

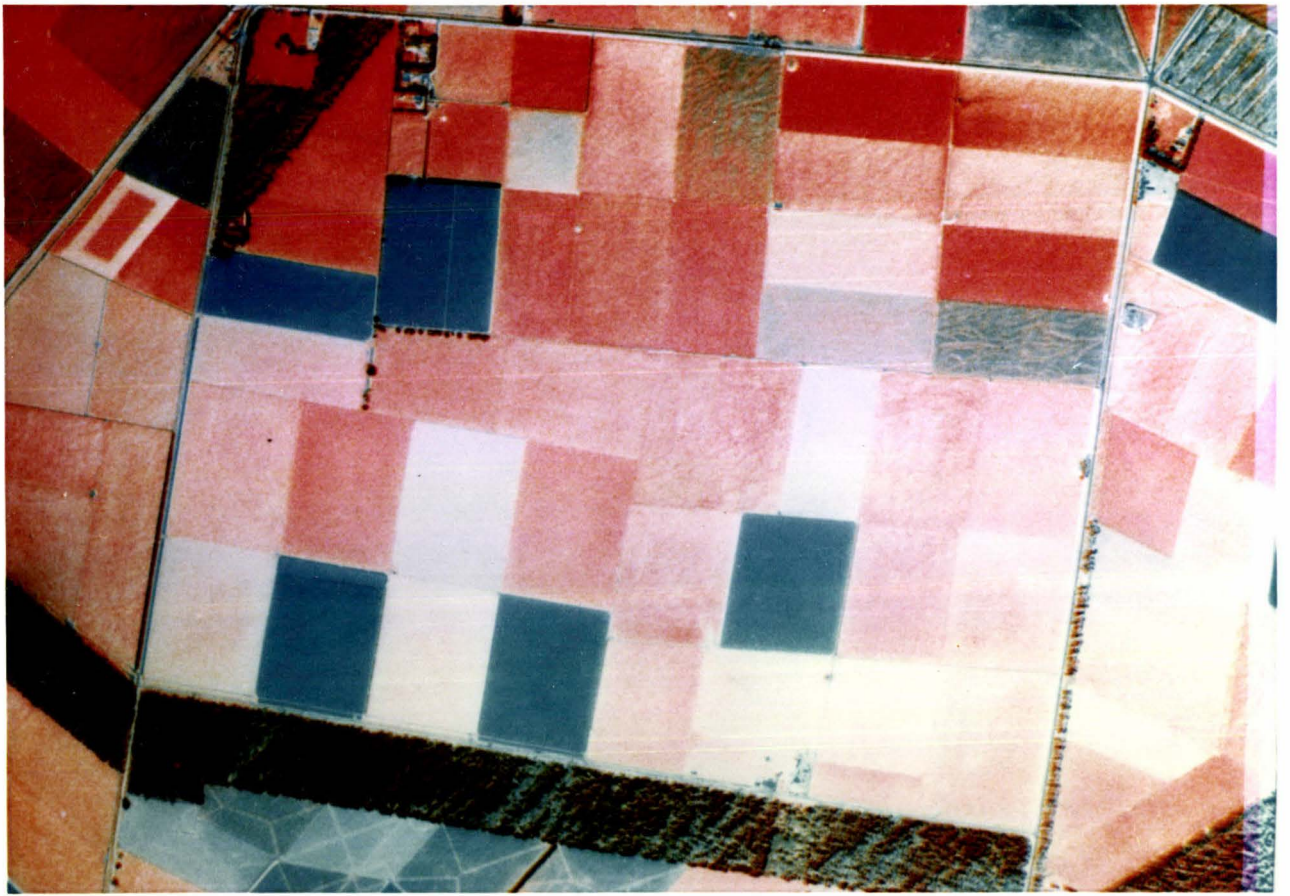
Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

UNSUPERVISED CLUSTERING OF
SPECTRAL SIGNATURES
IN LANDSAT IMAGERY

by Brian C. Clement B.E. (Hons)

December 1977

A Thesis presented in partial fulfilment
of the requirements for the degree of
Master of Arts in Computer Science
at Massey University



ABSTRACT

This thesis describes an investigation into automatic recognition of satellite imagery from the LANDSAT Project. Clustering techniques are shown to be the most suitable; of the three clustering algorithms investigated the k -means is shown to be the most effective. The need to perform edge detection on the images prior to clustering is also demonstrated. A suitable algorithm for edge detection is described.

Indexing terms: clustering, LANDSAT Satellite project, pattern recognition,
Satellite data

ACKNOWLEDGEMENTS

I wish to express my thanks to my project supervisor, Mr K.J. Hopper, for his encouraging assistance during the preparation of this thesis, and to Dr. B.E. Carpenter for introducing me to this area of Artificial Intelligence.

TABLE OF CONTENTS

1	Introduction	1
1.1	Data Requirements	1
1.2	Remote Sensing	2
1.3	LANDSAT	4
1.4	The Research Objective	5
2	Pattern Recognition	6
2.1	General	6
2.1	Choice of a Technique for LANDSAT Data	8
2.3	Supervised and Unsupervised Recognition	8
2.4	Clustering Principles	9
3	Current Satellite Imagery Research	12
3.1	Traditional Approaches	12
3.2	Important Aspects	14
4	Problem Approach	15
4.1	General Data Manipulation	15
4.2	Edge Detection	15
4.3	Classification	16
4.3.1	Shared Near Neighbour	16
4.3.2	The Divisive Approach	16
4.3.3	The <i>K</i> -means Algorithm	17
5	General Data Manipulation	20
5.1	File Handling	20
5.2	Display Routines	21
5.3	Analysis Routines	22
6	Edge Detection	28
6.1	Fuzziness	28
6.2	Differentiating	28
6.3	Finding Boundaries	29
6.4	Correlating Boundary Information	31

7	The Algorithms Used	.	.	.	41
7.1	Shared Near Neighbour	.	.	.	41
7.2	MAXFINDER - the Divisive Approach	.	.	.	42
7.3	CENTREFINDER - the <i>K</i> -means Approach	.	.	.	44
8	Project Assessment	.	.	.	48
8.1	Practical Problems	.	.	.	48
8.2	Boundary Points	.	.	.	48
8.3	Classification	.	.	.	49
8.4	Overall Success	.	.	.	49
8.5	Suggestions for Further Work	.	.	.	49
	References	.	.	.	50
	Bibliography	.	.	.	52
	Annex A	.	.	.	53
	Annex B	.	.	.	55
	Annex C	.	.	.	57
	Annex D	.	.	.	59
	Annex E	.	.	.	61
	Annex F	.	.	.	63
	Annex G	.	.	.	65
	Annex H	.	.	.	67
	Annex I	.	.	.	71
	Annex J	.	.	.	77
	Annex K	.	.	.	81
	Annex L	.	.	.	84

LIST OF ILLUSTRATIONS

Figure		Page
1	The LANDSAT Multispectral Scanner . . .	5
2	The <i>K</i> -means Algorithm . . .	18
3a	The output produced by ALLEVELS for Band 7 showing the intensity levels recorded for the entire test area . . .	23
3b	Comparison of Image registration . . .	24
4	The output produced by SHADES for Band 7. A shaded version of fig 3a. . .	25
5	The output produced by HISTOGRAMMER for Band 7 . . .	27
6	The output produced by DISTRIBUTION/ANALYSER for Bands 4 and 6 showing the correlation between the two . . .	27
7	The output produced by DIFFERENTIATOR for Band 7 showing the edges detected . . .	30
8a	Boundary points detected in Band 4 for $T = 10\%$. . .	32
8b	Boundary points detected in Band 5 for $T = 10\%$. . .	32
8c	Boundary points detected in Band 6 for $T = 10\%$. . .	33
8d	Boundary points detected in Band 7 for $T = 10\%$. . .	33
9a	Boundary points detected in Band 4 for $T = 15\%$. . .	34
9b	Boundary points detected in Band 5 for $T = 15\%$. . .	34
9c	Boundary points detected in Band 6 for $T = 15\%$. . .	35
9d	Boundary points detected in Band 7 for $T = 15\%$. . .	35
10a	Merged boundaries for $T = 10\%$. Boundary points appearing in at least two of the four files (fig 8) are included. . .	37
10b	Merged boundaries for $T = 10\%$. Boundary points appearing in at least three of the four files (fig 8) are included. . .	37
10c	Merged boundaries for $T = 10\%$. Only boundary points appearing in all four files (fig 8) are included. . .	38

11a	Merged boundaries for $T = 15\%$. Boundary points appearing in at least two of the four files (fig 9) are included.	39
11b	Merged boundaries for $T = 15\%$. Boundary points appearing in at least three of the four files (fig 9) are included.	39
11c	Merged boundaries for $T = 15\%$. Only boundary points appearing in all four files (fig 9) are included.	40
12a	Clustered output from SHAREDNN for $k = 20$, $k_t = 12$	43
12b	Clustered output from SHAREDNN for $k = 20$, $k_t = 13$	43
13a	Clustered output from CENTREFINDER	46
13b	Clustered output from CENTREFINDER with boundary points classified	46
13c	Ground truth for the test area	47
A.1	Structure diagram of FILEMAKER	54
B.1	Structure diagram of FLIPPER	56
C.1	Structure diagram of ALLLEVELS	58
D.1	Structure diagram of SHADES	60
E.1	Structure diagram of HISTOGRAMMER	62
F.1	Structure diagram of DISTRIBUTION/ANALYSER	64
G.1	Structure diagram of DIFFERENTIATOR	66
H.1	Structure diagram of BOUNDARYFINDER (i)	68
H.2	Structure diagram of BOUNDARYFINDER (ii)	69
H.3	Structure diagram of BOUNDARYFINDER (iii)	70
I.1	Structure diagram of BOUNDARYMERGER (i)	73
I.2	Structure diagram of BOUNDARYMERGER (ii)	74
I.3	Structure diagram of BOUNDARYMERGER (iii)	75
I.4	Structure diagram of BOUNDARYMERGER (iv)	76
J.1	Structure diagram of SHAREDNN (i)	78
J.2	Structure diagram of SHAREDNN (ii)	79
J.3	Structure diagram of SHAREDNN (iii)	80

K.1	Structure diagram of MAXFINDER (i)	.	.	82
K.2	Structure diagram of MAXFINDER (ii)	.	.	83
L.1	Structure diagram of CENTREFINDER (i)	.	.	86
L.2	Structure diagram of CENTREFINDER (ii)	.	.	87