/NN. This clearly can be

referred to the inadequate level of spatial information to distinguish between urban features. This can lead to significant misclassification, especially for OBIA/NN more than other both pixel-based approaches.

Compared to pixel-based classifiers such as ML and SVM, which rely on spectral differences for mapping, OBIA/NN benefits from the textural and contextual information in addition to spectral information. Perhaps, the influence of considering textural and contextual information on mapping accuracy becomes more apparent when imagery offers an adequate level of spatial and spectral information depending on the complexity of classification scheme.

According to results in Section 3.5.4, the superiority of OBIA/NN over both the parametric and non-parametric pixel-based classifiers becomes more evident on data sets using finest spatial and spectral resolutions. The highest accuracy of OBIA can be seen in the high quality extracted segments which are a more realistic and accurate representation of urban features. This only occurred when the most detailed spatial and spectral information was employed, which in this study was provided by the WV-2 full eight spectral bands of the VHR. Perhaps this finding reinforces the opinion that OBIA is particularly well suited for VHR imagery (Bruzzone and Carlin, 2006; Gholoobi and Kumar, 2015). At this fine scale of observation within-feature variation is likely, which may well lead to pixel-based misclassification, but may be mitigated by aggregation at the scale of the object.

The result highlighted that the accuracy of OBIA is more influenced by the number of spectral bands compared to pixel-based classifiers. OBIA/NN showed the most improvement in overall accuracy by involving additional

spectral bands compared to ML and SVM. For example, the overall accuracy of OBIA/NN improved by 9% and 8% using two and four additional spectral bands respectively. This improvement in mapping accuracy was not statistically different at 99% level of significance (Z value calculated as 2.9) among data sets with 4 and 8 spectral bands using SVM at 2m spatial resolution, while for similar 2m spatial resolution, there is statistical difference at 99% level of significance (Z value calculated as 4.5) using OBIA/NN. Perhaps, this can be explained through the quality of extracted segments. According to a preliminary study, using additional spectral bands improved the quality of extracted segments in OBIA and this improvement become more significant in data sets with 4m and 2m spatial resolution. Therefore, the results imply that contemporary VHR instruments with enhanced spectral capabilities have particular potential for urban mapping, holding a considerable advantage over early VHR sensors.

It should be noted that to assess the accuracy of pixel- and object-based classification accuracy, a point-based assessment procedure was adopted, since this enabled direct comparison between the different classification approaches. However, some practitioners have recently promoted the uptake of object-based assessment procedures, suggesting they may provide a more appropriate test of OBIA outputs (Zhan *et al.,* 2005; Radouxa *et al.,* 2011; Stehman and Wickham, 2011).

In this part of the research, broad-leaved trees, needle-leaved trees, and grass were three green space classes that were considered. Figure 7.2 shows the accuracy of pixel-based (i.e., ML and SVM) over the object-based (OBIA/NN) classification according to the user and producer accuracy of averaged three

vegetated classes (summarising the influence of spectral resolution). While both the pixel and object classifiers obtained similar user accuracy, the object-based classifier obtained more accurate producer accuracy compared to pixel-based classifiers (i.e., ML and SVM). This suggests an advantage of using object-based over pixel-based classifiers for mapping vegetated areas. Perhaps this can be addressed by spectral similarity among vegetated classes, and this becomes challenging when pixel-based classifiers merely rely on spectral information, while OBIA includes textural information in addition to spectral information to distinguish between classes with similar spectral diversity.

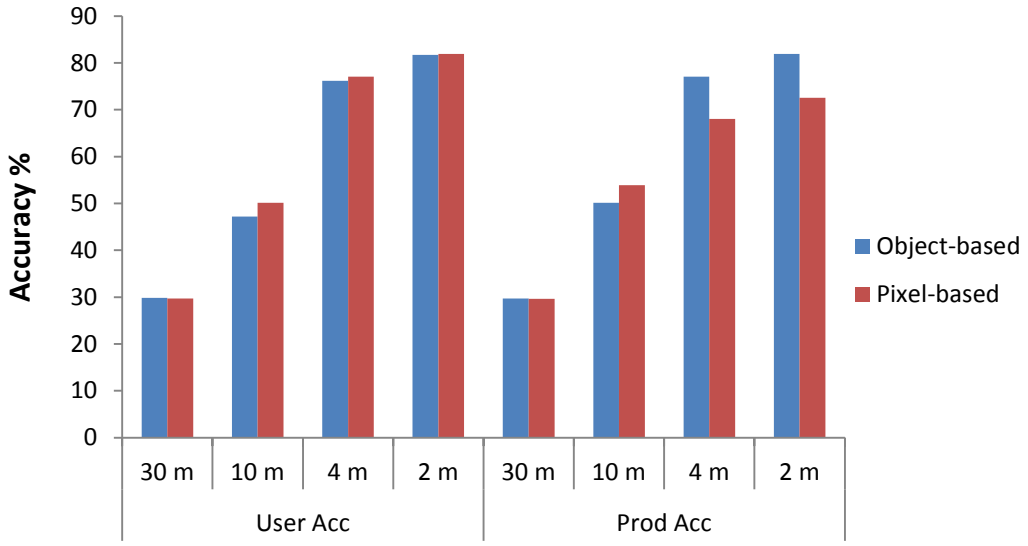


Figure 7.2. The comparison between the pixel-based and object-based accuracies on mapping vegetated areas by summarising the influence of spectral resolution.

To optimise OBIA accuracy, it is important to find the most optimised segments quality which most likely represents real ground features. Regarding the implementation procedure, SVM and ML are more straightforward than OBIA/NN. Because finding the most optimum parameters for segmentation are acquired through trial-and-error and segmentation assessment step. From a

wider perspective, it can be seen that each of these three classifiers had the most accurate accuracy when the finest resolution data was available at the time. For example, ML as earlier classification algorithm performed more successfully with data sets with medium resolution, while SVM was more successful data sets with finer resolution. Finally, OBIA/NN as most recently classification approach achieved the best accuracy when the most spatially and spectrally data set (WV-2 with 2m resolution and eight spectral band sets) were used.

7.5 Project requirements versus project resources

This research presents various data and analysis considerations for urban mapping projects. Other essential consideration relates to project resourcing since this will have a bearing on both the imagery acquired and the methodology employed. The results here show that VHR imagery is essential for accurate, detailed thematic mapping of urban land cover. Unfortunately, this imagery can be costly, unlike the free provision of all Landsat data. Moreover, the more advanced image products (i.e., 2m, eight bands WV-3 imagery) tends to be considerably more expensive than basic (i.e., 4m, two bands) products. Also regarding software and operator resources, OBIA approaches tend to require more resource than pixel-based classification approaches. For instance, OBIA involves considerably more operator input than ML or SVM analysis, some OBIA operations require substantial computer processing resources, and OBIA packages can be relatively costly. This experiment found OBIA classification of 2m spatial resolution, eight spectral

bands imagery most accurate, though this combination is perhaps the most expensive regarding resourcing. Satisfactory, cheaper alternatives may exist, depending on user requirements. Here, for instance, SVM classification of 2m, four band data resulted in only a relatively modest decrease in accuracy against the maximum, and this approach would incur considerable savings regarding data costs (4 bands WV-2 or WV-3 imagery), software requirements (no OBIA purchase) and workforce (reduced operator time).

7.6 Sparse representation and Landsat data

The overall accuracy of performing sparse representation on original Landsat data, and their corresponding pan-sharpened versions, plus original WV-2 imagery were demonstrated in both Figures 5.9 and 6.7. However, Figure 7.3 demonstrates on the overall accuracies by performing the sparse representation on Landsat data and their corresponding pan-sharpening versions with regards to three scenarios.

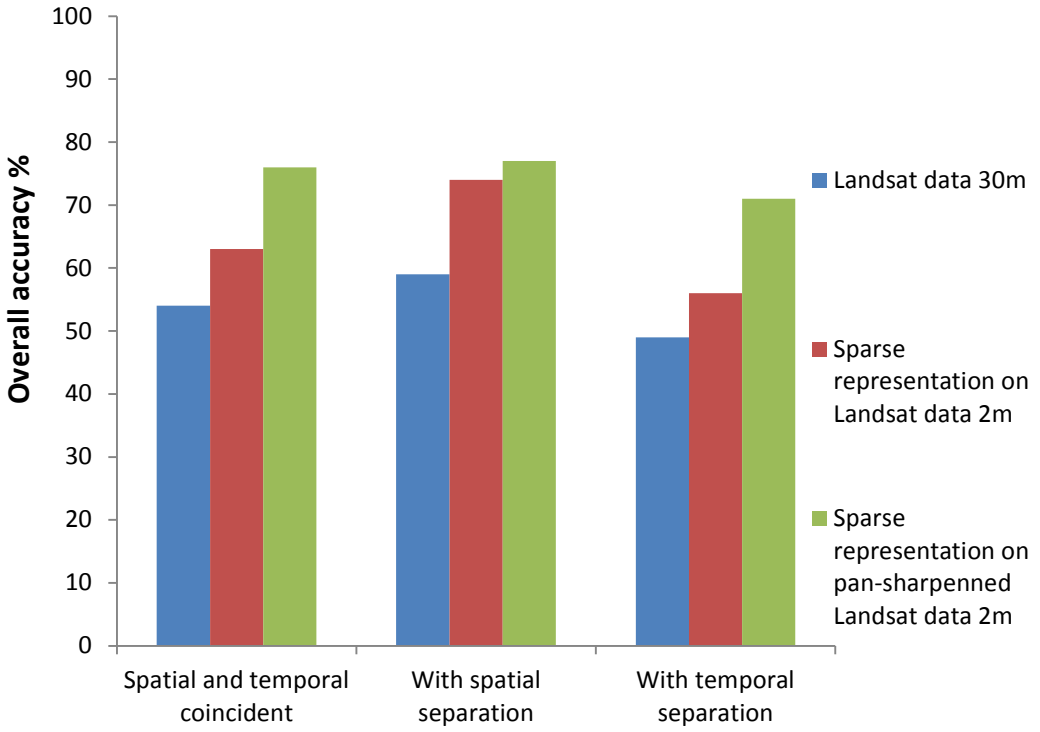


Figure7.3. the overall accuracies of Landsat data and its spatially enhanced versions (i.e., using SVM classier), representing through three scenarios.

As per Figure 7.3, performing sparse representation on Landsat data enhanced their mapping accuracy by 9%, 15%, and 7% in the first, second, and third scenarios respectively (Note: only SVM classifier). However, these mapping accuracies were improved more significantly by 22%, 18%, and 22% in the first, second, and third scenarios respectively (Note: only SVM classifier) when sparse representation was performed on pan-sharpened Landsat data. Referring to Tables 5.12 and 6.8, these mapping accuracy enhancements were statistically significant at a 99% level of confidence, whether performing the sparse representation of original or pen-sharpened Landsat data.

7.7 Sparse representation and pan-sharpened Landsat data

As Figure 7.3 shows, performing sparse representation on pan-sharpened Landsat was more successful than performing directly on original Landsat data for all three scenarios. Results in Chapters 5 and 6 showed that performing sparse representation on pan-sharpened Landsat data outperformed carrying out sparse representation on original Landsat data by 13%, 3%, and 15% for first, second, and third scenarios respectively. According to Figure 5.12, there is a statistically significant difference between performing sparse representation on Landsat data and its pan-sharpened version in the first scenario. However, the result of the McNemar test (Figure 6.8) revealed that there is no statistically significant difference between performing sparse representation on Landsat data and their pan-sharpened versions, by consideration of spatial or temporal separation between training image and Landsat data. This statistical significance only occurred when the training image was spatially and temporally coincident with Landsat data. This result suggests that while performing sparse representation on the pan-sharpened version is more likely to outperform its original data set, in such applications where there is a spatial and temporal separation between training and the input image, it is more likely that these advantages occur without significant statistical differences.

From a visual perspective, spatially enhanced land cover maps using sparse representation method offers greater thematic information compared to their original data sets (Figure 7.4).

The success of spatially enhanced both pan-sharpened Landsat data and the original Landsat data, using the sparse representation method depends on the level of spatial information provided by both pan-sharpened/original Landsat images. In this case, the spatial information, which is offered by pan-sharpened Landsat data, is more detailed than original Landsat data (Figure 5.7). For example, while each pan-sharpened Landsat data's pixel represents an area of 15m x 15m ($225\mathrm{m}^2$) on the ground, each pixel of Landsat data cover an area of $900\mathrm{m}^2$ (300m x 300m) on ground surface. Furthermore, the magnification degree is an important factor affecting the performance of the sparse representation methods. In fact, increasing the spatial resolution of Landsat data directly from 30m to 2m means to estimate the spatial information of 225 pixels with a 2m resolution from a 30m resolution pixel of Landsat imagery. However, increasing the spatial resolution of pan-sharpened Landsat data from 15m to 2m means to estimate the lost spatial information of 56 pixels with a 2m resolution from a 15m resolution pixel of pan-sharpened Landsat OLI. Because, the estimation of the spatial information of 225 pixels with 2m resolution from a single 30m resolution of Landsat imagery is a more challenging task compared to estimating the spatial information of 56 pixels with 2m resolution from a single 15m resolution of pan-sharpened Landsat imagery. It can be suggested that performing the sparse representation method on pan-sharpened Landsat imagery is more likely to obtain more

accurate results compared with performing sparse representation method on original Landsat data (Figure 7.4).

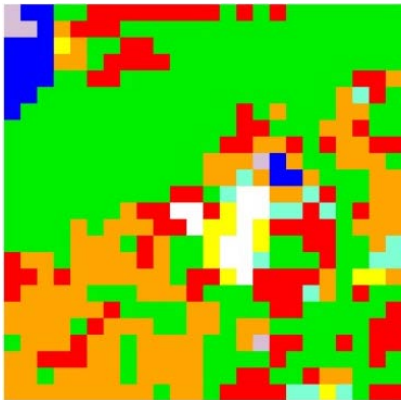


Figure 7.4.a, Landsat OLI 30m

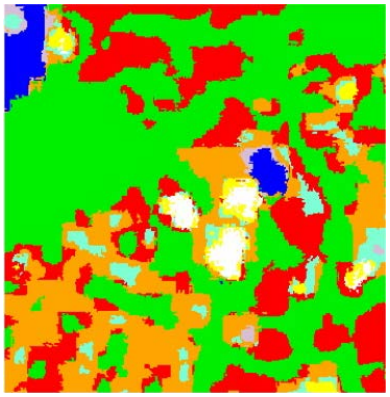


Figure 7.4.b, enhanced Landsat OLI 2m



Figure 7.4.c enhanced pan-sharpened Landsat OLI 2m



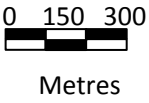




Figure 7.4. 3 small subsets with 750m × 750m size of land cover maps representing when there is spatial and temporal coincident between WV-2 and Landsat OLI (a); Landsat OLI with 30m, (b); enhanced pan-sharpened Landsat OLI using sparse representation method with 2m resolution, (c); actual WV-2 imagery with 2m resolution.

7.8 Classification scheme and the overall mapping accuracy

While performing sparse representation enhanced the mapping accuracy of Landsat data for three scenarios whether on original Landsat data or their

corresponding pan-sharpened versions, the mapping accuracy of sparse representation on the second scenario obtained the highest accuracy compared to their corresponding results of first and third scenarios. For example, the classified overall accuracy of Landsat data spatially enhanced through performing the sparse representation in the second scenario obtained 59% and 74%, while these figures were 54% and 63% in the first scenario and 49% and 56% in the third scenario, respectively.

This advantage of mapping Derby can be attributed to the different classification schemes used in the second scenario compared to first and third scenarios. The overall accuracies of the Landsat imagery and its two spatially enhanced versions achieved higher for Derby compared than Nottingham (Figure 6.7 on P.196). Since these superiorities are valid not only for spatially enhanced Landsat images, the overall accuracy of the classified land cover of the original Landsat imagery in Derby outperformed the classified land cover in Nottingham. This higher overall accuracy can be attributed through the different classification schemes which used in Nottingham and Derby. In fact, while the classification scheme was used for Nottingham involved eight-members, slate roof class was merged into other roof tiles class forming classification scheme with seven members for Derby. Since, there was noticeable misclassification between the slate roof and other roof tiles, merging these two classes helped to improve the overall accuracy of classification for Derby. For example, the producer accuracies of other roof class in the third scenario (i.e., classification scheme with 8 class members) obtained 0%, 68%, and 60% for Landsat data, spatially enhanced Landsat data

through sparse representation, and spatially enhanced pan-sharpened Landsat data through sparse representation, respectively. However, these figures improved to 53%, 70%, and 79% in the second scenario (i.e., classification scheme with 7 class members) for their three corresponding data sets.

According to Stehman (2009), increasing the detail of classification schemes can lead to a reduction in the classification accuracy. For example, as a part of the US National Land Cover Data set, Vogelmann *et al.* (1998) reported a 21% decrease in the overall mapping accuracy when the classification scheme changed from the very general Anderson level I to the more detailed Anderson level II.

7.9 Classifier choice

The influence of classifier type has been explained in more detail in Section 7.4, while this second part focuses on the compatibility of both non-parametric pixel and object-based classifiers (i.e., SVM and OBIA/SVM) to classify the spatially enhanced data set using sparse representation method. Although the accuracy of using SVM and OBIA/SVM were similar in both enhanced Landsat OLI through the pan-sharpening and sparse representation methods, SVM with 76% overall accuracy outperformed OBIA/SVM with 70% overall on classifying the enhanced Landsat OLI by performing sparse representation on pan-sharpened and original Landsat OLI.

Careful visual assessment on segmented spatially enhanced Landsat data revealed that using the sparse representation method allocates the values of the pixels (i.e., digital number) in the way that the pixel's values gradually

change from the centre of a homogeneous feature. As result, instead of representing an actual homogeneous feature as one object, it is more likely that an actual homogeneous feature represents with several smaller objects (i.e., over-segmentation). Furthermore, referring to the outcome of the first part of this these, OBIA/NN has the highest accuracy when the most spatially and spectrally data set was used (i.e., 2m resolution and full eight spectral bands) were used, while here the number of spectral bands was limited to four and the level of spatial information was not the same as the actual WV-2 imagery.

7.10 Training image characteristics

This section discusses the importance of the spatial properties of the training image (i.e., spatial resolution and spatial coincident) referring to input coarse resolution on the of sparse representation accuracy.

The influence of varying spatial resolution of the training image on sparse representation's accuracy has not been studied previously. Because Yang *et al,* (2010) assumed that the sparsity vector is similar to an H-resolution image and its degraded version; they claimed that the effect of using finer spatial resolution imagery for training could be marginalised, which means images with coarser resolution can be considered to be used as the training image. As a preliminary study, three small subsets of WV-2 imagery degraded to 15m and 30m spatial resolutions and were considered as training images to spatially enhanced Landsat OLI imagery. The initial visual assessment showed that using those three different training images successfully increased the spatial

information of Landsat OLI, while the level of spatial information became greater when a finer resolution training image was used. The training image with 2m spatial resolution offers sharper and clearer edges followed by training images with 15m and 30m spatial resolution. Perhaps, this result suggests that the level of spatial resolution of the training image is important for sparse representation in such detailed urban applications.

Next, as a result of dictionary training, both L\H-resolution dictionaries represent the structure of the training images. Since the training image is the only source used to train both L\H-resolution dictionaries, the statistical similarity (i.e., texture variations, land cover types, and structural patterns) between training and input coarse resolution images has a significant role in the success of sparse representation (Yang *et al.,* 2010). It would be useful to review a summarised concept of the sparse representation and the roles of L\H-resolution dictionaries (Note: refer to Section 4.8 for more detailed explanation). This explanation helps to understand the importance of statistical similarity between training and input coarse resolution image on the accuracy of the sparse representation method.

Referring to the concept of this method, both L\H-resolution dictionaries are generated within the training phase, and they are used to compute other key parts, including the sparsity vector and estimating the lost spatial information of the input coarse resolution image (i.e., Landsat data).The sparsity vector of patch k (s_k) can be determined through the L-resolution dictionary (D_l) and its corresponding patch k from input coarse resolution image (P_k^l) (formula number (12): $\|P_k^l - D_l\, s_k\|_2 \leq \varepsilon$). Next, the same computed sparsity vector for

patch k (s_k) is multiplied into the high-resolution dictionary (D_h) to estimate the lost spatial information of patch k (P_k^h) (assumption 4: $P_k^h = D_h\ s_k$). Yang *et al.* (2010) emphasised that formula number 12, and assumption 4 are validated when the isometry condition occurs between the training and input coarse resolution images.

According to Figures 5.7 and 6.4, the spatially enhanced Landsat data offered a greater level of spatial information when there is a spatial coincidence between training and input coarse resolution images. This highlighted the essentiality of the statistical similarity between the training and input coarse images. It could be, concluded that the training image represents similar statistical nature (i.e., texture variations, land cover types, and structural patterns) of the input coarse resolution image to obtain finer sparse representation accuracy. For example, using a training image with mostly coverage of vegetation to increase the spatial resolution imagery of imagery covering a dense urban landscape would be unlikely to satisfy the isometry condition which is required to validate the sparse representation assumptions, and therefore leading to the unsatisfactory result.

Next, it is an important to consider phenology when selecting the training imagery. The urban landscape contains vegetated areas, which are in their highest abundance in the summer (in general, in the temperate environment of the UK). It is important to consider the element of phenology between training and input coarse resolution images to optimise the statistical similarity between these two data sets.

7.11 Zoom factor

The zoom factor is another parameter with significant influence on the sparse representing accuracy. This factor acts as the spatial resolution magnification degree of input coarse resolution Imagery within the reconstruction phase. Conversely to previous studies (Song *et al.,* 2005; Yang *et al.,* 2010), in which used the maximum zoom factor was limited to 3; this research assessed the capability of sparse representation by using two zoom factors 7.5 and 15. According to results, the accuracy of sparse representation on pan-sharpened Landsat data with zoom factor 7.5 ($\frac{15m}{2m}$) outperformed the accuracy of sparse representation on original Landsat data with zoom factor 15 ($\frac{30m}{2m}$). This result suggests that the accuracy of the sparse representation method is more likely to degrade rapidly using a greater zoom factor.

Compared to increasing the spatial resolution of Landsat data from 30 to 2m (zoom factor =15), because of allocating lower zoom factor (i.e., 5); it is more likely that increasing the spatial resolution of more recent medium resolution data sets (i.e., Sentinel-2) with finer spatial resolution (10m) offers greater level of spatial information.

During this study, it has been discovered that allocating the same zoom factor for both training and reconstruction phases is more likely to produce a more accurate result. In fact, this condition helps the trained L-resolution dictionary match the spatial resolution of input coarse imagery. This is particularly important because the sparse vector is computed through L-resolution dictionary and input coarse resolution image (i.e., Landsat data) and this sparse

vector will be used to estimate the lost spatial information of input coarse resolution image.

7.12 Sparse representation algorithm

The sparse representation algorithm which is used in this thesis was locally based by performing it on patches, rather than the entire image. Compared with other learning-based SRR methods, this sparse representation method reduced the size of the training database leading to quicker processing time (Sun *et al.*, 2003). Yang *et al.* (2008) claimed that this size and dimensionality of training database, which is extracted as raw patches from training samples, is sufficient for the purpose of sparse representation accuracy.

While conventional dictionary training procedures generate low and H-resolution dictionaries in two separate steps (i.e., K-SVD algorithm (Aharon *et al.,* 2006), this sparse representation method employed a recently developed joint dictionary training (Pham *et al.,* 2008; Dong *et al.,* 2011), therefore training both low and H-resolution dictionaries together in a single processing step. Thus, joint learning dictionary methods require significantly less database space compared to K-SVD training dictionary, resulting in a quicker processing time.

Elad and Aharon (2006) claim that this sparse representation is more robust to the noise of the training image, and, due to linear programming, the computation process of this sparse representation is reasonably efficient and scalable (Yang *et al.,* 2008). Furthermore, unlike most of the other learning-based SRR methods, this sparse representation method can perform both

denoising and spatial enhancement tasks simultaneously and generates both realistic textures and sharp edges from the single input image (Yang *et al.,* 2010). However, this joint dictionary training method does not guarantee that the sparse representation of input coarse resolution image patch can be well reconstructed its underlying HR image patch. In response to this, Cheng (2015) reported that sparse representation methods remain extremely ill-posed because there are infinitely a significant number of H-resolution images which satisfy the reconstruction constraint assumption for input coarse resolution image (i.e., Landsat data).

7.13 Applications

7.13.1 Urban mapping

The dynamic development of urban areas requires frequent updating of existing databases. Heiden *et al.* (2007), reported the need for frequent mapping of the urban areas. Relying on VHR imagery for frequent mapping of urban areas is a costly task. As an alternative path, the sparse representation method can be adopted to increase imagery's spatial resolution. The images spatial characteristics can be defined according to the complexity of the classification scheme. According to the result of this thesis, Landsat data can be spatially enhanced using a WV-2 imagery to map generalised urban mapping applications. However, for urban application with a more complex classification scheme, perhaps SPOT with the panchromatic band would be a better option.

7.13.2 Historical applications

The thesis demonstrates the capability of sparse representation for historical applications. However, performing sparse representation on pan sharpened Landsat ETM+, enhanced mapping accuracy of Landsat ETM+ significantly (by 22%) compared to performing sparse representation on pan sharpened Landsat ETM+ (by 8%). However, the availability of the panchromatic band of Landsat data was limited to the ETM+ and OLI sensors, which have only been available since 2000. The practicality of the sparse representation for older Landsat data generations (i.e., Landsat MSS and TM) can be more significant for applications with more generalised classification scheme.

Additionally, the statistical similarity between training and coarse resolution image was highlighted (in training image characteristics Section) as an essential factor to the success of sparse representation method on mapping complex urban areas. Regarding historical applications, it would be recommended to consider the spatial coincident factor to optimising the sparse representation accuracy. For example, for spatially enhanced Landsat data of Sheffield, it would be better if a VHR image of Sheffield was used as the training image.

Conversely, one of the challenges of using sparse representation method for historical applications would be accessing the reference data for the validation procedure. The possible resources for collecting historical reference data are typically limited to historical maps and images which may not offer an adequate level of detail for selecting reference data.

7.13.3 Disaster management

The validity of sparse representation method can be extended to wider applications. For example, looking towards a Sentinel era where medium resolution imagery, with a spatial resolution three times finer than Landsat data (10m resolution in RGB and NIR spectral bands) are becoming available freely (i.e., Sentinel-2 and more than five-year-old SPOT). Furthermore, Sentinel-2 offers shorter revisit (5 days, but with different viewing angles) and wider coverage (290km swath width), compared to Landsat data with a 16-day revisit and 165km swath width. Because the sparse representation procedure involves two separate phases, the trained dictionaries could be reserved for future tasks. These available training dictionaries and the Sentinal-2 short revisit time (5 days) and spatial resolution (10m) can be used for disaster management applications. For example, in the case of disaster occurs the reserved trained dictionaries can be used to increase the spatial resolution of Sentinel-2 to map buildings damage.

7.14 Summary

Overall, sparse representation was successful in enhancing the mapping accuracy for Landsat data. This study demonstrated the potential of the sparse representation method for historical applications and applications which require spatial transferability. The accuracy of sparse representation would be finer if the input coarse resolution images offered a panchromatic band, which could then be utilised. This single band with finer spatial resolution ensures the possibility of performing sparse representation on the pan-sharpened version

of data sets rather than its original version. This mapping accuracy can be identified as improving the producer accuracies of each land cover class as demonstrated in Figures 7.5, 7.6, and 7.7, for first, second, and third scenarios, respectively. As discussed in Section 5.8, the overall mapping accuracy of spatially enhanced Landsat OLI imagery using sparse representation and pan-sharpening methods obtained 13% and 22% less accuracy when compared to WV-2 imagery. In fact, there is a clear visual difference between the level of spatial information produced by performing sparse representation on pan-sharpened Landsat OLI and level of spatial information offered by WV-2 imagery (refer to Figure 7.4). However, regarding the free cost of Landsat data compared to the significant cost of WV-2 imagery, the enhanced level of spatial information of Landsat data using sparse representation method, may satisfy several applications including urban mapping. It should be highlighted that the accuracy of sparse representation can be even more significant if using free remotely sensed multispectral products with finer spatial resolution (i.e., Sentinel-2 and SPOT with 10m) or applications which required more generalised classification system (compared to classification scheme used with eight class members).

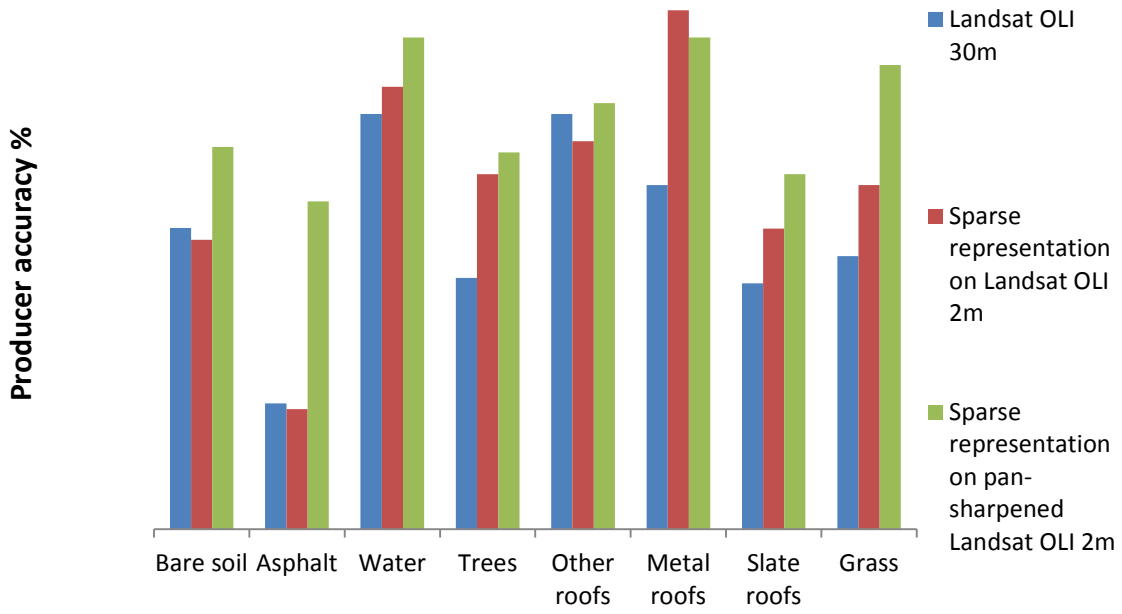


Figure7.5. The producer accuracies of Landsat OLI of Nottingham and theirs's spatially enhanced versions when there are spatial and temporal coincident between Landsat OLI and WV-2 (the first scenario).

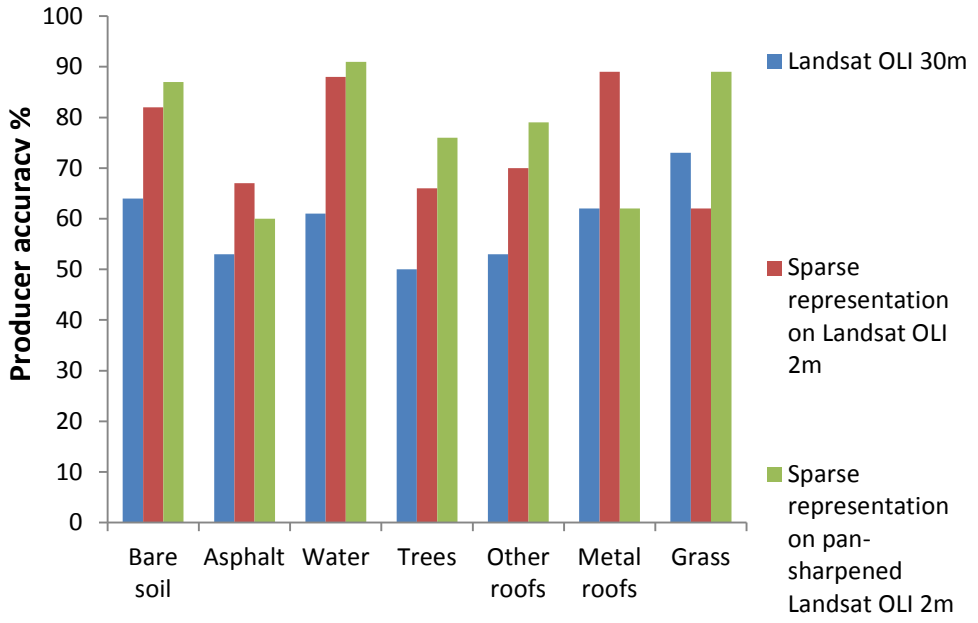


Figure7.6. The producer accuracies of Landsat OLI of Derby and theirs's spatially enhanced versions when there is a spatial separation between Landsat OLI and WV-2 (the second scenario).

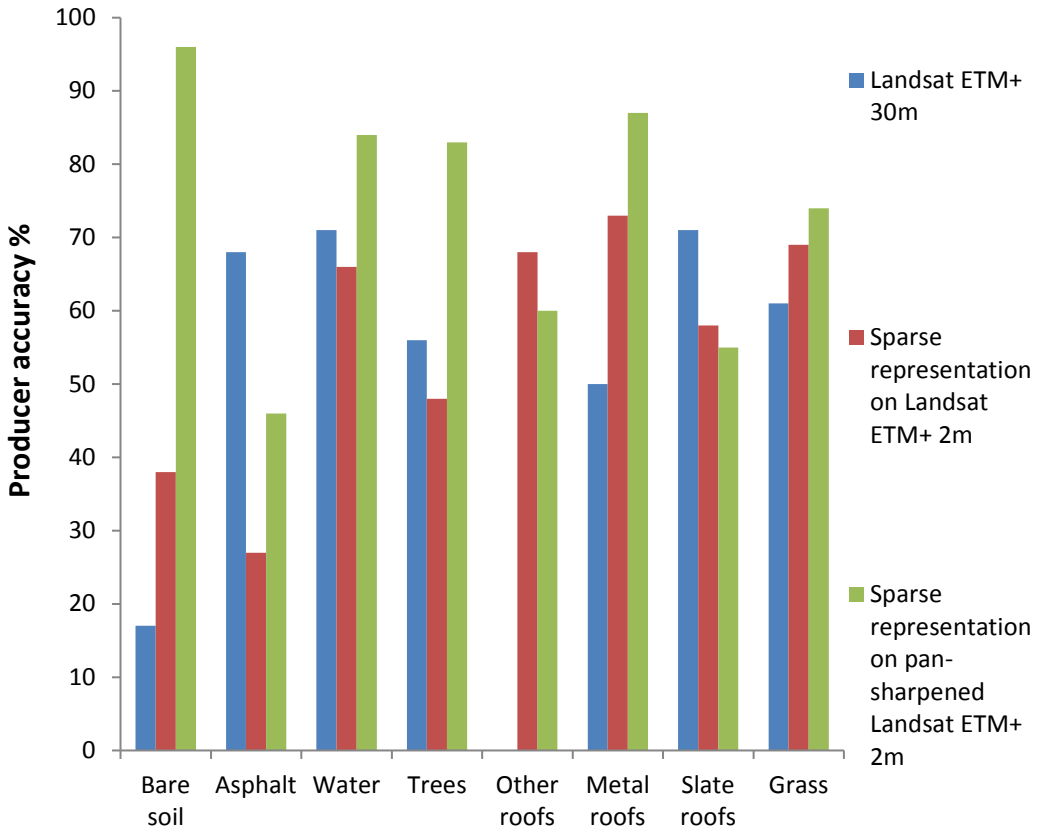


Figure7.7. The producer accuracies of Landsat ETM+ of Nottingham and theirs's spatially enhanced versions when there is a temporal separation between Landsat ETM+ and WV-2 (the third scenario).

7.15 Conclusions

In conclusion, this thesis has

- Reaffirmed the importance of the spatial resolution of remotely sensed data for urban land cover mapping.
- Demonstrated the advantages of non-parametric classifiers (i.e., NN and SVM) compared to parametric classifiers (i.e., ML) for urban land cover classification.
- Highlighted the influence of spectral improvements of the recent VHR sensor over to conventional multi-spectral seasons with four spectral bands sets.
- Illustrated the effectiveness of sparse representation as a method to enhance the spatial resolution of Landsat data for urban land cover mapping.

- Demonstrated that the sparse representation method can be transferred over space and time, highlighting the influence of the parameters around its use.

7.16 Future work

7.14.1 Applications

The thesis demonstrated the capability of sparse representation to spatially enhanced medium multispectral data sets (i.e., Landsat OLI and Landsat ETM+) for urban mapping. However, future work could assess the accuracy of sparse representation in more general urban applications such as urban creep, urban growth, urban imperviousness, and urban land cover monitoring. Most of these applications can be defined using more a generalised classification scheme (i.e., impermeable and permeable classes for imperviousness mapping) compared to the classification scheme which was used (with eight and seven members for Nottingham and Derby, respectively). As a result, it is more likely that performing sparse representation on Landsat data would be more successful for these applications compared to the result found in this thesis.

For more specific examples of urban applications, it would be interesting to demonstrate the sparse representation potential for mapping more detailed tree species and roofs materials. Mapping two tree types (i.e., broad-leaved and needle-leaved trees) using WV-2 imagery with 2m spatial resolution in the first part of the thesis. However, because of coarser spatial information offered by spatially enhanced Landsat data using sparse representation, these two tree species merged into a tree class. Perhaps, as a future work, it would be

suggested to perform sparse representation on Sentinel-2 and SPOT data sets with finer spatial resolution compared to Landsat data (10m against 30m) for mapping more detailed tree species. McHale *et al.* (2007) emphasised the importance of detailed inventories of urban vegetation for conservation, planting and maintenance applications of urban trees.

Identifying roof types using remotely sensing technology is becoming a new area of the interest (Taherzadeh *et al.,* 2014; Samsudin *et al.,* 2015; Osińska-Skotak and Ostrowski, 2015). The classification scheme of this thesis was limited to four types of roof materials (metal, concrete, slate, and clay). The potential future study includes an assessment of the accuracy of sparse representation on Sentinel-2 and SPOT data sets, for the identification of different roofing materials.

7.14.2 Sparse representation algorithm

Although both WV-2 and Landsat OLI offer eight multispectral optical bands; because of spectral dissimilarity, only four spectral bands of WV-2 were used to enhance its corresponding bands of Landsat data. The importance of using a higher number of spectral bands for detailed urban mapping was demonstrated within the first part of the thesis (Chapter 3). As a result, it would be valuable to investigate the possibility of utilising other dissimilar spectral bands between WV-2 and Landsat OLI (i.e., using NIR2 of WV-2 as training to enhance SWIR1 and 2 of Landsat OLI), for instance.

As a preliminary study, the capability of sparse representation was assessed by comparing WV-2 against raw and pan-sharpened versions of Landsat OLI, to

spatially enhance Landsat OLI. It was found that using finer spatial resolution imagery as a training database was more successful for increasing the spatial information of Landsat OLI. Since starting working on this thesis, new sensors with finer spatial resolution emerged. In medium resolution category Sentinel-2 with 10m spatial resolution on four spectral bands (i.e., RGB and NIR spectral bands), while in the VHR category VW-3 imagery with a spatial resolution of 1.36m on eight spectral bands (similar spectral variation as WV-2). The availability of such data sets with finer spatial resolution suggests the finer accuracy of sparse representation and becoming eligible for wider applications which are required a greater level of spatial information.

As previously discussed in Section 7.12, it is important that the dictionary represents the similar statistical nature (i.e., texture variations, land cover types, and structural patterns) of input coarse resolution imagery. The role of dictionary size becomes more important for large-scale urban applications, which use a single imagery as the training database. In this thesis, the number of samples and dictionary size were allocated as 100,000 and 2,048. It may be a good idea to explore the sparse representation accuracy by using a greater sample number and dictionary size, while it should be noted that increasing the values of these parameters would significantly increase the computation processing time.

This sparse representation method works in a patch-based manner and spatially enhanced imagery recovered through individual patches. Elad (2015) claimed that allocating a larger patch size can increase the possibility of overlapping between patches because it is expected that averaging two values

for a pixel offers more realistic results compared to a single estimation for the pixel. However, increasing the patch size can significantly increase the computation processing time. Two different patch sizes were used in the thesis, 15 for zoom factor 7.5 and 30 for zoom factor 15. As a future study, the influence of adopting greater patch size can be considered on the accuracy of sparse representation.

Bibliography

AHARON, M., ELAD, M. & BRUCKSTEIN, A. 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Ieee Transactions on Signal Processing,* 54**,** 4311-4322.

AKGUN, T., ALTUNBASAK, Y. & MERSEREAU, R. M. 2005. Super-resolution reconstruction of hyperspectral images. *Ieee Transactions on Image Processing,* 14**,** 1860-1875.

APLIN, P. 2003a. Comparison of simulated IKONOS and SPOT HRV imagery for classifying urban areas. *In:* MESEV, *In Remotely-Sensed Cities.* London, UK: Taylor and Francis.

APLIN, P. 2003b. Remote sensing: base mapping. *Progress in Physical Geography,* 27**,** 275-283.

APLIN, P. & ATKINSON, P. M. 2001. Sub-pixel land cover mapping for per-field classification. *International Journal of Remote Sensing,* 22**,** 2853-2858.

APLIN, P., ATKINSON, P. M. & CURRAN, P. J. 1999. Fine spatial resolution simulated satellite sensor imagery for land cover mapping in the United Kingdom. *Remote Sensing of Environment,* 68**,** 206-216.

APLIN, P. & SMITH, G. M. 2011. Introduction to object-based landscape analysis. *International Journal of Geographical Information Science,* 25**,** 869-875.

ARDILA, J. P., BIJKER, W., TOLPEKIN, V. A. & STEIN, A. 2012. Context-sensitive extraction of tree crown objects in urban areas using VHR satellite images. *International Journal of Applied Earth Observation and Geoinformation,* 15**,** 57-69.

ARNOLD, C. L. & GIBBONS, C. J. 1996. Impervious surface coverage - The emergence of a key environmental indicator. *Journal of the American Planning Association,* 62**,** 243-258.

ATKINSON, P. M. 2005. Sub-pixel target mapping from soft-classified, remotely sensed imagery. *Photogrammetric Engineering and Remote Sensing,* 71**,** 839-846.

ATKINSON, P. M. 2009. Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study. *International Journal of Remote Sensing,* 30**,** 5293-5308.

ATKINSON, P. M. & APLIN, P. 2004. Spatial variation in land cover and choice of spatial resolution for remote sensing. *International Journal of Remote Sensing,* 25**,** 3687-3702.

ATKINSON, P. M., CUTLER, M. E. J. & LEWIS, H. 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing,* 18**,** 917-935.

AZZARI, L. & FOI, A. 2016. Variance Stabilization for Noisy plus Estimate Combination in Iterative Poisson Denoising. *Ieee Signal Processing Letters,* 23**,** 1086-1090.

BABU, S. & Murthy, S. 2011. A Survey on the methods of Super-resolution Image Reconstruction, *International Journal of Computer Applications*, 15, 1-6.

BAKER, S. & KANADE, T. 2002. Limits on super-resolution and how to break them. *Ieee Transactions on Pattern Analysis and Machine Intelligence,* 24**,** 1167-1183.

BAN, Y., HU, H. & RANGEL, I. M. 2010. Fusion of Quickbird MS and RADARSAT SAR data for urban land-cover mapping: object-based and knowledge-based approach. *International Journal of Remote Sensing,* 31**,** 1391-1410.

BASAEED, E., BHASKAR, H., AL-MUALLA, M. & IEEE 2013. Comparative Analysis of Pan-sharpening Techniques on DubaiSat-1 images. *2013 16th International Conference on Information Fusion (Fusion)***,** 227-234.

BENZ, U. C., HOFMANN, P., WILLHAUCK, G., LINGENFELDER, I. & HEYNEN, M. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing,* 58**,** 239-258.

BHASKARAN, S., PARAMANANDA, S. & RAMNARAYAN, M. 2010. Per-pixel and object-oriented classification methods for mapping urban features using Ikonos satellite data. *Applied Geography,* 30**,** 650-665.

BLASCHKE, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65**,** 2-16.

BLASCHKE, T., HAY, G. J., KELLY, M., LANG, S., HOFMANN, P., ADDINK, E., FEITOSA, R. Q., VAN DER MEER, F., VAN DER WERFF, H., VAN COILLIE, F. & TIEDE, D. 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. *Isprs Journal of Photogrammetry and Remote Sensing,* 87**,** 180-191.

BLASCHKE, T., HAY, G. J., WENG, Q. & RESCH, B. 2011. Collective Sensing: Integrating Geospatial Technologies to Understand Urban Systems-An Overview. *Remote Sensing,* 3**,** 1743-1776.

BOSE, N. K. & AHUJA, N. A. 2006. Superresolution and noise filtering using moving least squares. *Ieee Transactions on Image Processing,* 15**,** 2239-2248.

BOSE, N. K., KIM, H. C., VALENZUELA, H. M. & IEEE 1993. *Recursive Implementation of Total Least-Squares Algoryhm for Image-Reconsruction From Noisy, Undersampled Multiframes*.

BOUCHER, A. 2009. Sub-pixel Mapping of Coarse Satellite Remote Sensing Images with Stochastic Simulations from Training Images. *Mathematical Geosciences,* 41**,** 265-290.

BOUCHER, A. & KYRIAKIDIS, P. C. 2006. Super-resolution land cover mapping with indicator geostatistics. *Remote Sensing of Environment,* 104**,** 264-282.

BOUZIANI, M., GOITA, K. & HE, D. C. 2010. Rule-Based Classification of a Very High Resolution Image in an Urban Environment Using Multispectral Segmentation Guided by Cartographic Data. *Ieee Transactions on Geoscience and Remote Sensing,* 48**,** 3198-3211.

BOYD, D. S., SANCHEZ-HERNANDEZ, C. & FOODY, G. M. 2006. Mapping a specific class for priority habitats monitoring from satellite sensor data. *International Journal of Remote Sensing,* 27**,** 2631-2644.

BRUZZONE, L. & CARLIN, L. 2006. A multilevel context-based system for classification of very high spatial resolution images. *Ieee Transactions on Geoscience and Remote Sensing,* 44**,** 2587-2600.

CAI, S. & LIU, D. 2013. A comparison of object-based and contextual pixel-based classifications using high and medium spatial resolution images. *Remote Sensing Letters,* 4**,** 998-1007.

CALLICO, G. M., LOPEZ, S., SOSA, O., LOPEZ, J. F. & SARMIENTO, R. 2008. Analysis of fast block matching motion estimation algorithms for video super-resolution systems. *Ieee Transactions on Consumer Electronics,* 54**,** 1430-1438.

CAMPBELL, J. B. 2007. *Introduction to remote sensing,* New York, The Guilford Press.

CAMPBELL, J. B. & WYNNE, R. H. 2011. *Introduction to remote sensing [electronic resource],* New York, Guilford Press.

CAND`ES, E. 2006. Compressive sampling. *in Proceedings of the International Congress of Mathematicians.*

CANDES, E. J. & TAO, T. 2006. Near-optimal signal recovery from random projections: Universal encoding strategies? *Ieee Transactions on Information Theory,* 52**,** 5406-5425.

CCRS 2016. Fundamental of Remote Sensing. *In:* TUTORIAL, Canada: Natural Resources Canada.

CHANG, H., YEUNG, D. Y., XIONG, Y. & SOCIETY, I. C. 2004. Super-resolution through neighbor embedding. *Proceedings of the 2004 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 1.*

CHATTERJEE, P. & MILANFAR, P. 2012. Patch-Based Near-Optimal Image Denoising. *Ieee Transactions on Image Processing,* 21**,** 1635-1649.

CHAUDHURI, S. & JOSHI, M. 2005. Research on image super-resolution. *In:* CHAUDHURI, S. & JOSHI, M. (eds.) *Motion-Free Super-resolution.* New York: Springer Science and Business Media.

CHEN, Y., SHI, P., FUNG, T., WANG, J. & LI, X. 2007. Object-oriented classification for urban land cover mapping with ASTER imagery. *International Journal of Remote Sensing,* 28**,** 4645-4651.

CHENG, H. 2015. *Sparse Representation, Modeling and Learning and Learning in Visual Recognition,* NLondon, Springer-Verlag.

CHI, M. M., FENG, R. & BRUZZONE, L. 2008. Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Advances in Space Research,* 41**,** 1793-1799.

CHOI, M. 2006. A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter. *Ieee Transactions on Geoscience and Remote Sensing,* 44**,** 1672-1682.

CHUBEY, M. S., FRANKLIN, S. E. & WULDER, M. A. 2006. Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering and Remote Sensing,* 72**,** 383-394.

CIVCO, D. L. & WITHARANA, C. 2012. Assessing the spatial fidelity of resolution-enhanced imagery using Fourier analysis: a proof-of-concept study. *Earth Resources and Environmental Remote Sensing/Gis Applications Iii,* 8538**,** 13.

CLEVE, C., KELLY, M., KEARNS, F. R. & MORITZ, M. 2008. Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. *Computers, Environment and Urban Systems,* 32**,** 317-326.

CLINTON, N., HOLT, A., SCARBOROUGH, J., YAN, L. & GONG, P. 2010. Accuracy Assessment Measures for Object-based Image Segmentation Goodness. *Photogrammetric Engineering and Remote Sensing,* 76**,** 289-299.

CONGALTON, R. G. & GREEN, K. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices,* Boca Raton, FL, USA, Taylor & Francis.

CORTES, C. & VAPNIK, V. 1995. Support-Vector Networks. *Machine Learning,* 20**,** 273-297.

COUNCIL, D. C. 2016. *Derby's population* [Online]. Available: http://www.derby.gov.uk/council-and-democracy/statistics-and-census-information/census-information/population/ [Accessed 2016].

CUSHNIE, J. L. 1987. The Interactive Effect of Spatial-Resolution and Degree of Internal Variability Within Land-cover Types on Classification Accuracies. *International Journal of Remote Sensing,* 8**,** 15-29.

DEAN, A. M. & SMITH, G. M. 2003. An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. *International Journal of Remote Sensing,* 24**,** 2905-2920.

DIXON, B. & CANDADE, N. 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? *International Journal of Remote Sensing,* 29**,** 1185-1206.

DONG, W., LI, X., ZHANG, L., SHI, G. & IEEE 2011. Sparsity-based Image Denoising via Dictionary Learning and Structural Clustering. *2011 Ieee Conference on Computer Vision and Pattern Recognition.*

DONOHO, D. L. 2006. Compressed sensing. *Ieee Transactions on Information Theory,* 52**,** 1289-1306.

DU, Q., YOUNAN, N. H., KING, R. & SHAH, V. P. 2007. On the performance evaluation of pan-sharpening techniques. *Ieee Geoscience and Remote Sensing Letters,* 4**,** 518-522.

DUNBAR, B. 2016. *Landsat 6* [Online]. National Aeronautics and Space Administration. Available: http://www.nasa.gov/ [Accessed 2016].

EGILMAN, D. & MENENDEZ, L. M. 2011. A Case of Occupational Peritoneal Mesothelioma From Exposure to Tremolite-Free Chrysotile in Quebec, Canada: A Black Swan Case. *American Journal of Industrial Medicine,* 54**,** 153-156.

ELAD, M. 2010. *Sparse and Redundant Representations,* Berlin, Germany, Springer.

ELAD, M. 2015. *Sparse and Redundant Representations From Theory to Applications in Signal and Image Processing*.

ELAD, M. & AHARON, M. 2006. Image denoising via sparse and redundant representations over learned dictionaries. *Ieee Transactions on Image Processing,* 15**,** 3736-3745.

ERBEK, F. S., ÖZKAN, C. & TABERNER, M. 2004. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing,* 25**,** 1733-1748.

ERENER, A. & DUZGUN, H. S. 2009. A methodology for land use change detection of high resolution pan images based on texture analysis. *Rivista Italiana Di Telerilevamento,* 41**,** 47-59.

FARSIU, S., ROBINSON, D., ELAD, M. & MILANFAR, P. 2004a. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology,* 14**,** 47-57.

FARSIU, S., ROBINSON, M. D., ELAD, M. & MILANFAR, P. 2004b. Fast and robust multiframe super resolution. *Ieee Transactions on Image Processing,* 13**,** 1327-1344.

FENG, X. 2016. *Enhancing the spatial resolution of satellite-derived land surface temperature mapping for urban areas.* Doctor of Philosophy, University of Nottingham.

FILIPOVIC, M. & KOPRIVA, I. 2011. A Comparison of Dictionary Based Approaches to Inpainting and Denoising With an Emphasis to Independent Component Analysis Learned Dictionaries. *Inverse Problems and Imaging,* 5**,** 815-841.

FIUMI, L. 2012. Surveying the roofs of Rome. *Journal of Cultural Heritage,* 13**,** 304-313.

FLANDERS, D., HALL-BEYER, M. & PEREVERZOFF, J. 2003. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing,* 29**,** 441-452.

FOODY, G. M. 1999. The continuum of classification fuzziness in thematic mapping. *Photogrammetric Engineering and Remote Sensing,* 65**,** 443-451.

FOODY, G. M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment,* 80**,** 185-201.

FOODY, G. M. 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing,* 70**,** 627-633.

FOODY, G. M. 2008. Harshness in image classification accuracy assessment. *International Journal of Remote Sensing,* 29**,** 3137-3158.

FOODY, G. M. & ARORA, M. K. 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing,* 18**,** 799-810.

FOODY, G. M., BOYD, D. S. & SANCHEZ-HERNANDEZ, C. 2007. Mapping a specific class with an ensemble of classifiers. *International Journal of Remote Sensing,* 28**,** 1733-1746.

FOODY, G. M., MUSLIM, A. M. & ATKINSON, P. M. 2005. Super-resolution mapping of the waterline from remotely sensed data. *International Journal of Remote Sensing,* 26**,** 5381-5392.

FRANKLIN, S. E. & WULDER, M. A. 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography,* 26**,** 173-205.

FREEDMAN, G. & FATTAL, R. 2011. Image and Video Upscaling from Local Self-Examples. *Acm Transactions on Graphics,* 30.

GAO, J., CHEN, H. F., ZHANG, Y. & ZHA, Y. 2004. Knowledge-based approaches to accurate mapping of mangroves from satellite data. *Photogrammetric Engineering and Remote Sensing,* 70**,** 1241-1248.

GE, Y., LI, S. & LAKHAN, V. C. 2009. Development and Testing of a Subpixel Mapping Algorithm. *Ieee Transactions on Geoscience and Remote Sensing,* 47**,** 2155-2164.

GHAHREMANI, M. & GHASSEMIAN, H. 2016. Nonlinear IHS: A Promising Method for Pan-Sharpening. *Ieee Geoscience and Remote Sensing Letters,* 13**,** 1606-1610.

GHOLOOBI, M. & KUMAR, L. 2015. Using object-based hierarchical classification to extract land use land cover classes from high-resolution satellite imagery in a complex urban area. *Journal of Applied Remote Sensing,* 9**,** 15.

GHOSH, A. & JOSHI, P. K. 2013. Assessment of pan-sharpened very high-resolution WorldView-2 images. *International Journal of Remote Sensing,* 34**,** 8336-8359.

GIRYES, R. & ELAD, M. 2014. Sparsity-Based Poisson Denoising With Dictionary Learning. *Ieee Transactions on Image Processing,* 23**,** 5057-5069.

GLASNER, D., BAGON, S., IRANI, M. & IEEE 2009. Super-Resolution from a Single Image. *2009 Ieee 12th International Conference on Computer Vision.*

GOBEL, P., DIERKES, C. & COLDEWEY, W. G. 2007. Storm water runoff concentration matrix for urban areas. *Journal of Contaminant Hydrology,* 91**,** 26-42.

GOETZ, S. J., WRIGHT, R. K., SMITH, A. J., ZINECKER, E. & SCHAUB, E. 2003. IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region. *Remote Sensing of Environment,* 88**,** 195-208.

GRIMM, N. B., FAETH, S. H., GOLUBIEWSKI, N. E., REDMAN, C. L., WU, J. G., BAI, X. M. & BRIGGS, J. M. 2008. Global change and the ecology of cities. *Science,* 319**,** 756-760.

GU, Y., ZHANG, Y. & ZHANG, J. 2008. Integration of spatial-spectral information for resolution enhancement in hyperspectral images. *Ieee Transactions on Geoscience and Remote Sensing,* 46**,** 1347-1358.

GU, Y. F., WANG, Q. W., JIA, X. P. & BENEDIKTSSON, J. A. 2015. A Novel MKL Model of Integrating LiDAR Data and MSI for Urban Area Classification. *Ieee Transactions on Geoscience and Remote Sensing,* 53**,** 5312-5326.

GUINDON, B., ZHANG, Y. & DILLABAUGH, C. 2004. Landsat urban mapping based on a combined spectral-spatial methodology. *Remote Sensing of Environment,* 92**,** 218-232.

HAMEDIANFAR, A. 2015. *Integrating of Object-based Image Analysis and Data Mining Techniques for Detailed Urban Mapping Using Very High*

Resolution Remote Sensing. Doctor of Philosophy, UTM University, Kuala Lumpur.

HAMEDIANFAR, A. & SHAFRI, H. Z. M. 2015. Detailed intra-urban mapping through transferable OBIA rule sets using WorldView-2 very-high-resolution satellite images. *International Journal of Remote Sensing,* 36**,** 3380-3396.

HAMEDIANFAR, A., SHAFRI, H. Z. M., MANSOR, S. & AHMAD, N. 2014. Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of Remote Sensing,* 35**,** 1876-1899.

HAN, W. S. & BURIAN, S. J. 2009. Determining Effective Impervious Area for Urban Hydrologic Modeling. *Journal of Hydrologic Engineering,* 14**,** 111-120.

HANSEN, M. C. & LOVELAND, T. R. 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment,* 122**,** 66-74.

HARDIE, R. C., BARNARD, K. J. & ARMSTRONG, E. E. 1997. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *Ieee Transactions on Image Processing,* 6**,** 1621-1633.

HAY, G. & CASTILLA, G. 2006. Object-based image analysis: strengths, weakness, Opportunities and threats (SWOT). *presented at the 1st Int. Conf. Object-Based Image Analysis (OBIA),.* Salzburg.

HEIDEN, U., SEGL, K., ROESSNER, S. & KAUFMANN, H. 2007. Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data. *Remote Sensing of Environment,* 111**,** 537-552.

HEROLD, M., GARDNER, M. & ROBERTS, D. 2003b. spectral resolution requirements for mapping the urban areas. *IEEE Transactions on Geoscience and Remote Sensing,* 14**,** 1907 - 1919.

HEROLD, M., LIU, X. H. & CLARKE, K. C. 2003c. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing,* 69**,** 991-1001.

HEROLD, M., ROBERTS, D. A., GARDNER, M. E. & DENNISON, P. E. 2004. Spectrometry for urban area remote sensing - Development and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment,* 91**,** 304-319.

HIRANO, A., WELCH, R. & LANG, H. 2003. Mapping from ASTER stereo image data: DEM validation and accuracy assessment. *Isprs Journal of Photogrammetry and Remote Sensing,* 57**,** 356-370.

HOLLAND, J. & APLIN, P. 2013. Super-resolution image analysis as a means of monitoring bracken (Pteridium aquilinum) distributions. *Isprs Journal of Photogrammetry and Remote Sensing,* 75**,** 48-63.

HOU, B., ZHU, Z., LI, G. & YU, A. 2016. An Efficient Algorithm for Overcomplete Sparsifying Transform Learning with Signal Denoising. *Mathematical Problems in Engineering*.

HSIEH, P. F., LEE, L. C. & CHEN, N. Y. 2001. Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *Ieee Transactions on Geoscience and Remote Sensing,* 39**,** 2657-2663.

HU, M. G., WANG, J. F. & GE, Y. 2009. Super-resolution reconstruction of remote sensing images using multifractal analysis. *Sensors (Basel),* 9**,** 8669-83.

HU, X. & WENG, Q. 2009. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sensing of Environment,* 113**,** 2089-2102.

HUANG, C., DAVIS, L. S. & TOWNSHEND, J. R. G. 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing,* 23**,** 725-749.

HUANG, L. & NI, L. 2008. *Object-oriented classification of high resolution satellite image for better accuracy*.

HUANG, X. & ZHANG, L. P. 2013. An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery. *Ieee Transactions on Geoscience and Remote Sensing,* 51**,** 257-272.

INGLADA, J. & MICHEL, J. 2009. Qualitative Spatial Reasoning for High-Resolution Remote Sensing Image Analysis. *Ieee Transactions on Geoscience and Remote Sensing,* 47**,** 599-612.

IPPOLITI-RAMILO, G. A., EPIPHANIO, J. C. N. & SHIMABUKURO, Y. E. 2003. Landsat-5 Thematic Mapper data for pre-planting crop area evaluation in tropical countries. *International Journal of Remote Sensing,* 24**,** 1521-1534.

JALAN, S. & SOKHI, B. S. 2012. Comparison of different pan-sharpening methods for spectral characteristic preservation: multi-temporal CARTOSAT-1 and IRS-P6 LISS-IV imagery. *International Journal of Remote Sensing,* 33**,** 5629-5643.

JANSSEN, L. L. F. & VANDERWEL, F. J. M. 1994. Accuracy Assessment of Satellite-Derived Land-Cover Data - A Review. *Photogrammetric Engineering and Remote Sensing,* 60**,** 419-426.

JAVAN, F. D., SAMADZADEGAN, F. & REINARTZ, P. 2013. Spatial Quality Assessment of Pan-Sharpened High Resolution Satellite Imagery Based

on an Automatically Estimated Edge Based Metric. *Remote Sensing,* 5**,** 6539-6559.

JEBUR, M. N., SHAFRI, H. Z. M., PRADHAN, B. & TEHRANY, M. S. 2014. Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery. *Geocarto International,* 29**,** 792-806.

KANAREK, M. S. 2011. Mesothelioma from Chrysotile Asbestos: Update. *Annals of Epidemiology,* 21**,** 688-697.

KARL, J. W. & MAURER, B. A. 2010. Multivariate correlations between imagery and field measurements across scales: comparing pixel aggregation and image segmentation. *Landscape Ecology,* 25**,** 591-605.

KASETKASEM, T., ARORA, M. & VARSHNEY, P. 2005. Super-resolution land cover mapping using a Markov random field based approach. *Remote Sensing of Environment,* 96**,** 302-314.

KAVZOGLU, T. & COLKESEN, I. 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation,* 11**,** 352-359.

KEERTHI, S. S. & LIN, C. J. 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation,* 15**,** 1667-1689.

KEUCHEL, J., NAUMANN, S., HEILER, M. & SIEGMUND, A. 2003. Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. *Remote Sensing of Environment,* 86**,** 530-541.

KHAN, M. M., CHANUSSOT, J., CONDAT, L. & MONTANVERT, A. 2008. Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *Ieee Geoscience and Remote Sensing Letters,* 5**,** 98-102.

KIM, M., WARNER, T. A., MADDEN, M. & ATKINSON, D. S. 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing,* 32**,** 2825-2850.

KIM, S. P., BOSE, N. K. & VALENZUELA, H. M. 1990. Recursive reconstrduction of high-resolution image from noisy undersampled multiframes. *Ieee Transactions on Acoustics Speech and Signal Processing,* 38**,** 1013-1027.

KIM, S. P. & SU, W.-Y. 1993. Recursive high-resolution reconstruction of blurred multiframe images. *Ieee Transactions on Image Processing,* 2**,** 534-539.

LALIBERTE, A. S., RANGO, A., HAVSTAD, K. M., PARIS, J. F., BECK, R. F., MCNEELY, R. & GONZALEZ, A. L. 2004. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment,* 93**,** 198-210.

LANG, S., SCHOPFER, E., HOLBLING, D., BLASCHKE, T., MOELLER, M., JEKEL, T. & KLOYBER, E. 2008. *In:* PETROSILLO, F., MÜLLER, B., JONES, G., ZURLINI,

K., KRAUZE, S., VICTOROV, L. & LI AND W.G. KEPNER. (eds.) *In Use of Landscape Sciences for the Assessment of Environmental Security.* Dordrecht, Germany: Springer.

LATTY, R. S., NELSON, R., MARKHAM, B., WILLIAMS, D., TOLL, D. & IRONS, J. 1985. Performance comparisons between information extraction techniques using variable spatial-resolution data. *Photogrammetric Engineering and Remote Sensing,* 51**,** 1459-1470.

LEE, H., BATTLE, A. & RAINA, R. 2007. Efficient sparse coding algorithms. *Advances in Neural Information Processing Systems***,** 801 – 808.

LI, F., TANG, L., LI, C., GUO, Y. & GAO, J. 2013a. A new super resolution method based on combined sparse representations for remote sensing imagery. *In:* BRUZZONE, L. (ed.) *Image and Signal Processing for Remote Sensing Xix.*

LI, F., TANG, L. L., LI, C. R., GUO, Y. & GAO, J. B. 2013b. A new super resolution method based on combined sparse representations for remote sensing imagery. *Image and Signal Processing for Remote Sensing Xix,* 8892**,** 7.

LI, X., TIAN, L., ZHAO, X. & CHEN, X. 2011. A super resolution approach for spectral unmixing of remote sensing images. *International Journal of Remote Sensing,* 32**,** 6091-6107.

LI, X. X. & SHAO, G. F. 2014. Object-based land-cover mapping with high resolution Aerial photography at a county scale in Midwestern USA. *Remote Sensing,* 6**,** 11372-11390.

LILLESAND, T. M., KIEFER, R. W. & CHIPMAN, J. W. 2008. *Remote sensing and image interpretation,* Hoboken, N.J., Wiley.

LING, F., DU, Y., XIAO, F., XUE, H. & WU, S. 2010. Super-resolution land-cover mapping using multiple sub-pixel shifted remotely sensed images. *International Journal of Remote Sensing,* 31**,** 5023-5040.

LING, F., LI, W. B., DU, Y. & LI, X. D. 2011. Land cover change mapping at the subpixel scale with different spatial-resolution remotely sensed imagery. *Ieee Geoscience and Remote Sensing Letters,* 8**,** 182-186.

LING, F., LI, X., XIAO, F., FANG, S. & DU, Y. 2012. Object-based sub-pixel mapping of buildings incorporating the prior shape information from remotely sensed imagery. *International Journal of Applied Earth Observation and Geoinformation,* 18**,** 283-292.

LING, F., LI, X. D., DU, Y. & XIAO, F. 2013. Sub-pixel mapping of remotely sensed imagery with hybrid intra- and inter-pixel dependence. *International Journal of Remote Sensing,* 34**,** 341-357.

LIU, C., SHUM, H.-Y. & FREEMAN, W. T. 2007. Face hallucination: Theory and practice. *International Journal of Computer Vision,* 75**,** 115-134.

LIU, D. & XIA, F. 2010. Assessing object-based classification: advantages and limitations. *Remote Sensing Letters,* 1**,** 187-194.

LIU, J. J., SHI, L. S., ZHANG, C., YANG, J. Y., ZHU, D. H. & YANG, J. H. 2013. A Variable Multi-Scale Segmentation Method for Spatial Pattern Analysis Using Multispectral World View-2 Images. *Sensor Letters,* 11**,** 1055-1061.

LOW, F. & DUVEILLER, G. 2014. Defining the Spatial Resolution Requirements for Crop Identification Using Optical Remote Sensing. *Remote Sensing,* 6**,** 9034-9063.

LU, D. & WENG, Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing,* 28**,** 823-870.

LU, D. S., HETRICK, S. & MORAN, E. 2010. Land Cover Classification in a Complex Urban-Rural Landscape with QuickBird Imagery. *Photogrammetric Engineering and Remote Sensing,* 76**,** 1159-1168.

LU, D. S., HETRICK, S. & MORAN, E. 2011a. Impervious surface mapping with Quickbird imagery. *International Journal of Remote Sensing,* 32**,** 2519-2533.

LU, D. S., MORAN, E. & HETRICK, S. 2011b. Detection of impervious surface change with multitemporal Landsat images in an urban-rural frontier. *Isprs Journal of Photogrammetry and Remote Sensing,* 66**,** 298-306.

LU, D. S. & WENG, Q. H. 2004. Spectral mixture analysis of the urban landscape in Indianapolis with landsat ETM plus imagery. *Photogrammetric Engineering and Remote Sensing,* 70**,** 1053-1062.

MAHGOUB ELSAYED, I. 2006. *The Effect of Urbanization on the Intensity of the Urban Heat Island: A case study on the City of the Kuala Lumpur, Malaysia.* Doctor of Philosophy, International Islamic University Malaysia.

MAIRAL, J. & BACH, F. 2012. Sparse Modeling for Image and Vision Processing. *Computer Graphics and Vision,* 8**,** 85–283.

MAIRAL, J., BACH, F. & PONCE, J. 2012. Task-Driven Dictionary Learning. *Ieee Transactions on Pattern Analysis and Machine Intelligence,* 34**,** 791-804.

MAIRAL, J., ELAD, M. & SAPIRO, G. 2008a. Sparse representation for color image restoration. *Ieee Transactions on Image Processing,* 17**,** 53-69.

MAIRAL, J., SAPIRO, G. & ELAD, M. 2008b. Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling & Simulation,* 7**,** 214-241.

MAKIDO, Y., SHORTRIDGE, A. & MESSINA, J. P. 2007. Assessing alternatives for Modeling the spatial distribution of multiple land-cover classes at sub-pixel scales. *Photogrammetric Engineering and Remote Sensing,* 73**,** 935-943.

MARCELLO, J., MEDINA, A. & EUGENIO, F. 2013. Evaluation of Spatial and Spectral Effectiveness of Pixel-Level Fusion Techniques. *Ieee Geoscience and Remote Sensing Letters,* 10**,** 432-436.

MARCHISIO, G., PACIFICI, F., PADWICK, C. & IEEE 2010. On the relative predective value of the new spectral bands in the WorldView-2 sensor. *2010 Ieee International Geoscience and Remote Sensing Symposium***,** 2723-2726.

MARJANOVIC, M., KOVACEVIC, M., BAJAT, B. & VOZENILEK, V. 2011. Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology,* 123**,** 225-234.

MATHER, P. & KOCK, M. 2011. *Computer Processing of Remotely-Sensed Images: An Introduction,* UK, John Wiley & Sons.

MATHERON, G. 1965. *Les variables régionalisées et leur estimation,* Paris, Masson et Cie.

MCCOY, R. M. 2005. *Field Methods in Remote Sensing*.

MCHALE, M. R., MCPHERSON, E. G. & BURKE, I. C. 2007. The potential of urban tree plantings to be cost effective in carbon credit markets. *Urban Forestry & Urban Greening,* 6**,** 49-60.

MEENAKSHISUNDARAM, V. 2005. *Quality Assessment of Ikonos and Quickbird Fused Images for Urban Mapping.* PhD, Calgary, Alberta.

MENON, D. & CALVAGNO, G. 2011. Color image demosaicking: An overview. *Signal Processing-Image Communication,* 26**,** 518-533.

MERINO TERESA, M. & NUNEZ, J. 2007. Super-resolution of remotely sensed images with variable-pixel linear reconstruction. *Ieee Transactions on Geoscience and Remote Sensing,* 45**,** 1446-1457.

MERTENS, K. C., DE BAETS, B., VERBEKE, L. P. C. & DE WULF, R. R. 2006. A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models. *International Journal of Remote Sensing,* 27**,** 3293-3310.

MERTENS, K. C., VERBEKE, L. P. C., DUCHEYNE, E. I. & DE WULF, R. R. 2003. Using genetic algorithms in sub-pixel mapping. *International Journal of Remote Sensing,* 24**,** 4241-4247.

MERTENS, K. C., VERBEKE, L. P. C., WESTRA, T. & DE WULF, R. R. 2004. Sub-pixel mapping and sub-pixel sharpening using neural network predicted wavelet coefficients. *Remote Sensing of Environment,* 91**,** 225-236.

MITCHELL, H. B. 2010. *Image Fusion: Theories, Techniques and Applications,* Berlin, Germany, Springer-Verlag.

MOMENI, R., APLIN, P. & BOYD, D. S. 2016. Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach. *Remote Sensing,* 8**,** 88-88.

MOREIRA, R. C. & GALVAO, L. S. 2010. Variation in spectral shape of urban materials. *Remote Sensing Letters,* 1**,** 149-158.

MORTON, D., ROWLAND, C., WOOD, C., MEEK, L., MARSTON, C., SMITH, G., WADSWORTH, R. & I, S. 2011. Final Report for LCM2007 - the new UK Land Cover Map. *In:* HYDROLOGY, C. F. E. (ed.).

MOUNTRAKIS, G., IM, J. & OGOLE, C. 2011. Support vector machines in remote sensing: A review. *Isprs Journal of Photogrammetry and Remote Sensing,* 66**,** 247-259.

MUAD, A. 2011. *Super-Resolution Mapping.* Doctor of Philosophy, University og Nottingham.

MÜLLER, K., STEINMEIER, C. & KÜCHLER, M. 2010. Urban growth along motorways in Switzerland. *Landscape and Urban Planning,* 98**,** 3-12.

MUMBY, P. J. & EDWARDS, A. J. 2002. Mapping marine environments with IKONOS imagery: enhanced spatial resolution can deliver greater thematic accuracy. *Remote Sensing of Environment,* 82**,** 248-257.

MYINT, S. W., GALLETTI, C. S., KAPLAN, S. & KIM, W. K. 2013. Object vs. pixel: a systematic evaluation in urban environments. *Geocarto International,* 28**,** 657-678.

MYINT, S. W., GOBER, P., BRAZEL, A., GROSSMAN-CLARKE, S. & WENG, Q. H. 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment,* 115**,** 1145-1161.

MYINT, S. W. & MESEV, V. 2012. A comparative analysis of spatial indices and wavelet-based classification. *Remote Sensing Letters,* 3**,** 141-150.

NAIK, S. & PATEL, N. 2013. Single Image Super Resolution in Spatial and Wavelet Domain. *The International Journal of Multimedia & Its Applications,* 5**,** 23 - 32.

NAKANISHI-OHNO, Y., OBUCHI, T., OKADA, M. & KABASHIMA, Y. 2016. Sparse approximation based on a random overcomplete basis. *Journal of Statistical Mechanics-Theory and Experiment*.

NASROLLAHI, K. & MOESLUND, T. B. 2014. Super-resolution: a comprehensive survey. *Machine Vision and Applications,* 25**,** 1423-1468.

NEWMAN, M. E., MCLAREN, K. P. & WILSON, B. S. 2011. Comparing the effects of classification techniques on landscape-level assessments: pixel-based versus object-based classification. *International Journal of Remote Sensing,* 32**,** 4055-4073.

NGUYEN, N. & MILANFAR, P. 2000. A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution). *Circuits Systems and Signal Processing,* 19**,** 321-338.

OJHA, C., FUSCO, A., MANUNTA, M. & IEEE 2015. Denosing of full resolution differential SAR interferogram based on K-SVD technique. *2015 Ieee International Geoscience and Remote Sensing Symposium.*

ORDNANCESURVEY. 2009. OS MasterMap Topography Layer, User guide contents. 1.8. [Accessed 04/05/2016].

OSIŃSKA-SKOTAK, K. & OSTROWSKI, W. 2015. Use of Satellite and ALS Data for Classification of Roofing Materials on The Example of Asbestos Roof Tile Identofocation. *Technical Sciences,* 18**,** 283–298.

OTUKEI, J. R. & BLASCHKE, T. 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation,* 12**,** S27-S31.

PAL, M. & MATHER, P. M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment,* 86**,** 554-565.

PAL, M. & MATHER, P. M. 2004. Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems,* 20**,** 1215-1225.

PAL, M. & MATHER, P. M. 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing,* 26**,** 1007-1011.

PAL, N. R. & PAL, S. K. 1993. A review on image image segmentation techniques. *Pattern Recognition,* 26**,** 1277-1294.

PALUBINSKAS, G. 2013. Fast, simple, and good pan-sharpening method. *Journal of Applied Remote Sensing,* 7.

PAPYAN, V. & ELAD, M. 2016. Multi-Scale Patch-Based Image Restoration. *Ieee Transactions on Image Processing,* 25**,** 249-261.

PARK, S. C., PARK, M. K. & KANG, M. G. 2003. Super-resolution image reconstruction: A technical overview. *Ieee Signal Processing Magazine,* 20**,** 21-36.

PATTI, A. J., SEZAN, M. I. & TEKALP, A. M. 1997. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *Ieee Transactions on Image Processing,* 6**,** 1064-1076.

PAUL, M. J. & MEYER, J. L. 2001. Streams in the urban landscape. *Annual Review of Ecology and Systematics,* 32**,** 333-365.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* 12**,** 2825-2830.

PELEG, T. & ELAD, M. 2014. A Statistical Prediction Model Based on Sparse Representations for Single Image Super-Resolution. *Ieee Transactions on Image Processing,* 23**,** 2569-2582.

PHAM, D.-S., VENKATESH, S. & IEEE 2008. Joint learning and dictionary construction for pattern recognition. *2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12.*

PIOTR, W., PAWEL, H., MARTA, S., MARCIN, P. & ROELAND, D., KOKB 2014. land use and land cover map of water catchments areas in south poland, based on geobia multi stage approach. *South Eastern European Journal of Earth Observation and Geomatics* 3.

PLATT, R. V. & RAPOZA, L. 2008. An Evaluation of an Object-Oriented Paradigm for Land Use/Land Cover Classification∗. *The Professional Geographer,* 60**,** 87-100.

PLAZA, A., BENEDIKTSSON, J. A., BOARDMAN, J. W., BRAZILE, J., BRUZZONE, L., CAMPS-VALLS, G., CHANUSSOT, J., FAUVEL, M., GAMBA, P., GUALTIERI, A., MARCONCINI, M., TILTON, J. C. & TRIANNI, G. 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment,* 113**,** S110-S122.

POURSANIDIS, D., CHRYSOULAKIS, N. & MITRAKA, Z. 2015. Landsat 8 vs. Landsat 5: A comparison based on urban and pen-urban land cover mapping. *International Journal of Applied Earth Observation and Geoinformation,* 35**,** 259-269.

POWELL, R. L., ROBERTS, D. A., DENNISON, P. E. & HESS, L. L. 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment,* 106**,** 253-267.

PRICE, J. C. 1997. Spectral band selection for visible near infrared remote sensing: Spectral-spatial resolution tradeoffs. *Ieee Transactions on Geoscience and Remote Sensing,* 35**,** 1277-1285.

PRIEM, F. & CANTERS, F. 2016. Synergistic Use of LiDAR and APEX Hyperspectral Data for High-Resolution Urban Land Cover Mapping. *Remote Sensing,* 8**,** 787-787.

PROTTER, M. & ELAD, M. 2009. Super Resolution With Probabilistic Motion Estimation. *Ieee Transactions on Image Processing,* 18**,** 1899-1904.

PROTTER, M., ELAD, M., TAKEDA, H. & MILANFAR, P. 2009. Generalizing the Nonlocal-Means to Super-Resolution Reconstruction. *Ieee Transactions on Image Processing,* 18**,** 36-51.

PU, R., LANDRY, S. & YU, Q. 2011. Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery. *International Journal of Remote Sensing,* 32**,** 3285-3308.

QIN, Y. C., LI, S. H., VU, T. T., NIU, Z. & BAN, Y. F. 2015. Synergistic application of geometric and radiometric features of LiDAR data for urban land cover mapping. *Optics Express,* 23**,** 13761-13775.

QUACKENBUSH, L. J., HOPKINS, P. F. & KINN, G. J. 2000. Developing forestry products from high resolution digital aerial imagery. *Photogrammetric Engineering and Remote Sensing,* 66**,** 1337-1346.

RADOUXA, J., BOGAERTA, P., FASBENDERB, D. & DEFOURNYA, P. 2011. Thematic accuracy assessment of geographic object-based image classification. *International Journal of Geographical Information Science* 25.

RHEE, S. & KANG, M. G. 1999. Discrete cosine transform based regularized high-resolution image reconstruction algorithm. *Optical Engineering,* 38**,** 1348-1356.

RICHARDS, J. & JIA, X. 1999. *Remote Sensing Digital Image Analysis: An Introduction,* New York, Springer-Verlag.

RICOTTA, C. & AVENA, G. C. 1999. The influence of fuzzy set theory on the areal extent of thematic map classes. *International Journal of Remote Sensing,* 20**,** 201-205.

ROBINSON, M. D., TOTH, C. A., LO, J. Y. & FARSIU, S. 2010. Efficient Fourier-Wavelet Super-Resolution. *Ieee Transactions on Image Processing,* 19**,** 2669-2681.

ROGAN, J. & CHEN, D. M. 2004. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning,* 61**,** 301-325.

RUBINSTEIN, R., BRUCKSTEIN, A. M. & ELAD, M. 2010. Dictionaries for Sparse Representation Modeling. *Proceedings of the Ieee,* 98**,** 1045-1057.

SADEGHI, M., BABAIE-ZADEH, M. & JUTTEN, C. 2013. Dictionary Learning for Sparse Representation: A Novel Approach. *Ieee Signal Processing Letters,* 20**,** 1195-1198.

SAMADZADEGAN, F., FATHOLLAHI, F., REZAEE, M. & IEEE 2012. Comparing different IHS-based pan-sharpening techniques for WorldView-2 high resolution satellite imagery. *2012 Ieee International Geoscience and Remote Sensing Symposium.*

SAMSUDIN, S. H., SHAFRI, H. Z. M., HAMEDIANFAR, A. & MANSOR, S. 2015. Spectral feature selection and classification of roofing materials using field spectroscopy data. *Journal of Applied Remote Sensing,* 9.

SANCHEZ, T. W. 2004. Land use and growth impacts from highway capacity increases. *Journal of Urban Planning and Development-Asce,* 130**,** 75-82.

SCHOWENGERDT, R. A. 1997. *Remote sensing, models, and methods for image processing,* San Diego, Calif. ; London, Academic Press.

SHAHDOOSTI, H. R. & GHASSEMIAN, H. 2015. Fusion of MS and PAN Images Preserving Spectral Quality. *Ieee Geoscience and Remote Sensing Letters,* 12**,** 611-615.

SHAN, J., HUSSAIN, E., KIM, K. & BIEHL, L. 2010. Flood Mapping with Satellite Images and its Web Service. *Photogrammetric Engineering and Remote Sensing,* 76**,** 102-105.

SHEN, H., NG, M. K., LI, P. & ZHANG, L. 2009. Super-Resolution Reconstruction Algorithm To MODIS Remote Sensing Images. *Computer Journal,* 52**,** 90-100.

SIU, W.-C., HUNG, K.-W. & IEEE 2012. Review of Image Interpolation and Super-resolution. *2012 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (Apsipa Asc)*.

SMALL, C. & LU, J. W. T. 2006. Estimation and vicarious validation of urban vegetation abundance by spectral mixture analysis. *Remote Sensing of Environment,* 100**,** 441-456.

SONG, H., HUANG, B., LIU, Q. & ZHANG, K. 2015a. Improving the Spatial Resolution of Landsat TM/ETM plus Through Fusion With SPOT5 Images via Learning-Based Super-Resolution. *Ieee Transactions on Geoscience and Remote Sensing,* 53**,** 1195-1204.

SONG, H. H., HUANG, B., LIU, Q. S. & ZHANG, K. H. 2015b. Improving the Spatial Resolution of Landsat TM/ETM plus Through Fusion With SPOT5 Images via Learning-Based Super-Resolution. *Ieee Transactions on Geoscience and Remote Sensing,* 53**,** 1195-1204.

SONG, M., CIVCO, D. L. & HURD, J. D. 2005. A competitive pixel-object approach for land cover classification. *International Journal of Remote Sensing,* 26**,** 4981-4997.

STARK, H. & OSKOUI, P. 1989. High-Resolution Image RecoveryY From Image-Plane Arrays, Using Convex Projections. *Journal of the Optical Society of America a-Optics Image Science and Vision,* 6**,** 1715-1726.

STEHMAN, S. V. 2009. Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing,* 30**,** 5243-5272.

STEHMAN, S. V. & WICKHAM, J. D. 2011. Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. *Remote Sensing of Environment,* 115**,** 3044-3055.

STRAHLER, A. H., WOODCOCK, C. E. & SMITH, J. A. 1986. ON THE NATURE OF MODELS IN REMOTE-SENSING. *Remote Sensing of Environment,* 20**,** 121-139.

STRAIT, M., RAHMANI, S. & MARKURJEV, D. 2008. Evaluation of Pan-Sharpening Methods. *In:* WITTMAN, T. (ed.) *UCLA Department of Mathematics.*

SUN, H. S., WU, J. Q. & XU, X. C. 2016. A Method to Determine Appropriate Spatial Resolution for Hard Image Classification. *Journal of the Indian Society of Remote Sensing,* 44**,** 11-19.

SUN, J., SUN, J., XU, Z., SHUM, H.-Y. & IEEE 2008. Image super-resolution using gradient profile prior. *2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12.*

SUN, J., ZHENG, N. N., TAO, H., SHUM, H. Y., IEEE COMPUTER, S. & IEEE COMPUTER, S. 2003. Image hallucination with primal sketch priors. *2003 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol Ii, Proceedings.*

SURVEY, O. 2015. *OS MasterMap Topgraphy Layer* [Online]. Available: https://www.ordnancesurvey.co.uk/business-and-government/products/topography-layer.html [Accessed 24 September 2015].

SWAIN, P. & DAVIS, S. 1979. *Remote Sensing: The Quantitative Approach,* London, McGraw Hill Higher Education.

TAHERZADEH, E. & SHAFRI, H. 2011. Using hyperspectral remote sensing data in urban mapping over Kuala Lumpur. *Joint Urban Remote Sensing Event.* Munich, Germany: JURSE.

TAHERZADEH, E., SHAFRI, H. & SHAHI, K. 2014. Roof Material Detection Based on Object-BasedApproach Using WorldView-2 Satellite Imagery. *International Journal of Computer, Electrical, Automation, Control and Information Engineering,* 8.

TATEM, A. J., LEWIS, H. G., ATKINSON, P. M. & NIXON, M. S. 2001. Super-resolution target identification from remotely sensed images using a Hopfield neural network. *Ieee Transactions on Geoscience and Remote Sensing,* 39**,** 781-796.

TATEM, A. J., LEWIS, H. G., ATKINSON, P. M. & NIXON, M. S. 2002. Super-resolution land cover pattern prediction using a Hopfield neural network. *Remote Sensing of Environment,* 79**,** 1-14.

TATEM, A. J., LEWIS, H. G., ATKINSON, P. M. & NIXON, M. S. 2003. Increasing the spatial resolution of agricultural land cover maps using a Hopfield neural network. *International Journal of Geographical Information Science,* 17**,** 647-672.

TAUBENBOCK, H., ESCH, T., WURM, M., ROTH, A. & DECH, S. 2010. Object-based feature extraction using high spatial resolution satellite data of urban areas. *Journal of Spatial Science,* 55**,** 117-132.

TEHRANY, M. S., PRADHAN, B. & JEBUV, M. N. 2014. A comparative assessment between object and pixel-based classification approaches for land-use/land-cover mapping using SPOT 5 imagery. *Geocarto International,* 29**,** 351-369.

THOMAS, N., HENDRIX, C. & CONGALTON, R. G. 2003a. A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering and Remote Sensing,* 69**,** 963-972.

THOMAS, V., TREITZ, P., JELINSKI, D., MILLER, J., LAFLEUR, P. & MCCAUGHEY, J. H. 2003b. Image classification of a northern peatland complex using spectral and plant community data. *Remote Sensing of Environment,* 84**,** 83-99.

TIAN, J. & MA, K.-K. 2011. A survey on super-resolution imaging. *Signal, Image and Video Processing,* 5**,** 329-342.

TIPPING, M. & BISHOP, C. 2002. Bayesian image super-resolution. *Advances in Neural Information Processing Systems,* 16**,** 1279–1286.

TOLL, D. L. 1985. Effect of Landsat Thematic Mapper Sensor Parameters on Land Cover Classification. *Remote Sensing of Environment,* 17**,** 129-140.

TOSIC, I. 2011. Dictionary learning: What is the right representation for my signal? *Ieee Signal Processing Letters***,** 27-38.

TOUTIN, T. 2009. The SAGE Handbook of Remote Sensing. *In:* WARNER, A., NELLIS, D. & FOODY, M. (eds.) *The SAGE Handbook of Remote Sensing.* London: SAGE.

TREITZ, P. M., HOWARTH, P. J., SUFFLING, R. C. & SMITH, P. 1992. Application of detailed ground information to vegetation mapping with high spatial-resolution digital imagery. *Remote Sensing of Environment,* 42**,** 65-82.

TRIMBLE. 2016. *eCognition Suite Version 9.1* [Online]. Available: http://www.ecognition.com/ [Accessed 05/06/2016].

TSAI, R. & HUANG, T. 1984. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing,* 1**,** 317-339.

TSO, B. & MATHER, P. M. 2009. *Classification methods for remotely sensed data,* Boca Raton ; London, CRC Press.

TSUTSUMIDA, N. & COMBER, A. J. 2015. Measures of spatio-temporal accuracy for time series land cover data. *International Journal of Applied Earth Observation and Geoinformation,* 41**,** 46-55.

TUIA, D., PACIFICI, F., KANEVSKI, M. & EMERY, W. J. 2009. Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines. *Ieee Transactions on Geoscience and Remote Sensing,* 47**,** 3866-3879.

UNITED NATIONS. 2012. *World urbanization prospects: The 2011 revision* [Online]. Available: http://esa.un.org/unpd/wup/index.htm [Accessed Jully 2012].

UNITED NATIONS. 2015. World Urbanization Prospects: The 2014 Revision. New York, NY, USA,: United Nations.

VAIOPOULOS, A. D. 2011. Developing Matlab scripts for image analysis and quality assessment. *In:* MICHEL, U. & CIVCO, D. L. (eds.) *Earth Resources and Environmental Remote Sensing/Gis Applications Ii.*

VATSAVAI, R. R. 2011. High-resolution Urban Image Classification Using Extended Features. *Data Mining Workshops (ICDMW), IEEE 11th International.* Vancouver, BC: Oak Ridge Nat. Lab., Oak Ridge, TN, USA

VERHOEYE, J. & DE WULF, R. 2002. Land cover mapping at sub-pixel scales using linear optimization techniques. *Remote Sensing of Environment,* 79**,** 96-104.

VOGELMANN, J. E., SOHL, T. & HOWARD, S. M. 1998. Regional characterization of land cover using multiple. *Photogrammetric Engineering and Remote Sensing,* 64**,** 45-57.

VOSS, M. & SUGUMARAN, R. 2008. Seasonal effect on tree species classification in an urban environment using hyperspectral data, LiDAR, and an object-oriented approach. *Sensors,* 8**,** 3020-3036.

WALKER, J. S. & BLASCHKE, T. 2008a. Object-based land-cover classification for the Phoenix metropolitan area: optimization vs. transportability. *International Journal of Remote Sensing,* 29**,** 2021-2040.

WALKER, J. S. & BLASCHKE, T. 2008b. Object-based land-cover classification for the Phoenix metropolitan area: optimization vs. transportability. *International Journal of Remote Sensing,* 29**,** 2021-2040.

WANG, P. & WANG, Y. S. 2015. Malware behavioural detection and vaccine development by using a support vector model classifier. *Journal of Computer and System Sciences,* 81**,** 1012-1026.

WANG, Z. & BOVIK, A. C. 2002. A universal image quality index. *Ieee Signal Processing Letters,* 9**,** 81-84.

WARNER, A., NELLIS, D. & FOODY, M. 2009. *In:* WARNER, A., NELLIS, D. & FOODY, M. (eds.) *The SAGE Handbook of Remote Sensing.* London: SAGE.

WEIDNER, U. 2008. contribution to the assessment of segmentation quality for remote sensing applications. *the 21st Congress for the International Society for Photogrammetry and Remote Sensing.* Beijing, China.

WENG, Q. 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment,* 117**,** 34-49.

WENG, Q. 2014. *Scale Issues in Remote Sensing,* US, Wiley-Interscience.

WHITESIDE, T. G., BOGGS, G. S. & MAIER, S. W. 2011. Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation,* 13**,** 884-893.

WOODCOCK, C. E. & GOPAL, S. 2000. Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science,* 14**,** 153-172.

WULDER, M. A., MASEK, J. G., COHEN, W. B., LOVELAND, T. R. & WOODCOCK, C. E. 2012. Opening the archive: How free data has enabled the science

and monitoring promise of Landsat. *Remote Sensing of Environment,* 122**,** 2-10.

YAN, G., MAS, J. F., MAATHUIS, B. H. P., XIANGMIN, Z. & VAN DIJK, P. M. 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,* 27**,** 4039-4055.

YANG, J. & HUANG, T. 2010. *Image Super-Resolution: Historical Overview and Future Challenges*, CRC Press, Taylor&Francis Group.

YANG, J., LI, P. & HE, Y. 2014. A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation. *Isprs Journal of Photogrammetry and Remote Sensing,* 94**,** 13-24.

YANG, J., WANG, Z., LIN, Z., COHEN, S. & HUANG, T. 2012. Coupled Dictionary Training for Image Super-Resolution. *Ieee Transactions on Image Processing,* 21**,** 3467-3478.

YANG, J., WRIGHT, J., HUANG, T., MA, Y. & IEEE 2008. Image super-resolution as sparse representation of raw image patches. *2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12*.

YANG, J., WRIGHT, J., HUANG, T. S. & MA, Y. 2010. Image Super-Resolution Via Sparse Representation. *Ieee Transactions on Image Processing,* 19**,** 2861-2873.

YIN, H. T. 2015. Sparse representation based pansharpening with details injection model. *Signal Processing,* 113**,** 218-227.

YOCKY, D. A. 1996. Multiresolution wavelet decomposition image merger of landsat thematic mapper and SPOT panchromatic data. *Photogrammetric Engineering and Remote Sensing,* 62**,** 1067-1074.

ZEYDE, R., ELAD, M. & PROTTER, M. 2012. On Single Image Scale-up Using Sparse-Representations. *In:* BOISSONNAT, J., CHENIN, P., COHEN, A., GOUT, C., LYCHE, T., MAZURE, M. & SCHUMAKER, L. (eds.) *7th International Conference Curves Surfaces.* Avignon, France: Springer-Verlag Berlin Heidelberg.

ZHAN, Q., MOLENAAR, M., TEMPFLI, K. & SHI, W. 2005a. Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing,* 26**,** 2953-2974.

ZHAN, Q. M., MOLENAAR, M., TEMPFLI, K. & SHI, W. Z. 2005b. Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing,* 26**,** 2953-2974.

ZHANG, H., FRITTS, J. E. & GOLDMAN, S. A. 2008. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding,* 110**,** 260-280.

ZHANG, H., YANG, J., ZHANG, Y. & HUANG, T. S. 2010. Non-Local Kernel Regression for Image and Video Restoration. *In:* DANIILIDIS, K., MARAGOS, P. & PARAGIOS, N. (eds.) *Computer Vision-Eccv 2010, Pt Iii.*

ZHANG, J., LI, P. & WANG, J. 2014. Urban Built-Up Area Extraction from Landsat TM/ETM+ Images Using Spectral Information and Multivariate Texture. *Remote Sensing,* 6**,** 7339-7359.

ZHANG, Y. 2008. Methods for Image Fusion Quality Assessment - A Review Comparison and Analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science.* Beijing.

ZHOU, W. & TROY, A. 2008. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing,* 29**,** 3119-3135.

ZHOU, W. Q., HUANG, G. L., PICKETT, S. T. A. & CADENASSO, M. L. 2011. 90 years of forest cover change in an urbanizing watershed: spatial and temporal dynamics. *Landscape Ecology,* 26**,** 645-659.

ZHOU, W. Q., TROY, A. & GROVE, M. 2008. Object-based land cover classification and change analysis in the Baltimore metropolitan area using multitemporal high resolution remote sensing data. *Sensors,* 8**,** 1613-1636.

ZHOU, X., JANCSÓ, T., CHEN, C. & MALGORZATA, W. 2012. Urban Land Cover Mapping Based on Object Oriented Classification Using WorldView 2 Satellite Remote Sensing Images. *International Scientific Conference on Sustainable Development & Ecological Footprint.* Sopron, Hungary.

ZHOU, Y. H. & QIU, F. 2015. Fusion of high spatial resolution WorldView-2 imagery and LiDAR pseudo-waveform for object-based image analysis. *Isprs Journal of Photogrammetry and Remote Sensing,* 101**,** 221-232.

ZIAEI, Z., PRADHAN, B. & BIN MANSOR, S. 2014. A rule-based parameter aided with object-based classification approach for extraction of building and roads from WorldView-2 images. *Geocarto International,* 29**,** 554-569.

ZORAN, D., WEISS, Y. & IEEE 2011. From Learning Models of Natural Image Patches to Whole Image Restoration. *2011 Ieee International Conference on Computer Vision (Iccv)***,** 479-486.

8. Appendix

8.1 OBIA procedure for data sets with 30, 10, and 4m spatial resolution from Chapter 3

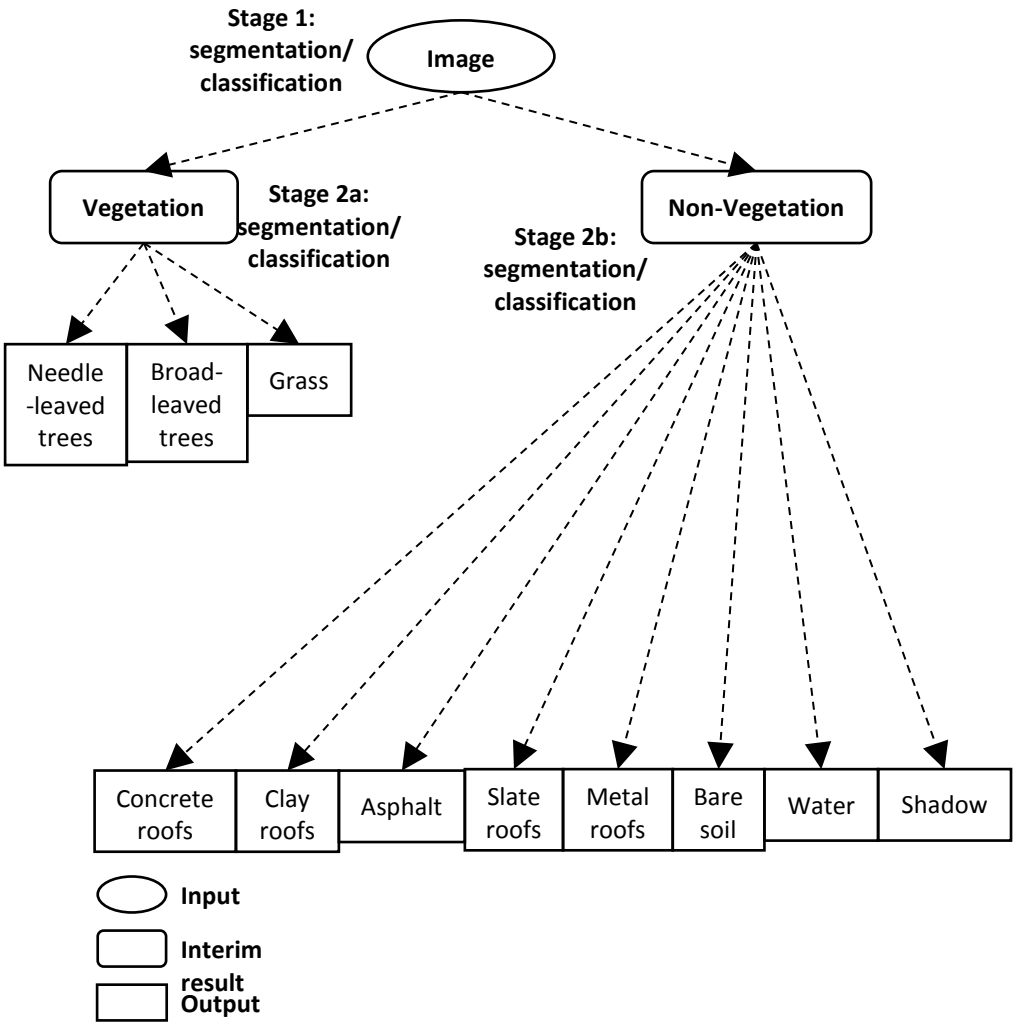


Figure A.1. Multi-stage object-based classification (OBIA) procedure for data sets with 30, 10, and 4m spatial resolution (Regarding Chapter3).

8.2 The remaining 30 error matrices from Chapter 3

Overall Accu: 83% Kappa Coeff: 0.82 Class	Reference data: Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**24**	0	0	0	0	1	4	0	0	0	0	**83**
Asphalt	0	**34**	0	0	0	1	0	1	0	1	0	**92**
Water	0	0	**43**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	1	0	0	**38**	20	1	0	0	0	0	2	**61**
Needle L trees*	0	0	0	3	**18**	0	0	0	0	0	0	**86**
Clay roof	3	1	0	0	0	**29**	0	1	0	6	1	**71**
Metal roof	0	0	0	0	0	**0**	**36**	0	0	0	0	**100**
Slate roof	0	2	0	0	0	1	0	**34**	0	2	0	**87**
Shadow	0	0	0	0	0	0	0	0	**39**	0	0	**100**
Concrete roof	1	6	0	0	0	7	0	4	0	**32**	0	**64**
Grass	1	0	0	1	0	0	0	0	1	0	**38**	**93**
Producer accuracy %	**80**	**79**	**100**	**90**	**47**	**73**	**90**	**85**	**98**	**78**	**93**	

Classification result

Table A.1. The error matrix for performing OBIA on WV-2 with 2m spatial resolution and six spectral bands.

Overall Accu: 74% Kappa Coeff: 0.72 Class	Reference data: Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**20**	2	0	0	0	4	1	0	0	2	0	**69**
Asphalt	0	**32**	0	0	0	1	1	2	0	3	0	**82**
Water	0	0	**43**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	0	0	**35**	17	0	0	0	0	0	5	**61**
Needle L trees*	0	0	0	2	**19**	0	0	0	0	0	0	**90**
Clay roof	10	1	0	1	1	**23**	1	5	1	9	1	**43**
Metal roof	0	0	0	0	0	**0**	**35**	0	0	0	0	**100**
Slate roof	0	4	0	0	0	1	0	**23**	2	1	0	**74**
Shadow	0	0	0	0	0	0	0	0	**33**	0	0	**100**
Concrete roof	0	4	0	1	0	11	0	9	0	**26**	0	**51**
Grass	0	0	0	3	1	0	2	0	4	0	**35**	**78**
Producer accuracy %	**67**	**74**	**100**	**83**	**50**	**58**	**88**	**59**	**83**	**63**	**85**	

Classification result

Table A.2. The error matrix for performing OBIA on WV-2 with 2m spatial resolution and four spectral bands.

Overall Accu: 71% Kappa Coeff: 0.68 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**18**	0	0	0	0	3	4	0	0	4	0	**62**
Asphalt	1	**33**	0	0	0	1	4	14	2	7	0	**53**
Water	0	0	**35**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	2	1	**34**	8	2	0	1	1	0	5	**63**
Needle L trees*	0	0	0	6	**26**	0	0	0	1	0	0	**79**
Clay roof	6	0	0	1	0	**22**	1	0	1	5	0	**61**
Metal roof	2	0	0	0	0	1	**31**	1	0	0	0	**89**
Slate roof	0	2	2	0	0	1	0	**21**	1	3	0	**70**
Shadow	0	0	3	0	0	1	0	1	**32**	0	0	**86**
Concrete roof	1	3	1	0	1	7	0	4	2	**22**	0	**54**
Grass	2	2	0	1	3	2	0	0	0	0	**35**	**78**
Producer accuracy %	**60**	**79**	**83**	**81**	**68**	**55**	**78**	**50**	**80**	**54**	**88**	

Table A.3. The error matrix for performing OBIA on degraded WV-2 with 4m spatial resolution and six spectral bands.

Overall Accu: 68% Kappa Coeff: 0.48 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**26**	2	0	0	0	8	7	0	0	2	0	**58**
Asphalt	0	**27**	0	0	0	2	1	3	2	2	0	**73**
Water	0	0	**34**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	1	2	**38**	18	3	0	1	2	0	16	**47**
Needle L trees*	0	1	0	3	**17**	1	0	0	1	0	0	**74**
Clay roof	1	1	0	0	1	**13**	0	0	1	5	0	**59**
Metal roof	0	0	2	0	0	0	**29**	0	0	0	0	**94**
Slate roof	1	4	1	0	0	0	0	**29**	1	3	0	**74**
Shadow	0	0	3	0	0	1	0	1	**31**	0	0	**86**
Concrete roof	0	5	0	0	0	12	1	8	1	**29**	0	**52**
Grass	2	1	0	1	2	0	2	0	1	0	**24**	**73**
Producer accuracy %	**87**	**64**	**81**	**90**	**45**	**33**	**73**	**69**	**78**	**71**	**60**	

Table A.4. The error matrix for performing OBIA on degraded WV-2 with 4m spatial resolution and four spectral bands.

Overall Accu: 49% Kappa Coeff: 0.44 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**19**	1	0	0	0	3	9	0	0	3	0	**54**
Asphalt	1	**23**	3	0	0	9	1	9	8	5	1	**38**
Water	0	0	**33**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	3	3	1	**19**	19	5	1	5	3	13	5	**25**
Needle L trees*	1	2	2	9	**17**	3	0	1	5	4	2	**37**
Clay roof	1	3	0	4	0	**18**	0	5	2	4	0	**49**
Metal roof	0	0	0	0	0	0	**24**	0	2	0	0	**92**
Slate roof	1	4	2	0	0	0	0	**14**	9	1	0	**45**
Shadow	0	0	1	0	0	0	0	0	**5**	0	0	**83**
Concrete roof	0	5	0	2	0	1	1	6	4	**11**	2	**34**
Grass	4	1	0	8	1	1	4	1	2	0	**30**	**58**
Producer accuracy %	**63**	**55**	**79**	**45**	**46**	**45**	**60**	**34**	**13**	**27**	**75**	

Table A.5. The error matrix for performing OBIA on degraded WV-2 with 10m spatial resolution and six spectral bands.

Overall Accu: 43% Kappa Coeff: 0.37 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**13**	1	0	0	0	4	4	0	2	2	1	**48**
Asphalt	0	**13**	10	0	0	2	0	10	5	2	0	**31**
Water	0	0	**26**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	1	1	1	**13**	13	0	0	0	0	0	11	**33**
Needle L trees*	3	1	1	16	**23**	5	0	3	5	7	7	**32**
Clay roof	7	5	0	7	0	**23**	4	7	8	13	0	**31**
Metal roof	1	1	1	0	0	0	**29**	0	2	0	0	**85**
Slate roof	1	0	2	0	0	0	0	**11**	4	2	0	**55**
Shadow	0	0	0	0	0	0	0	0	**6**	0	0	**100**
Concrete roof	0	16	1	3	0	5	0	8	7	**11**	2	**21**
Grass	4	4	0	3	1	1	3	2	1	4	**19**	**45**
Producer accuracy %	**43**	**31**	**62**	**31**	**62**	**58**	**73**	**27**	**15**	**27**	**48**	

Table A.6. The error matrix for performing OBIA on degraded WV-2 with 10m spatial resolution and four spectral bands.

Overall Accu: 34% Kappa Coeff: 0.27 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**12**	2	1	0	0	2	19	4	4	0	0	**27**
Asphalt	1	**5**	1	1	0	6	1	3	6	6	1	**16**
Water	0	0	**24**	0	0	0	0	0	0	0	1	**96**
Broad L trees*	4	2	4	**20**	19	8	0	5	5	5	11	**24**
Needle L trees*	3	7	8	4	**8**	5	0	3	6	6	3	**15**
Clay roof	0	7	1	2	0	**7**	1	3	2	1	0	**29**
Metal roof	2	0	0	0	0	0	**12**	0	2	1	0	**71**
Slate roof	2	4	1	2	0	0	0	**18**	9	1	0	**49**
Shadow	0	0	0	0	0	0	0	0	**0**	0	0	**0**
Concrete roof	3	15	0	6	3	12	6	5	5	**22**	4	**27**
Grass	3	0	2	7	7	0	1	0	1	1	**20**	**48**
Producer accuracy %	**40**	**12**	**57**	**48**	**22**	**18**	**30**	**44**	**0**	**51**	**50**	

Table A.7. The error matrix for performing OBIA on degraded WV-2 with 30m spatial resolution and six spectral bands.

Overall Accu: 31% Kappa Coeff: 0.24 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**12**	1	1	0	0	0	15	1	1	0	0	**39**
Asphalt	3	**8**	3	2	0	5	8	8	10	4	2	**15**
Water	0	23	**0**	0	0	0	0	0	0	0	1	**96**
Broad L trees*	2	1	1	**13**	19	5	0	4	3	6	6	**22**
Needle L trees*	3	7	10	4	**6**	6	0	3	8	6	5	**10**
Clay roof	0	7	1	3	1	**6**	1	3	4	6	2	**18**
Metal roof	0	0	0	0	0	0	**12**	0	3	0	0	**80**
Slate roof	1	3	0	0	0	0	0	**17**	6	2	0	**59**
Shadow	0	0	1	0	0	0	0	0	**0**	0	0	**0**
Concrete roof	3	14	0	7	2	17	3	5	3	**17**	1	**24**
Grass	6	1	2	13	9	1	1	0	2	2	**23**	**38**
Producer accuracy %	**40**	**19**	**55**	**31**	**16**	**15**	**30**	**41**	**0**	**40**	**58**	

Table A.8. The error matrix for performing OBIA on degraded WV-2 with 30m spatial resolution and four spectral bands.

Overall Accu: 51% Kappa Coeff: 0.46 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**17**	0	0	0	0	4	2	0	0	1	0	**71**
Asphalt	1	**21**	1	0	0	3	1	12	7	6	2	**39**
Water	0	0	**37**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	2	3	0	**21**	16	3	0	3	2	9	6	**32**
Needle L trees*	2	2	2	10	**17**	5	0	2	5	8	5	**29**
Clay roof	1	5	0	4	0	**21**	0	1	2	6	0	**53**
Metal roof	1	0	0	0	0	1	**33**	0	2	0	0	**89**
Slate roof	1	2	2	0	0	0	0	**14**	10	1	0	**47**
Shadow	0	0	0	0	0	0	0	0	**5**	0	0	**100**
Concrete roof	0	8	0	2	0	2	2	8	6	**9**	1	**24**
Grass	5	1	0	5	4	1	2	1	1	1	**26**	**55**
Producer accuracy %	**57**	**50**	**88**	**50**	**46**	**53**	**83**	**34**	**13**	**22**	**65**	

Table A.9. The error matrix for performing OBIA on degraded WV-2 with 10m spatial resolution and eight spectral bands.

Overall Accu: 35% Kappa Coeff: 0.29 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**17**	2	1	1	0	0	9	0	0	0	3	**52**
Asphalt	3	**11**	14	5	1	9	10	14	12	5	1	**13**
Water	0	1	**17**	0	0	0	1	1	4	0	2	**65**
Broad L trees*	1	0	0	**10**	6	0	0	0	1	0	2	**50**
Needle L trees*	1	3	6	5	**12**	1	0	1	2	1	6	**32**
Clay roof	2	7	1	1	5	**20**	2	5	4	10	3	**33**
Metal roof	1	0	0	0	0	0	**5**	0	1	0	0	**71**
Slate roof	1	3	0	0	0	0	0	**15**	5	2	0	**58**
Shadow	0	0	0	0	0	0	0	0	**0**	0	0	**0**
Concrete roof	0	14	1	11	1	10	1	5	9	**24**	3	**30**
Grass	4	1	2	9	12	0	12	0	2	1	**20**	**32**
Producer accuracy %	**57**	**26**	**40**	**24**	**32**	**50**	**13**	**37**	**0**	**56**	**50**	

Table A.10. The error matrix for performing OBIA on degraded WV-2 with 30m spatial resolution and eight spectral bands.

Overall Accu: 86% Kappa Coeff: 0.85 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**28**	0	0	0	0	0	0	0	0	0	0	**100**
Asphalt	2	**39**	4	0	0	1	0	2	0	0	0	**81**
Water	0	0	**32**	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	0	0	**25**	0	0	0	0	0	0	2	**93**
Needle L trees*	0	0	0	14	**25**	0	0	0	0	0	0	**64**
Clay roof	0	0	0	0	0	**35**	0	0	0	1	0	**97**
Metal roof	0	0	0	0	0	**0**	**40**	0	0	0	0	**100**
Slate roof	0	3	4	0	0	1	0	**35**	1	0	0	**80**
Shadow	0	0	3	0	13	0	0	1	**39**	0	0	**70**
Concrete roof	0	1	0	0	0	3	0	2	0	**40**	0	**87**
Grass	0	0	0	3	0	0	0	0	0	0	**39**	**93**
Producer accuracy %	**93**	**91**	**74**	**60**	**66**	**88**	**100**	**88**	**98**	**98**	**95**	

Table A.11. The error matrix for performing SVM on WV-2 with 2m spatial resolution and eight spectral bands.

Overall Accu: 83% Kappa Coeff: 0.81 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	27	0	0	0	0	0	0	0	0	0	0	**100**
Asphalt	2	38	5	0	0	1	0	2	0	1	0	**78**
Water	0	0	31	0	0	0	0	0	0	0	0	**100**
Broad L trees*	0	0	0	23	1	0	0	0	0	0	2	**88**
Needle L trees*	0	0	0	17	24	0	0	0	0	0	0	**59**
Clay roof	0	0	0	0	0	33	0	0	0	2	0	**94**
Metal roof	0	0	0	0	0	0	40	0	0	0	0	**100**
Slate roof	0	4	4	0	0	0	0	32	0	1	0	**78**
Shadow	0	0	3	0	13	0	0	1	39	0	0	**70**
Concrete roof	1	1	0	0	0	6	0	5	1	37	0	**73**
Grass	0	0	0	2	0	0	0	0	0	0	39	**95**
Producer accuracy %	**90**	**88**	**72**	**55**	**63**	**83**	**100**	**80**	**98**	**90**	**95**	

Table A.12. The error matrix for performing SVM on WV-2 with 2m spatial resolution and four spectral bands.

Overall Accu: 76% Kappa Coeff: 0.76 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**26**	0	0	0	0	1	4	0	0	0	0	**84**
Asphalt	1	**35**	6	0	0	1	1	6	1	4	0	**64**
Water	0	0	**30**	0	0	0	1	0	0	0	0	**97**
Broad L trees*	0	0	0	**23**	1	0	0	0	0	0	0	**96**
Needle L trees*	0	0	0	14	**21**	0	0	0	0	0	1	**58**
Clay roof	2	1	0	0	0	**31**	0	0	0	5	0	**79**
Metal roof	0	0	0	0	0	0	**34**	0	0	0	0	**100**
Slate roof	0	1	4	0	0	1	0	**26**	4	0	0	**72**
Shadow	0	0	2	1	16	1	0	2	**35**	0	0	**61**
Concrete roof	0	5	0	0	0	5	0	8	0	**32**	0	**64**
Grass	1	0	0	4	0	0	0	0	0	0	**39**	**89**
Producer accuracy %	**87**	**83**	**71**	**55**	**55**	**78**	**85**	**62**	**88**	**78**	**98**	

Table A.13. The error matrix for performing SVM on degraded WV-2 with 4m spatial resolution and eight spectral bands.

Overall Accu: 75% Kappa Coeff: 0.73 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**26**	0	0	0	0	1	4	0	0	0	0	**84**
Asphalt	1	**36**	9	0	0	1	1	6	1	4	0	**61**
Water	0	0	**26**	0	0	0	1	0	1	0	0	**93**
Broad L trees*	0	0	0	**21**	1	0	0	0	0	0	1	**91**
Needle L trees*	0	0	0	15	**24**	0	0	0	0	0	1	**60**
Clay roof	2	0	0	0	0	**31**	0	0	0	4	0	**84**
Metal roof	0	0	0	0	0	0	**34**	0	0	0	0	**100**
Slate roof	0	1	5	0	0	1	0	**26**	2	0	0	**74**
Shadow	0	0	2	1	13	1	0	2	**34**	0	0	**64**
Concrete roof	0	5	0	0	0	5	0	8	2	**33**	0	**62**
Grass	1	0	0	5	0	0	0	0	0	0	**38**	**86**
Producer accuracy %	**87**	**86**	**62**	**50**	**63**	**78**	**85**	**62**	**85**	**80**	**95**	

Table A.14. The error matrix for performing SVM on degraded WV-2 with 4m spatial resolution and four spectral bands.

Overall Accu: 53% Kappa Coeff: 0.48 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**20**	1	0	0	0	3	6	0	0	1	1	**63**
Asphalt	0	**26**	4	2	0	4	1	16	8	7	3	**37**
Water	0	0	**32**	0	0	0	0	0	4	0	0	**89**
Broad L trees*	0	1	0	**20**	12	0	0	2	2	5	3	**44**
Needle L trees*	2	1	1	3	**19**	2	0	1	4	7	3	**44**
Clay roof	2	0	1	2	1	**25**	0	3	0	3	2	**64**
Metal roof	0	0	0	0	0	0	**31**	0	2	0	0	**94**
Slate roof	1	3	2	0	0	0	0	**9**	8	0	0	**39**
Shadow	0	0	2	0	4	0	0	0	**5**	0	0	**45**
Concrete roof	1	9	0	9	0	6	1	9	7	**17**	1	**28**
Grass	4	1	0	6	1	0	1	1	0	1	**27**	**64**
Producer accuracy %	**67**	**62**	**76**	**48**	**51**	**63**	**78**	**22**	**13**	**41**	**68**	

Table A.15. The error matrix for performing SVM on degraded WV-2 with 10m spatial resolution and four spectral bands.

Overall Accu: 53% Kappa Coeff: 0.48 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**20**	1	0	0	0	3	6	0	0	1	1	**63**
Asphalt	0	**26**	4	2	0	4	1	16	8	6	3	**37**
Water	0	0	**32**	0	0	0	1	0	3	0	0	**89**
Broad L trees*	1	1	0	**19**	12	1	0	3	2	6	3	**40**
Needle L trees*	1	1	1	2	**19**	1	0	0	4	6	3	**50**
Clay roof	2	0	1	2	1	**26**	0	4	0	4	1	**63**
Metal roof	0	0	0	0	0	0	**30**	0	2	0	0	**94**
Slate roof	1	3	2	0	0	0	0	**9**	8	0	0	**39**
Shadow	0	0	2	0	4	0	0	0	**6**	1	0	**46**
Concrete roof	1	9	0	9	0	5	1	9	7	**16**	1	**28**
Grass	4	1	0	8	1	0	1	0	0	1	**28**	**64**
Producer accuracy %	**67**	**62**	**76**	**45**	**51**	**65**	**75**	**22**	**15**	**39**	**70**	

Table A.16. The error matrix for performing SVM on degraded WV-2 with 10m spatial resolution and six spectral bands.

Overall Accu: 52% Kappa Coeff: 0.47 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**20**	0	0	0	0	3	6	0	0	2	2	**61**
Asphalt	0	**27**	5	2	0	5	1	19	8	8	3	**35**
Water	0	0	**30**	0	0	0	1	0	3	0	0	**88**
Broad L trees*	0	1	0	**19**	12	0	0	2	2	5	2	**44**
Needle L trees*	2	1	1	3	**19**	2	0	0	4	6	4	**45**
Clay roof	1	0	0	0	0	**21**	0	0	1	0	0	**91**
Metal roof	0	0	0	0	0	0	**30**	0	2	0	0	**94**
Slate roof	1	3	3	0	0	0	0	**9**	9	1	0	**35**
Shadow	0	0	2	0	4	0	0	0	**5**	0	0	**45**
Concrete roof	2	9	1	10	0	7	1	10	6	**18**	1	**28**
Grass	4	1	0	8	2	2	1	1	0	1	**28**	**58**
Producer accuracy %	**67**	**64**	**71**	**45**	**51**	**53**	**75**	**22**	**13**	**44**	**70**	

Table A.17. The error matrix for performing SVM on degraded WV-2 with 10m spatial resolution and eight spectral bands.

Overall Accu: 33% Kappa Coeff: 0.26 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**14**	2	2	0	1	0	10	1	1	0	3	**41**
Asphalt	6	**6**	8	2	0	7	8	9	9	6	1	**10**
Water	0	2	**19**	1	0	0	0	2	5	1	1	**61**
Broad L trees*	3	1	3	**17**	18	6	0	2	4	4	14	**24**
Needle L trees*	0	0	1	0	**1**	0	0	0	1	0	0	**33**
Clay roof	1	10	1	4	2	**17**	2	2	2	13	1	**31**
Metal roof	0	0	0	0	0	0	**18**	0	3	0	0	**86**
Slate roof	1	1	2	0	0	0	0	**15**	6	4	1	**50**
Shadow	0	0	3	2	3	1	0	1	**3**	1	1	**20**
Concrete roof	1	20	3	7	2	9	2	9	5	**14**	0	**19**
Grass	4	0	0	9	10	0	0	0	1	0	**18**	**43**
Producer accuracy %	**47**	**14**	**45**	**40**	**3**	**43**	**45**	**37**	**8**	**33**	**45**	

Table A.18. The error matrix for performing SVM on degraded WV-2 with 30m spatial resolution and four spectral bands.

Overall Accu: 33% Kappa Coeff: 0.26 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**14**	1	2	0	1	0	11	1	2	0	2	**41**
Asphalt	6	**8**	7	2	0	8	7	11	12	4	1	**12**
Water	0	2	**19**	1	0	0	0	1	5	1	1	**63**
Broad L trees*	4	1	2	**17**	21	6	0	2	5	5	13	**22**
Needle L trees*	0	0	1	0	**0**	1	0	0	0	0	0	**0**
Clay roof	1	12	2	5	2	**19**	2	4	2	15	1	**29**
Metal roof	0	0	0	0	0	0	**18**	0	3	0	0	**86**
Slate roof	2	2	2	0	0	0	0	**15**	3	4	1	**52**
Shadow	0	0	3	2	3	0	0	1	**2**	1	1	**15**
Concrete roof	0	16	4	7	2	6	2	6	6	**13**	0	**21**
Grass	3	0	0	8	8	0	0	0	0	0	**20**	**51**
Producer accuracy %	**47**	**47**	**47**	**47**	**47**	**47**	**47**	**47**	**47**	**47**	**47**	

Table A.19. The error matrix for performing SVM on degraded WV-2 with 30m spatial resolution and six spectral bands.

Overall Accu: 37% Kappa Coeff: 0.3 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**16**	0	2	0	1	0	9	1	1	0	1	**52**
Asphalt	4	**8**	6	3	0	5	6	8	10	5	2	**14**
Water	0	1	**20**	1	0	0	0	1	6	0	1	**67**
Broad L trees*	3	0	3	**19**	23	3	0	2	4	3	12	**26**
Needle L trees*	0	1	2	1	**1**	0	0	1	0	0	1	**14**
Clay roof	1	13	1	5	2	**22**	4	5	3	14	2	**31**
Metal roof	0	0	0	0	0	0	**21**	0	3	0	0	**88**
Slate roof	2	3	2	0	0	0	0	**16**	4	4	1	**50**
Shadow	0	0	2	2	2	1	0	2	**1**	1	0	**9**
Concrete roof	0	16	4	4	2	9	0	5	7	**16**	0	**25**
Grass	4	0	0	7	6	0	0	0	1	0	**20**	**53**
Producer accuracy %	**53**	**19**	**48**	**45**	**3**	**55**	**53**	**39**	**3**	**37**	**50**	

Table A.20. The error matrix for performing SVM on degraded WV-2 with 30m spatial resolution and eight spectral bands.

Overall Accu: 80% Kappa Coeff: 0.79 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**26**	0	0	0	0	1	3	0	0	0	0	**87**
Asphalt	1	**37**	4	0	0	0	0	2	0	1	0	**82**
Water	0	0	**31**	0	0	0	0	1	0	0	0	**97**
Broad L trees*	0	0	0	**25**	1	0	0	0	0	0	3	**86**
Needle L trees*	0	0	0	15	**25**	0	0	0	0	0	1	**61**
Clay roof	1	0	0	0	0	**33**	0	0	1	1	0	**92**
Metal roof	2	0	1	0	0	0	**37**	0	0	0	0	**93**
Slate roof	0	3	4	0	0	2	0	**36**	5	6	0	**64**
Shadow	0	0	3	0	12	0	0	0	**34**	0	0	**69**
Concrete roof	0	3	0	0	0	4	0	1	0	**33**	0	**80**
Grass	0	0	0	2	0	0	0	0	0	0	**37**	**95**
Producer accuracy %	**87**	**86**	**72**	**60**	**66**	**83**	**93**	**90**	**85**	**80**	**90**	

Table A.21. The error matrix for performing ML on WV-2 with 2m spatial resolution and four spectral bands.

Overall Accu: 83% Kappa Coeff: 0.81 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**26**	1	0	0	0	1	2	0	0	0	0	**87**
Asphalt	1	**37**	1	0	0	0	0	2	0	1	0	**88**
Water	0	1	**35**	0	0	0	0	1	0	0	0	**95**
Broad L trees*	0	0	0	**28**	1	0	0	0	0	0	4	**85**
Needle L trees*	0	0	0	12	**26**	0	0	0	0	0	0	**68**
Clay roof	1	0	0	0	0	**31**	0	0	0	1	0	**94**
Metal roof	2	0	1	0	0	0	**38**	0	0	0	0	**93**
Slate roof	0	3	3	0	0	2	0	**35**	6	4	0	**66**
Shadow	0	0	3	0	11	0	0	0	**34**	0	0	**71**
Concrete roof	0	1	0	0	0	6	0	2	0	**35**	0	**80**
Grass	0	0	0	2	0	0	0	0	0	0	**37**	**95**
Producer accuracy %	**87**	**86**	**81**	**67**	**68**	**78**	**95**	**88**	**85**	**85**	**90**	

Table A.22. The error matrix for performing ML on WV-2 with 2m spatial resolution and six spectral bands.

Overall Accu: 83% Kappa Coeff: 0.82 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**28**	0	0	0	0	1	0	0	0	0	0	**97**
Asphalt	1	**33**	4	0	0	0	0	2	0	2	0	**79**
Water	0	1	**30**	0	0	0	0	0	0	0	0	**97**
Broad L trees*	0	0	0	**28**	2	0	0	0	0	0	5	**80**
Needle L trees*	0	0	0	11	**24**	1	0	0	0	0	0	**67**
Clay roof	0	0	0	0	0	**36**	0	0	0	0	0	**100**
Metal roof	1	1	0	0	0	**0**	**40**	0	0	0	0	**95**
Slate roof	0	6	6	0	0	1	0	**36**	1	4	0	**67**
Shadow	0	0	3	0	12	0	0	0	**39**	0	0	**72**
Concrete roof	0	2	0	0	0	1	0	2	0	**35**	0	**88**
Grass	0	0	0	3	0	0	0	0	0	0	**36**	**92**
Producer accuracy %	**93**	**77**	**70**	**67**	**63**	**90**	**100**	**90**	**98**	**85**	**88**	

Table A.23. The error matrix for performing ML on WV-2 with 2m spatial resolution and eight spectral bands.

Overall Accu: 74% Kappa Coeff: 0.72 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**28**	0	0	0	0	2	1	0	0	2	1	**82**
Asphalt	1	**34**	1	0	0	0	1	4	1	6	0	**71**
Water	0	2	**34**	0	0	0	1	2	3	0	0	**81**
Broad L trees*	0	0	0	**18**	2	0	0	0	0	0	1	**86**
Needle L trees*	0	0	0	20	**24**	0	0	0	0	1	1	**52**
Clay roof	0	1	0	0	0	**31**	0	0	0	3	0	**89**
Metal roof	0	0	2	0	0	1	**37**	0	0	0	0	**93**
Slate roof	0	3	2	0	0	1	0	**30**	6	1	0	**70**
Shadow	0	0	2	0	12	0	0	0	**23**	0	0	**62**
Concrete roof	0	2	1	0	0	5	0	6	7	**28**	0	**57**
Grass	1	0	0	4	0	0	0	0	0	0	**37**	**88**
Producer accuracy %	**93**	**81**	**81**	**43**	**63**	**78**	**93**	**71**	**58**	**68**	**93**	

Table A.24. The error matrix for performing ML on degraded WV-2 with 4m spatial resolution and four spectral bands.

Overall Accu: 74% Kappa Coeff: 0.73 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**26**	0	0	0	0	1	2	0	0	1	1	**84**
Asphalt	1	**31**	1	0	0	1	0	3	0	6	0	**72**
Water	0	3	**35**	0	0	0	2	2	4	0	0	**76**
Broad L trees*	0	0	0	**21**	2	0	0	0	0	1	4	**75**
Needle L trees*	0	0	0	16	**25**	0	0	0	0	0	0	**61**
Clay roof	0	2	0	0	0	**29**	0	0	0	1	0	**91**
Metal roof	2	0	2	0	0	3	**36**	0	0	0	0	**84**
Slate roof	0	2	2	0	0	1	0	**29**	6	0	0	**73**
Shadow	0	0	2	0	11	0	0	0	**23**	0	0	**64**
Concrete roof	0	4	0	0	0	5	0	8	7	**32**	0	**57**
Grass	1	0	0	5	0	0	0	0	0	0	**35**	**85**
Producer accuracy %	**87**	**74**	**83**	**50**	**66**	**73**	**90**	**69**	**58**	**78**	**88**	

Table A.25. The error matrix for performing ML on degraded WV-2 with 4m spatial resolution and six spectral bands.

Overall Accu: 74% Kappa Coeff: 0.71 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**27**	1	0	0	0	3	2	0	0	1	1	**77**
Asphalt	1	**32**	4	0	0	1	1	3	0	5	0	**68**
Water	0	2	**32**	0	0	0	1	1	2	0	0	**84**
Broad L trees*	0	0	0	**24**	2	0	0	0	0	1	6	**73**
Needle L trees*	0	0	0	13	**24**	0	0	0	0	0	0	**65**
Clay roof	0	0	0	1	0	**30**	0	0	0	4	0	**86**
Metal roof	1	0	1	0	0	0	**36**	0	0	0	0	**95**
Slate roof	0	1	2	0	0	0	0	**24**	2	0	0	**83**
Shadow	0	0	2	0	12	1	0	1	**31**	0	0	**66**
Concrete roof	0	6	1	0	0	5	0	13	5	**30**	0	**50**
Grass	1	0	0	4	0	0	0	0	0	0	**33**	**87**
Producer accuracy %	**90**	**76**	**76**	**57**	**63**	**75**	**90**	**57**	**78**	**73**	**83**	

Table A.26. The error matrix for performing ML on degraded WV-2 with 4m spatial resolution and eight spectral bands.

Overall Accu: 55% Kappa Coeff: 0.5 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**21**	1	0	1	0	5	4	3	2	6	3	**46**
Asphalt	1	**28**	4	5	0	4	1	13	6	5	3	**40**
Water	0	0	**35**	0	0	0	0	0	5	0	0	**88**
Broad L trees*	1	0	0	**15**	12	0	0	0	1	5	3	**41**
Needle L trees*	1	2	2	7	**20**	0	0	2	5	3	3	**44**
Clay roof	3	0	0	3	1	**28**	0	5	2	7	1	**56**
Metal roof	0	1	0	0	0	0	**35**	0	4	0	0	**88**
Slate roof	1	4	0	0	0	2	0	**13**	8	2	0	**43**
Shadow	0	0	1	0	2	0	0	0	**3**	0	0	**50**
Concrete roof	0	5	0	2	0	1	0	4	4	**12**	0	**43**
Grass	2	1	0	9	2	0	0	1	0	1	**27**	**63**
Producer accuracy %	**70**	**67**	**83**	**36**	**54**	**70**	**88**	**32**	**8**	**29**	**68**	

Table A.27. The error matrix for performing ML on degraded WV-2 with 10m spatial resolution and six spectral bands.

Overall Accu: 55% Kappa Coeff: 0.5 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**21**	1	0	1	0	5	4	3	2	6	3	**46**
Asphalt	1	**28**	4	5	0	4	1	13	6	5	3	**40**
Water	0	0	**35**	0	0	0	0	0	5	0	0	**88**
Broad L trees*	1	0	0	**15**	12	0	0	0	1	5	3	**41**
Needle L trees*	1	2	2	7	**20**	0	0	2	5	3	3	**44**
Clay roof	3	0	0	3	1	**28**	0	5	2	7	1	**56**
Metal roof	0	1	0	0	0	0	**35**	0	4	0	0	**88**
Slate roof	1	4	0	0	0	2	0	**13**	8	2	0	**43**
Shadow	0	0	1	0	2	0	0	0	**3**	0	0	**50**
Concrete roof	0	5	0	2	0	1	0	4	4	**12**	0	**43**
Grass	2	1	0	9	2	0	0	1	0	1	**27**	**63**
Producer accuracy %	**70**	**67**	**83**	**36**	**54**	**70**	**88**	**32**	**8**	**29**	**68**	

Table A.28. The error matrix for performing ML on degraded WV-2 with 10m spatial resolution and eight spectral bands.

Overall Accu: 37% Kappa Coeff: 0.31 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**17**	5	4	1	9	5	7	0	4	5	4	**28**
Asphalt	3	**12**	5	7	0	4	5	7	10	7	2	**19**
Water	0	0	**21**	0	0	0	0	0	4	1	2	**75**
Broad L trees*	2	1	1	**14**	9	2	0	0	0	2	11	**33**
Needle L trees*	0	0	1	1	**4**	1	0	1	5	0	1	**29**
Clay roof	0	8	0	4	2	**22**	2	9	3	6	3	**37**
Metal roof	2	0	2	0	0	0	**25**	2	9	0	0	**63**
Slate roof	2	4	0	0	0	0	0	**15**	2	2	0	**60**
Shadow	2	1	6	5	4	0	0	4	**0**	3	0	**0**
Concrete roof	1	11	2	2	1	6	1	3	3	**16**	0	**35**
Grass	1	0	0	8	8	0	0	0	0	1	**17**	**49**
Producer accuracy %	**57**	**29**	**50**	**33**	**11**	**55**	**63**	**37**	**0**	**37**	**43**	

Table A.29. The error matrix for performing ML on degraded WV-2 with 30m spatial resolution and six spectral bands.

Overall Accu: 35% Kappa Coeff: 0.28 Class	Bare soil	Asphalt	Water	Broad L trees *	Needle L tree*	Clay roof	Metal roof	Slate roof	Shadow	Concrete roof	Grass	User accuracy %
Bare soil	**19**	4	3	4	11	4	6	1	6	7	9	**26**
Asphalt	3	**20**	12	7	2	6	5	17	19	21	3	**17**
Water	0	0	**21**	0	0	0	0	0	3	0	1	**84**
Broad L trees*	3	1	1	**16**	14	2	0	1	2	2	9	**31**
Needle L trees*	0	0	1	0	**1**	1	0	1	0	0	1	**20**
Clay roof	0	5	1	4	3	**18**	0	8	3	7	3	**35**
Metal roof	3	1	1	1	0	4	**29**	2	4	0	0	**64**
Slate roof	0	1	0	0	0	0	0	**11**	2	0	0	**79**
Shadow	0	1	2	2	1	0	0	0	**0**	2	0	**0**
Concrete roof	0	9	0	2	1	4	0	0	1	**3**	0	**15**
Grass	2	0	0	6	4	1	0	0	0	1	**14**	**50**
Producer accuracy %	**63**	**48**	**50**	**38**	**3**	**45**	**73**	**27**	**0**	**7**	**35**	

Table A.30. The error matrix for performing ML on degraded WV-2 with 30m spatial resolution and eight spectral bands.

8.3 The remaining four error matrices from Chapter 5

Overall Accuracy: 56% Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**7**	1	0	0	0	7	0	0	**47**
Asphalt	4	**9**	0	0	7	5	1	1	**33**
Water	0	1	**35**	0	0	0	0	0	**97**
Trees	0	1	1	**22**	1	0	0	2	**81**
Roof tiles	6	25	5	11	**32**	0	16	5	**32**
Metal roof	4	0	0	0	0	**28**	0	0	**88**
Slate roof	1	4	1	0	2	0	**22**	0	**73**
Grass	8	2	1	15	7	0	1	**32**	**48**
Producer accuracy %	**23**	**21**	**81**	**46**	**65**	**70**	**55**	**80**	

Table A.31. The error matrix of performing SVM on Pan-sharpened Landsat OLI of Nottingham with 15m spatial resolution.

Overall Accuracy: 58% Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**8**	1	0	0	0	6	0	0	**53**
Asphalt	7	**10**	0	0	11	5	3	3	**26**
Water	0	1	**39**	0	0	0	0	0	**98**
Trees	0	1	0	**21**	1	0	0	2	**84**
Roof tiles	4	24	2	12	**33**	1	15	3	**35**
Metal roof	4	0	0	0	0	**28**	0	0	**88**
Slate roof	2	4	1	0	2	0	**21**	0	**70**
Grass	5	2	1	15	2	0	1	**32**	**55**
Producer accuracy %	**27**	**23**	**91**	**44**	**67**	**70**	**53**	**80**	

Table A.32. The error matrix of performing OBIA\SVM on Pan-sharpened Landsat OLI of Nottingham with 15m spatial resolution.

Overall Accuracy: 54% Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**16**	2	2	0	0	9	1	4	**47**
Asphalt	2	**10**	0	1	3	4	3	0	**43**
Water	0	0	**31**	0	0	0	0	0	**100**
Trees	1	1	4	**22**	2	0	0	9	**56**
Roof tiles	2	25	2	12	**38**	0	16	6	**38**
Metal roof	2	0	0	0	0	**25**	0	0	**93**
Slate roof	1	5	2	0	2	0	**18**	1	**62**
Grass	5	1	0	13	5	2	2	**20**	**42**
Producer accuracy %	**55**	**23**	**76**	**46**	**76**	**63**	**45**	**50**	

Table A.33. The error matrix of performing SVM on Landsat OLI of Nottingham with 30m spatial resolution.

Overall Accuracy: 54% Class	Bare soil	Asphalt	Water	Trees	Roof tiles	Metal roof	Slate roof	Grass	User accuracy %
Bare soil	**16**	2	2	0	0	9	1	4	**47**
Asphalt	2	**9**	0	1	3	4	3	0	**41**
Water	0	0	**31**	0	0	0	0	0	**100**
Trees	1	1	4	**22**	2	0	0	9	**56**
Roof tiles	2	26	2	12	**38**	0	16	6	**37**
Metal roof	2	0	0	0	0	**25**	0	0	**93**
Slate roof	1	5	2	0	2	0	**18**	1	**62**
Grass	5	1	0	13	5	2	2	**20**	**42**
Producer accuracy %	**55**	**20**	**76**	**46**	**76**	**63**	**45**	**50**	

Table A.34. The error matrix of performing OBIA\SVM on Landsat OLI of Nottingham with 30m spatial resolution.

8.4 List of all satellite images

8.4.1 Images were used in the chapter 3

No	Name	Location	Acquisition date	Spatial resolution	Spectral bands	Geographical scale
1	WV-2	Nottingham	26-June 2014	2m	Costal blue, Blue, Green, Yellow, Red, Red edge, NIR1, and NIR2	11 x 11 km
2	WV-2	Nottingham	26-June 2014	2m	Blue, Green, Red, Red edge, NIR1, and NIR2	11 x 11 km
3	WV-2	Nottingham	26-June 2014	2m	Blue, Green, Red, and NIR	11 x 11 km
4	Simulated IKONOS by degrading the WV-2	Nottingham	26-June 2014	4m	Blue, Green, Red, Red edge, NIR1, and NIR2	11 x 11 km
5	Simulated IKONOS by degrading the WV-2	Nottingham	26-June 2014	4m	Blue, Green, Red, Red edge, NIR1, and NIR2	11 x 11 km
6	Simulated IKONOS by degrading the WV-2	Nottingham	26-June 2014	4m	Blue, Green, Red, and NIR	11 x 11 km
7	Simulated SPOT by degrading the WV-2	Nottingham	26-June 2014	10m	Costal blue, Blue, Green, Yellow, Red, Red edge, NIR1, and NIR2	11 x 11 km
8	Simulated SPOT by degrading the WV-2	Nottingham	26-June 2014	10m	Blue, Green, Red, Red edge, NIR1, and NIR2	11 x 11 km
9	Simulated SPOT by degrading the WV-2	Nottingham	26-June 2014	10m	Blue, Green, Red, and NIR	11 x 11 km

10	Simulated Landsat by degrading the WV-2	Nottingham	26-June 2014	30m	Costal blue, Blue, Green, Yellow, Red, Red edge, NIR1, and NIR2	11 x 11 km
11	Simulated Landsat by degrading the WV-2	Nottingham	26-June 2014	30m	Blue, Green, Red, Red edge, NIR1, and NIR2	11 x 11 km
12	Simulated Landsat by degrading the WV-2	Nottingham	26-June 2014	30m	Blue, Green, Red, and NIR	11 x 11 km

8.4.2 Images were used in the chapter 5

No	Name	Location	Acquisition date	Spatial resolution	Spectral bands	Geographical scale
1	WV-2	Nottingham	26-June 2014	2m	Blue, Green, Red, and NIR	11 x 11 km
2	Landsat OLI	Nottingham	15-Jully 2013	30m	Blue, Green, Red, and NIR	11 x 11 km

8.4.3 Images were used in the chapter 6

No	Name	Location	Acquisition date	Spatial resolution	Spectral bands	Geographical scale
1	WV-2	Nottingham	26-June 2014	2m	Blue, Green, Red, and NIR	11 x 11 km
2	Landsat ETM+	Nottingham	16-May 2000	30m	Blue, Green, Red, and NIR	11 x 11 km
3	Landsat OLI	Derby	15-Jully 2013	30m	Blue, Green, Red, and NIR	12 x 14 km

8.5 MATLAB scripts

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Demo Codes For Image Super-resolution via Sparse Representation
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Reference

J. Yang et al. Image super-resolution as sparse representation of raw image patches. CVPR 2008.

J. Yang et al. Image super-resolution via sparse representation. IEEE Transactions on Image Processing, Vol 19, Issue 11, pp2861-2873, 2010

For any problems, send email to jyang29@uiuc.edu

```
============================================================
```

Demo_SR.m: demo code for image super-resolution via sparse recovery

1. The demo code is for upscaling factor of 2. For larger magnification factors, run the function "ScSR.m" multiple times. Note the code is a little different from what presented in the TIP10 paper. Please find the previous codes and results in folder "Previous".

2. Two pre-trained dictionaries are provided in directory "Dictionary". The dictionaries are for zoom factor of 2. You can train your own dictionary based on function "Demo_Dictionary_Training.m" talked below.

==========
Demo_Dictionary_Training.m: demo code for training the dictionary

1. If you want to train your own dictionary, replace the training images in subfolder "Data/Training" by yours.

2. You need to inspect the statistics of your sampled patches to prune those smooth patches.

Demo dictionary training

```
Demo codes for dictionary training by joint sparse coding
%
% Reference
%   J. Yang et al. Image super-resolution as sparse representation of raw
%   image patches. CVPR 2008.
%   J. Yang et al. Image super-resolution via sparse representation. IEEE
%   Transactions on Image Processing, Vol 19, Issue 11, pp2861-2873, 2010
%
% Jianchao Yang
% ECE Department, University of Illinois at Urbana-Champaign
% For any questions, send email to jyang29@uiuc.edu
%
==========

clear; clc; close all;
addpath(genpath('RegularizedSC'));

TR_IMG_PATH = 'D:\SRM\Imagery\Study Area-2m 4bands\Band 4 16bit';

dict_size   = 2048;         % dictionary size
lambda      = 0.15;         % sparsity regularization
patch_size  = 5;            % image patch size
nSmp        = 100000;       % number of patches to sample
upscale     = 7.5;          % upscaling factor

% randomly sample image patches
[Xh, Xl] = rnd_smp_patch(TR_IMG_PATH, '*.tif', patch_size, nSmp, upscale);

% prune patches with smaclll variances, threshould chosen based on the
% training data
[Xh, Xl] = patch_pruning(Xh, Xl, 0.005);
```

```
% joint sparse coding
[Dh, Dl] = train_coupled_dict(Xh, Xl, dict_size, lambda,upscale);
dict_path = ['Dictionary/D_' num2str(dict_size) '_' num2str(lambda) '_'
num2str(patch_size) '.mat' ];
save(dict_path, 'Dh', 'Dl');
```

Demo sparse representation

```
% Simple demo codes for image super-resolution via sparse representation
%
% Reference
%   J. Yang et al. Image super-resolution as sparse representation of raw
%   image patches. CVPR 2008.
%   J. Yang et al. Image super-resolution via sparse representation. IEEE
%   Transactions on Image Processing, Vol 19, Issue 11, pp2861-2873, 2010
%
% Jianchao Yang
% ECE Department, University of Illinois at Urbana-Champaign
% For any questions, send email to jyang29@uiuc.edu
%
============================================================

clear all; clc;

% path='C:\Users\lgxxf\Desktop\TM25scenes_UK\test\';
% outpath='C:\Users\lgxxf\Desktop\TM25scenes_UK\test\output\';
% filelist=dir(fullfile(path,'*.tif'));
% n=length(filelist);
% for tt=1:n

% read test image
%fname=F:\SRR\Imagery\Input LR Imagery\Degraded 2m to 30m\Band
2\Study area 30m degraded by Envi band 3;
% fname=strcat(path,filelist(tt).name);
im_l = imread('D:\SRM\Result\Pan-sharpenning of Landsat SW 15.7\HCS
Mege Resolution\Study Area\Band 3\HCS Band 3.tif');

% set parameters
lambda = 0.15;              % sparsity regularization
overlap = 1;               % the more overlap the better (patch size 5x5)
up_scale = 7.5;              % scaling factor, depending on the trained
dictionary
maxIter = 20;              % if 0, do not use backprojection
```

```
% load dictionary
load('D:\SRM\Matlab Codes\Matlab Codes
ScSR\Dictionary\D_2048_0.15_5.mat');

% change color space, work on illuminance only
%im_l_ycbcr = rgb2ycbcr(im_l);
%im_l_y = im_l_ycbcr(:, :, 1);
% im_l_cb = im_l_ycbcr(:, :, 2);
% im_l_cr = im_l_ycbcr(:, :, 3);

% image super-resolution based on sparse representation
[im_h] = ScSR(im_l, up_scale, Dh, Dl, lambda, overlap);
%[im_h_y] = backprojection(im_h_y, im_l, maxIter);

% upscale the chrominance simply by "bicubic"
% [nrow, ncol] = size(im_h);
% im_h_cb = imresize(im_l_cb, [nrow, ncol], 'bicubic');
% im_h_cr = imresize(im_l_cr, [nrow, ncol], 'bicubic');

% im_h_ycbcr = zeros([nrow, ncol, 3]);
% im_h_ycbcr(:, :, 1) = im_h_y;
% im_h_ycbcr(:, :, 2) = im_h_cb;
% im_h_ycbcr(:, :, 3) = im_h_cr;
% im_h = ycbcr2rgb(uint8(im_h_ycbcr));

% bicubic interpolation for reference
% im_b = imresize(im_l, [nrow, ncol], 'bicubic');

% read ground truth image
%im =
imread('C:\Users\lgxxf\desktop\ScSR\Data\Testing\p203r023_7dk20020911_
z30_61.tif');

% compute PSNR for the illuminance channel
%bb_rmse = compute_rmse(im, im_b);
%sp_rmse = compute_rmse(im, im_h);

%bb_psnr = 20*log10(255/bb_rmse);
%sp_psnr = 20*log10(255/sp_rmse);

% fprintf('PSNR for Bicubic Interpolation: %f dB\n', bb_psnr);
% fprintf('PSNR for Sparse Representation Recovery: %f dB\n', sp_psnr);

% show the images
% figure, imshow(im_h);
% title('Sparse Recovery');
% figure, imshow(im_b);
```

```
% title('Bicubic Interpolation');

imwrite2tif(im_h,[],'D:\SRM\Result\Sparse-representation super-resolution\Pan-sharpening + SR\Using WV-2 with 16 bit as training\Using HCS Resolution Merge by Erdas and Study Area 2m WV-2 16bit as training image with zoom factor 7.5\band 3.tif','single');
% imwrite2tif(im_h,[],strcat(outpath,filelist(tt).name,'50m.tif'),'single');
% end
```

Dist RMSE

```
clear;
clc;

[num1,txt1,raw1]=xlsread('C:\Users\lgxxf\Downloads\OS Street View TQ\boundary points\b9.xlsx');
[num2,txt2,raw2]=xlsread('C:\Users\lgxxf\Downloads\OS Street View TQ\boundary points\b9_USGS.xlsx');
sum=0;
[row,col] = size(num1);

for i=1:row
   dist_squ(i)=(num1(i,1)-num2(i,1))^2+(num1(i,2)-num2(i,2))^2;
   dist(i)=sqrt(dist_squ(i));


end

m=mean(dist);

sum=0;
for i=1:row
   sum=sum+(dist(i)-m)^2;


end
 stdev=sqrt(sum/row);
stdeor=stdev/sqrt(row);
```

Back projection

```
function [im_h] = backprojection(im_h, im_l, maxIter)

[row_l, col_l] = size(im_l);
[row_h, col_h] = size(im_h);

p = fspecial('gaussian', 5, 1);
p = p.^2;
```

```
p = p./sum(p(:));

im_l = double(im_l);
im_h = double(im_h);

for ii = 1:maxIter,
   im_l_s = imresize(im_h, [row_l, col_l], 'bicubic');
   im_diff = im_l - im_l_s;

   im_diff = imresize(im_diff, [row_h, col_h], 'bicubic');
   im_h = im_h + conv2(im_diff, p, 'same');
end
```

Compute RMSE

```
function [rmse] = compute_rmse(im1, im2)

[row,col]=size(im1);

imdff =(im1 - im2).^2;
% sum1=0;
total=sum(sum(imdff));
% for i=1:col
%
%       sum1=sum1+total(i);
%
% end

rmse = sqrt(total/(row*col));
```

Sample patches

```
function [HP, LP] = sample_patches(im, patch_size, patch_num, upscale)

if size(im, 3) == 3,
   hIm = rgb2gray(im);
else
   hIm = im;
end

% generate low resolution counter parts
lIm = imresize(hIm, 1/upscale, 'bicubic');
lIm = imresize(lIm, size(hIm), 'bicubic');
[nrow, ncol] = size(hIm);

x = randperm(nrow-2*patch_size-1) + patch_size;
y = randperm(ncol-2*patch_size-1) + patch_size;
```

```
[X,Y] = meshgrid(x,y);

xrow = X(:);
ycol = Y(:);

if patch_num < length(xrow),
   xrow = xrow(1:patch_num);
   ycol = ycol(1:patch_num);
end

patch_num = length(xrow);

hIm = double(hIm);
lIm = double(lIm);

%H = zeros(patch_size^2,     length(xrow));
%L = zeros(4*patch_size^2,   length(xrow));

% compute the first and second order gradients
hf1 = [-1,0,1];
vf1 = [-1,0,1]';

lImG11 = conv2(lIm, hf1,'same');
lImG12 = conv2(lIm, vf1,'same');

hf2 = [1,0,-2,0,1];
vf2 = [1,0,-2,0,1]';

lImG21 = conv2(lIm,hf2,'same');
lImG22 = conv2(lIm,vf2,'same');

for ii = 1:patch_num,
   row = xrow(ii);
   col = ycol(ii);

   Hpatch = hIm(row:row+patch_size-1,col:col+patch_size-1);

   Lpatch1 = lImG11(row:row+patch_size-1,col:col+patch_size-1);
   Lpatch2 = lImG12(row:row+patch_size-1,col:col+patch_size-1);
   Lpatch3 = lImG21(row:row+patch_size-1,col:col+patch_size-1);
   Lpatch4 = lImG22(row:row+patch_size-1,col:col+patch_size-1);

   Lpatch = [Lpatch1(:),Lpatch2(:),Lpatch3(:),Lpatch4(:)];
   Lpatch = Lpatch(:);

   HP(:,ii) = Hpatch(:)-mean(Hpatch(:));
```

```
   LP(:,ii) = Lpatch;
End
```

RND SMP Patch

```
function [Xh, Xl] = rnd_smp_patch(img_path, type, patch_size, num_patch,
upscale)

img_dir = dir(fullfile(img_path, type));

Xh = [];
Xl = [];

img_num = length(img_dir);
nper_img = zeros(1, img_num);

for ii = 1:length(img_dir),
   im = imread(fullfile(img_path, img_dir(ii).name));
   nper_img(ii) = prod(size(im));
end

nper_img = floor(nper_img*num_patch/sum(nper_img));

for ii = 1:img_num,
   patch_num = nper_img(ii);
   im = imread(fullfile(img_path, img_dir(ii).name));
   [H, L] = sample_patches(im, patch_size, patch_num, upscale);
   Xh = [Xh, H];
   Xl = [Xl, L];
end

patch_path = ['Training/rnd_patches_' num2str(patch_size) '_'
num2str(num_patch) '_s' num2str(upscale) '.mat'];
save(patch_path, 'Xh', 'Xl');
```

Training coupled dict

```
function [Dh, Dl] = train_coupled_dict(Xh, Xl, dict_size, lambda, upscale)

addpath(genpath('RegularizedSC'));

hDim = size(Xh, 1);
lDim = size(Xl, 1);

% should pre-normalize Xh and Xl !
hNorm = sqrt(sum(Xh.^2));
lNorm = sqrt(sum(Xl.^2));
Idx = find( hNorm & lNorm );
```

```
Xh = Xh(:, Idx);
Xl = Xl(:, Idx);

Xh = Xh./repmat(sqrt(sum(Xh.^2)), size(Xh, 1), 1);
Xl = Xl./repmat(sqrt(sum(Xl.^2)), size(Xl, 1), 1);

% joint learning of the dictionary
X = [sqrt(hDim)*Xh; sqrt(lDim)*Xl];
Xnorm = sqrt(sum(X.^2, 1));

clear Xh Xl;

X = X(:, Xnorm > 1e-5);
X = X./repmat(sqrt(sum(X.^2, 1)), hDim+lDim, 1);

idx = randperm(size(X, 2));

% dictionary training
[D] = reg_sparse_coding(X, dict_size, [], 0, lambda, 40);

Dh = D(1:hDim, :);
Dl = D(hDim+1:end, :);

% normalize the dictionary
% Dh = Dh./repmat(sqrt(sum(Dh.^2, 1)), hDim, 1);
% Dl = Dl./repmat(sqrt(sum(Dl.^2, 1)), lDim, 1);

patch_size = sqrt(size(Dh, 1));

dict_path = ['Dictionary/D_' num2str(dict_size) '_' num2str(lambda) '_'
num2str(patch_size) '_s' num2str(upscale) '.mat' ];
save(dict_path, 'Dh', 'Dl');
```

ScSR

```
function [hIm] = ScSR(lIm, up_scale, Dh, Dl, lambda, overlap)

% normalize the dictionary
norm_Dl = sqrt(sum(Dl.^2, 1));
Dl = Dl./repmat(norm_Dl, size(Dl, 1), 1);

patch_size = sqrt(size(Dh, 1));

% bicubic interpolation of the low-resolution image
mIm = single(imresize(lIm, up_scale, 'bicubic'));
```

```
hIm = zeros(size(mIm));
cntMat = zeros(size(mIm));

[h, w] = size(mIm);

% extract low-resolution image features
lImfea = extr_lIm_fea(mIm);

% patch indexes for sparse recovery (avoid boundary)
gridx = 3:patch_size - overlap : w-patch_size-2;
gridx = [gridx, w-patch_size-2];
gridy = 3:patch_size - overlap : h-patch_size-2;
gridy = [gridy, h-patch_size-2];

A = Dl'*Dl;
cnt = 0;

% loop to recover each low-resolution patch
for ii = 1:length(gridx),
   for jj = 1:length(gridy),

      cnt = cnt+1;
      xx = gridx(ii);
      yy = gridy(jj);

      mPatch = mIm(yy:yy+patch_size-1, xx:xx+patch_size-1);
      mMean = mean(mPatch(:));
      mPatch = mPatch(:) - mMean;
      mNorm = sqrt(sum(mPatch.^2));

      mPatchFea = lImfea(yy:yy+patch_size-1, xx:xx+patch_size-1, :);
      mPatchFea = mPatchFea(:);
      mfNorm = sqrt(sum(mPatchFea.^2));

      if mfNorm > 1,
         y = mPatchFea./mfNorm;
      else
         y = mPatchFea;
      end

      b = -Dl'*y;

      % sparse recovery
      w = L1QP_FeatureSign_yang(lambda, A, b);

      % generate the high resolution patch and scale the contrast
      hPatch = Dh*w;
```

```
        hPatch = lin_scale(hPatch, mNorm);

        hPatch = reshape(hPatch, [patch_size, patch_size]);
        hPatch = hPatch + mMean;

        hIm(yy:yy+patch_size-1, xx:xx+patch_size-1) = hIm(yy:yy+patch_size-1,
xx:xx+patch_size-1) + hPatch;
        cntMat(yy:yy+patch_size-1, xx:xx+patch_size-1) =
cntMat(yy:yy+patch_size-1, xx:xx+patch_size-1) + 1;
    end
end

% fill in the empty with bicubic interpolation
idx = (cntMat < 1);
hIm(idx) = mIm(idx);

cntMat(idx) = 1;
hIm = hIm./cntMat;
%hIm = uint8(hIm);
```

Patch pruning

```
function [Xh, Xl] = patch_pruning(Xh, Xl, threshold)

pvars = var(Xh, 0, 1);

idx = pvars > threshold;

Xh = Xh(:, idx);
Xl = Xl(:, idx);
```

Lin scale

```
function [xh] = lin_scale( xh, mNorm )

hNorm = sqrt(sum(xh.^2));

if hNorm,
    s = mNorm*1.2/hNorm;
    xh = xh.*s;
end
```

L1QP FeatureSign Yang

```
%% L1QP_FeatureSign solves nonnegative quadradic programming
```

```
%% using Feature Sign.
%%
%%    min  0.5*x'*A*x+b'*x+\lambda*|x|
%%
%% [net,control]=NNQP_FeatureSign(net,A,b,control)
%%
%%
%%

function [x]=L1QP_FeatureSign_yang(lambda,A,b)

A = double(A);
b = double(b);

EPS = 1e-9;
x=zeros(size(A, 1), 1);          %coeff

grad=A*sparse(x)+b;
[ma mi]=max(abs(grad).*(x==0));

while true,


 if grad(mi)>lambda+EPS,
  x(mi)=(lambda-grad(mi))/A(mi,mi);
 elseif grad(mi)<-lambda-EPS,
  x(mi)=(-lambda-grad(mi))/A(mi,mi);
 else
  if all(x==0)
   break;
  end
 end

 while true,
  a=x~=0;   %active set
  Aa=A(a,a);
  ba=b(a);
  xa=x(a);

  %new b based on unchanged sign
  vect = -lambda*sign(xa)-ba;
  x_new= Aa\vect;
  idx = find(x_new);
  o_new=(vect(idx)/2 + ba(idx))'*x_new(idx) + lambda*sum(abs(x_new(idx)));

  %cost based on changing sign
  s=find(xa.*x_new<=0);
```

```
   if isempty(s)
     x(a)=x_new;
     loss=o_new;
     break;
   end
   x_min=x_new;
   o_min=o_new;
   d=x_new-xa;
   t=d./xa;
   for zd=s',
     x_s=xa-d/t(zd);
     x_s(zd)=0;  %make sure it's zero
%      o_s=L1QP_loss(net,Aa,ba,x_s);
     idx = find(x_s);
     o_s = (Aa(idx, idx)*x_s(idx)/2 +
ba(idx))'*x_s(idx)+lambda*sum(abs(x_s(idx)));
     if o_s<o_min,
       x_min=x_s;
       o_min=o_s;
     end
   end

   x(a)=x_min;
   loss=o_min;
  end

  grad=A*sparse(x)+b;

  [ma mi]=max(abs(grad).*(x==0));
  if ma <= lambda+EPS,
    break;
  end
end
```

Imwrite2tif

```
function imwrite2tif(varargin)
% IMWRITE2TIF Write image to tif file with specified datatype.
%   IMWRITE2TIF(IMGDATA,HEADER,IMFILE,DATATYPE) exports IMGDATA
with HEADER
%   to TIF file named IMFILE. HEADER is usally obtained by IMFINFO from
%   original image file, and it can also be left empty. String DATATYPE
%   specifies data type for the export. Supported data types include
%   logical, uint8, int8, uint16, int16, uint32, int32, uint64, int64,
%   single and double.
%
```

```
%   IMWRITE2TIF(IMGDATA,HEADER,IMFILE,DATATYPE,TAG NAME1,TAG
VALUE1,TAG NAME2,
%   TAG VALUE2,...) writes with specified Matlab supported TIF tag values.
%   These new tag values overide those already defined in HEADER.
%
%   Note 1:
%      to avoid errors such as '??? Error using ==> tifflib The value for
%      MaxSampleValue must be ...', overide tag MaxSampleValue by Matlab
%      supported values. Or simply remove the tag from HEADER.
%
%   Note 2:
%      Overwriting of the existing image files is not checked. Be cautious
%      with the export image file name.
%
%   Example 1:
%      imgdata = imread('ngc6543a.jpg');
%      header  = imfinfo('ngc6543a.jpg');
%      imwrite2tif(imgdata,header,'new_ngc6543a.tif','uint8');
%
%   Example 2:
%      imgdata = imread('mri.tif');
%      imwrite2tif(imgdata,[],'new_mri.tif','int32','Copyright','MRI',
%      'Compression',1);
%
%   More information can be found by searching for 'Exporting Image Data
%   and Metadata to TIFF Files' in Matlab Help.

%   Zhang Jiang
%   $Revision: 1.0 $  $Date: 2011/02/23 $

if nargin<4 || mod(nargin,2)==1
   error('Invalid number of input arguments.');
end

% assign input argument
imgdata  = varargin{1};
header   = varargin{2};
imfile   = varargin{3};
datatype = varargin{4};
if nargin>4
   tag_name  = cell((nargin-4)/2,1);
   tag_value = cell((nargin-4)/2,1);
   for ii=1:(nargin-4)/2
      tag_name{ii}  = varargin{2*ii+3};
      tag_value{ii} = varargin{2*ii+4};
   end
end
```

```
% check errors
if ~isnumeric(imgdata)
    error('The first input argument (image data) must be a numeric array.');
end
if ~isempty(header) && ~isstruct(header)
   error('The second input argument (header info) must be empty or a
structure.');
end
if ~ischar(imfile)
   error('The third input argument (output tif file name) must be string.');
end
switch lower(datatype)
   case 'logical'
      BitsPerSample = 1;
      SampleFormat  = 1;
      imgdata = logical(imgdata);
   case 'uint8'
      BitsPerSample = 8;
      SampleFormat  = 1;
      imgdata = uint8(imgdata);
   case 'int8'
      BitsPerSample = 8;
      SampleFormat  = 2;
      imgdata = int8(imgdata);
   case 'uint16'
      BitsPerSample = 16;
      SampleFormat  = 1;
      imgdata = uint16(imgdata);
   case 'int16'
      BitsPerSample = 16;
      SampleFormat  = 2;
      imgdata = int16(imgdata);
   case 'uint32'
      BitsPerSample = 32;
      SampleFormat  = 1;
      imgdata = uint32(imgdata);
   case 'int32'
      BitsPerSample = 32;
      SampleFormat  = 2;
      imgdata = int32(imgdata);
    case 'single'
      BitsPerSample = 32;
      SampleFormat  = 3;
      imgdata = single(imgdata);
   case {'uint64','int64','double'}
      BitsPerSample = 64;
```

```
      SampleFormat  = 3;
      imgdata = double(imgdata);
   otherwise
      error('Invalid output data type.');
end

% creat a Tiff object
t = Tiff(imfile,'w');

% duplicate tags from header
tagnamelist = Tiff.getTagNames();
tagnamelist_delete = {...
   'StripByteCounts';...
   'StripOffsets';
   'TileByteCounts';...
   'TileOffsets';...
   'MaxSampleValue';...
   'MinSampleValue';...
   'ResolutionUnit'};
for ii=1:length(tagnamelist_delete)    % remove read only tag names
   tagnamelist(strcmpi(tagnamelist_delete{ii},tagnamelist)) = [];
end
if ~isempty(header)
   for ii=1:length(tagnamelist)
      if isfield(header,tagnamelist{ii}) && ~isempty(header.(tagnamelist{ii}))
        tagstruct.(tagnamelist{ii}) = header.(tagnamelist{ii});
      end
   end
end

% update tags determined from imgdata and datatype
tagstruct.ImageLength = size(imgdata,1);
tagstruct.ImageWidth = size(imgdata,2);
tagstruct.SamplesPerPixel = size(imgdata,3);
tagstruct.SampleFormat = SampleFormat;
tagstruct.BitsPerSample = BitsPerSample;

% update some default tags (these tags can be overriden by user input below)
tagstruct.Photometric = Tiff.Photometric.MinIsBlack;
tagstruct.Compression = 1;
tagstruct.PlanarConfiguration = Tiff.PlanarConfiguration.Chunky;
tagstruct.Software = 'MATLAB';

% update user input tag values
if nargin>4
   for ii=1:length(tag_name)
      tagstruct.(tag_name{ii}) = tag_value{ii};
```

```
   end
end

% set tags and write to tif file
t.setTag(tagstruct)
t.write(imgdata);
t.close();
```

Extra LIM Fea

```
function [lImFea] = extr_lIm_fea( lIm )

[nrow, ncol] = size(lIm);

lImFea = zeros([nrow, ncol, 4]);

% first order gradient filters
hf1 = [-1,0,1];
vf1 = [-1,0,1]';

lImFea(:, :, 1) = conv2(lIm, hf1, 'same');
lImFea(:, :, 2) = conv2(lIm, vf1, 'same');

% second order gradient filters
hf2 = [1,0,-2,0,1];
vf2 = [1,0,-2,0,1]';

lImFea(:, :, 3) = conv2(lIm,hf2,'same');
lImFea(:, :, 4) = conv2(lIm,vf2,'same');
```