

Interactive Visualisation Techniques for Data Mining of Satellite Imagery

By

Samuel John Welch, BComp

A dissertation submitted to the

School of Computing

In partial fulfilment of the requirements for the degree of

Bachelor of Computing with Honours



University of Tasmania

November 2006

Declaration

I, Samuel John Welch declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution. To my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis.

Samuel John Welch

Abstract

Supervised classification of satellite imagery largely removes the user from the information extraction process. Visualisation is an often ignored means by which users may interactively explore the complex patterns and relationships in satellite imagery. Classification can be considered a “hypothesis testing” form of analysis. Visual Data Mining allows for dynamic hypothesis generation, testing and revision based on a human user’s perception. In this study Visual Data Mining was applied to the classification of satellite imagery.

After reviewing appropriate techniques and literature a tool was developed for the visual exploration and mining of satellite image data. This tool augments existing semi-automatic data mining techniques with visualisation capabilities. The tool was developed in IDL as an extension to ENVI, a popular remote sensing package.

The tool developed was used to conduct a visual data mining analysis of high-resolution imagery of Heard Island. This process demonstrated the positive impacts of visualisation and visual data mining when used in the analysis of satellite imagery. These impacts consist of: increased opportunity for understanding and hence confidence in classification results, increased opportunity for the discovery of subtle patterns in satellite imagery and the ability to create, test and revise hypotheses based on visual assessment.

Acknowledgements

Firstly, I would like to thank my supervisors, Dr. Ray Williams and Dr. Arko Lucieer. Ray and Arko's feedback and guidance this year helped immensely. Arko's work in preparing the Heard Island data for Humphrey and I was very much appreciated. Thanks go to both Arko and Ray for weathering my storm of last minute re-writes.

Thanks are due to my fellow honours room residents, Luke, Tristan, Matt, Jaidev, and Mathew, and to the rest of the Hobart honours crew. Some humour and a few games made for an enjoyable work environment. Thanks also to Humphrey in Launceston. It's been good working with you this year and it's been interesting to see how our two projects developed independently and yet, in the end, each has much to contribute to the other.

Many thanks are also due to my parents. Mum and Dad, your support and encouragement have been greatly appreciated.

Finally, my thanks go to my girlfriend, Jessica. Thank you for helping me make it through this year. Your comforting support kept me sane and helped me through this stressful time.

Table of Contents

DECLARATION	II
ABSTRACT	III
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS	V
LIST OF FIGURES.....	VII
LIST OF TABLES.....	IX
LIST OF EQUATIONS.....	X
CHAPTER 1 INTRODUCTION.....	1
1.1 PROBLEM DESCRIPTION	1
1.2 RESEARCH OBJECTIVES.....	2
1.2.1 Hypothesis	2
1.2.2 Aim.....	2
CHAPTER 2 BACKGROUND	3
2.1 REMOTE SENSING.....	3
2.1.1 Introduction	3
2.1.2 Remotely Sensed Data	3
2.1.3 Pre-processing.....	6
2.1.4 Interpretation.....	6
2.2 DATA MINING	8
2.2.1 Introduction	8
2.2.2 Feature Space.....	10
2.3 SUPERVISED REMOTE SENSING IMAGE CLASSIFICATION.....	11
2.3.1 Introduction	11
2.3.2 Maximum Likelihood	12
2.3.3 Minimum Distance.....	13
2.3.4 Level-Slice and Parallelepiped.....	13
2.3.5 Instance Based Learning	14
2.3.6 Soft Classification.....	14
2.4 TEXTURE	15
2.5 SCIENTIFIC VISUALISATION.....	16
2.5.1 Introduction	16
2.5.2 Volumes	17
2.5.3 Computer Graphics for Displaying Feature Space	18
2.6 VISUAL DATA MINING	20
2.6.1 Introduction	20
2.6.2 Exploratory Data Analysis.....	22
2.6.3 Summary.....	22
2.7 VISUAL DATA MINING OF REMOTELY SENSED DATA	23
CHAPTER 3 METHODS.....	24
3.1 OVERVIEW	24
3.2 CONSTRUCTING VOLUME REPRESENTATIONS	25
3.3 INTERSECTION OF VOLUMES	28
3.4 DIRECT VOLUME RENDERING	29
3.5 BOUNDARY REPRESENTATION AND SURFACES	30
3.5.1 Problems with Direct Volume Rendering and Point Clouds	30
3.5.2 Ellipsoids	32
3.5.3 α -shapes.....	32
3.5.4 Isosurfaces	35
3.6 HIGHLIGHTING CONFLICTING TRAINING DATA.....	39
3.6.1 Linking Data Spaces	39
3.6.2 Limitations	40

3.6.3	<i>Determining Pixel Values</i>	42
3.6.4	<i>Displaying Offenders</i>	43
3.7	CONFIGURING THE VISUALISATION	44
3.8	IMPLEMENTATION OF THE PROTOTYPE	45
CHAPTER 4 CASE STUDY		46
4.1	INTRODUCTION	46
4.2	STUDY AREA	46
4.2.1	<i>Paddick Valley, Heard Island, Australia</i>	46
4.2.2	<i>Dataset</i>	47
4.2.3	<i>Statistics</i>	48
4.3	SIMPLE SPECTRAL CLASSIFICATION	49
4.3.1	<i>Overview</i>	49
4.3.2	<i>Iteration 1</i>	49
4.3.3	<i>Iteration 2</i>	56
4.4	VEGETATION CLASSIFICATION	66
4.5	VISUALISING TEXTURE	67
4.6	VISUALISING PRINCIPAL COMPONENTS BANDS	68
CHAPTER 5 DISCUSSION		70
CHAPTER 6 CONCLUSION		72
CHAPTER 7 FURTHER WORK		74
7.1	PROTOTYPE ENHANCEMENTS	74
7.2	VISUAL DATA MINING IN REMOTE SENSING	74
REFERENCES		76
APPENDIX A IMAGERY		79
	HOBART LANDSAT TM IMAGE	79
	PADDICK VALLEY IKONOS IMAGE	81
APPENDIX B PROGRAM CODE		87

List of Figures

FIGURE 2-1: REFLECTANCE IN EACH OF THE LANDAT TM SATELLITES 6 BANDS .	5
FIGURE 2-2: LANDAT TM IMAGE OF HOBART, TASMANIA IN VISIBLE LIGHT (BANDS 3, 2 AND 1).	6
FIGURE 2-3: FALSE COLOUR COMPOSITE USING BANDS 4, 1 AND 5 OF THE LANDSAT TM IMAGE OF HOBART.	6
FIGURE 2-4: THEMATIC MAP OF THE LANDSAT TM HOBART IMAGE	8
FIGURE 2-5: A SAMPLE FEATURE SPACE PLOT.	10
FIGURE 2-6: FEATURE SPACE PLOTS.	11
FIGURE 2-7: EXAMPLES OF DIFFERENT SCIENTIFIC VISUALISATIONS.	17
FIGURE 2-8: MRI OF THE HUMAN HEAD AND 2 SLICES EXTRACTED FROM THE VOLUME.	18
FIGURE 2-9: VOLUME RENDERING OF A CT SCAN OF AN ENGINE COMPONENT	20
FIGURE 3-1: EXAMPLE VOLUME RENDERING SHOWING VARYING VOXEL VALUES.	26
FIGURE 3-2: RANGES OF PIXEL VALUES TO BE SORTED TO FIVE COLUMNS.	27
FIGURE 3-3: RANGES OF PIXEL VALUES TO BE SORTED TO TEN COLUMNS.	27
FIGURE 3-4: SCREEN CAPTURES FROM AN EXAMPLE USAGE SCENARIO.	31
FIGURE 3-5: ELLIPSOIDS FOR REGIONS DEFINED IN FIGURE 3-4A AT 60% TRANSPARENCY	32
FIGURE 3-6: A-SHAPES FOR A SAMPLE POINT SET	33
FIGURE 3-7: EXAMPLE FORMATION OF A 2D A-SHAPE.	34
FIGURE 3-8: THREE ISOSURFACES FOR A CLUSTER OF POINTS IN FEATURE SPACE.	36
FIGURE 3-9: MARCHING CUBE (LORENSSEN & CLINE 1987)	39
FIGURE 3-10: SYMMETRICALLY DISTINCT TRIANGULATIONS (LORENSSEN & CLINE 1987).	39
FIGURE 3-11: ENVI DISPLAY SHOWING A LANDSAT TM IMAGE OF HOBART, TAS.	41
FIGURE 3-12: FEATURE SPACE INTERSECTION WITH A VOLUME SIZE OF 64.	41
FIGURE 3-13: FEATURE SPACE INTERSECTION WITH A VOLUME SIZE OF 32.	42
FIGURE 3-14: HIGHLIGHTED CONFLICT BETWEEN REGIONS.	44
FIGURE 4-1: IMAGE BAND STATISTICS OF THE STUDY AREA	49
FIGURE 4-2: ENVI USER INTERFACE WITH DISPLAY OF PADDICK VALLEY WITH REGIONS.	52
FIGURE 4-3: ISOSURFACES FOR ALL FOUR ROIS IN THE 1C-ALL ROI SET	52
FIGURE 4-4: ISOSURFACES FOR THREE ROIS IN THE 1C-ALL ROI SET	53
FIGURE 4-5: POTENTIAL INTERSECTION IN FEATURE SPACE BETWEEN THE ROCK AND WATER ROIS.	53
FIGURE 4-6: ISOSURFACE REPRESENTING THE INTERSECTION BETWEEN ROCK AND WATER ROIS.	54
FIGURE 4-7: FEATURE SPACE PLOT SHOWING NO INTERSECTION BETWEEN ROIS.	54
FIGURE 4-8: OVERLAYED FALSE COLOUR COMPOSITE OF PADDICK VALLEY DEMONSTRATING HIGHLIGHTING OF CONFLICTING PIXEL VALUES.	55
FIGURE 4-9: PIXELS CAUSING OVERLAP	55
FIGURE 4-10: AVERAGE SPECTRUM PLOT SHOWING HIGHER VARIANCE IN BANDS 4 AND 1.	56
FIGURE 4-11: INTERSECTION BETWEEN ROCK AND WATER BASED ON BANDS 4, 2 AND 1. NOTE NOW THAT THE INTERSECTION FORMS TWO BLOBS IN SPACE (WHEN RENDERED AS AN ISOSURFACE) IMPLYING TWO VOXELS OF OVERLAP.	57
FIGURE 4-12: PIXELS CAUSING OVERLAP IN THE WATER AND ROCK REGIONS.	58
FIGURE 4-13: COMPARISON OF DECISION SURFACES.	59
FIGURE 4-14: MEAN POINTS FOR EACH REGION IN FEATURE.	60
FIGURE 4-15: FALSE COLOUR COMPOSITE SHOWING BANDS 4, 2 AND 1 OF A SUBSET OF THE PADDICK VALLEY IMAGE.	60
FIGURE 4-16: THEMATIC IMAGE SHOWING THE RESULT OF A MINIMUM DISTANCE CLASSIFICATION OF THE IMAGE SUBSET IN FIGURE 4-15.	61
FIGURE 4-17: MEANS AND EXTENT FOR EACH REGION.	61
FIGURE 4-18: ELLIPSOIDS REPRESENTING THE DECISION BOUNDARY OF A MAXIMUM LIKELIHOOD CLASSIFIER.	62
FIGURE 4-19: THEMATIC IMAGE SHOWING THE RESULT OF A MAXIMUM LIKELIHOOD CLASSIFICATION OF THE IMAGE SUBSET IN FIGURE 4-15.	63
FIGURE 4-20: COMPARISON OF A-SHAPES AND ISOSURFACES	64
FIGURE 4-21: THEMATIC IMAGE SHOWING THE RESULT OF A K-NEAREST NEIGHBOUR CLASSIFICATION OF THE IMAGE SUBSET IN FIGURE 4-15	64
FIGURE 4-22: FEATURE SPACE PLOT SHOWING EXTENT OF REGIONS FOR A SIX VEGETATION CLASS ANALYSIS.	66
FIGURE 4-23: FEATURE SPACE PLOT SHOWING THE EFFECT OF TEXTURE ON CLASSIFICATION OF VEGETATION TYPES.	67

FIGURE 4-24: FALSE COLOUR COMPOSITE OF THREE DERIVED PRINCIPAL COMPONENT BANDS	68
FIGURE 4-25: FEATURE SPACE PLOT BASED ON 3 GENERATED PRINCIPAL COMPONENT BANDS.	69

List of Tables

TABLE 2-1: SAMPLE FROM THE IRIS PLANTS DATABASE (FISHER 1988)	9
TABLE 2-2 - COMMON TEXTURE MEASURES DERIVED FROM THE CO-OCCURRENCE MATRIX	15
TABLE 4-1: MULTI-SPECTRAL WAVELENGTHS RECORDED PER BAND BY THE IKONOS SATELLITE.....	47
TABLE 4-2: THE THREE CLASSIFICATION LEVELS.	48
TABLE 4-3: STATISTICS FOR PADDICK VALLEY IMAGE.....	48
TABLE 4-4: STATISTICS FOR ORANGE REGION SHOWN IN FIGURE 4-8.	50
TABLE 4-5: ERROR MATRICES FOR EACH CLASSIFICATION..	65

List of Equations

EQUATION 3-1: POTENTIAL ENERGY IN A GRAVITY FIELD	37
EQUATION 3-2: POTENTIAL ENERGY IN FEATURE SPACE.....	37