The automatic hierarchical decomposition of images into sub-images for use in image recognition and classification

Paul Stefan Williams

This thesis is presented for the degree of Doctor of Philosophy of
The University of Western Australia

Centre for Intelligent Information Processing Systems
Department of Electrical and Electronic Engineering
The University of Western Australia
Nedlands, WA 6907, Australia

September 1999

Abstract

This thesis addresses the problem of image recognition and classification with the aim of developing a generic framework applicable to a wide range of image types and recognition problems. The approach is to use an algorithm, or set of algorithms, to automatically infer and extract structure which possibly corresponds to meaningful objects. Extracted structure is examined for further structure, hence providing a hierarchical process.

The investigated framework, referred to as the UpWrite, consists in its ideal form of localisation, amalgamation and representation phases. Initially, an image is examined locally to extract features which are used as low level primitives. These are analysed using one or several algorithms and amalgamated into higher level entities based on structure. The entities, possibly corresponding to meaningful objects, are represented in a chosen standard form. They may be classified or analysed locally in this new form for amalgamation and representation at an increasingly higher level of abstraction. Hence, the process is hierarchical, transforming numerous low level primitives into increasingly higher level objects. It is also intended to be generic, making few assumptions without image specific parameters, and while being automated, also allow a human to determine why a particular result was produced. The application of the framework is demonstrated on three restricted decomposition problems.

1. **Anomaly detection:** Grey scale images containing a few isolated defects or anomalous regions on an otherwise homogeneous background are investigated. With the focus on painted steel images containing subtle surface defects, the background is successfully identified and represented. The isolated anomalies are subsequently investigated by locally analysing the corresponding feature vectors. Extracted structure is sufficient to classify images by defect type.

2. **Grey scale newspaper segmentation:** The second restricted problem aims to decompose grey scale newspaper images into regions corresponding to background, pictures and text. Using the same feature extraction framework applied to anomaly detection, and identifying multiple objects rather than one single homogeneous background, the image is segmented. Results are very good, although subjective, across all images examined. The method compares favourably to segmentation using Gabor filters based on region prototypes.

3. **Coloured natural scenes:** Finally, the less restricted problem of segmenting coloured natural scenes is examined. Using the same type of low level primitives with the inclusion of a novel colour representation, a region growing segmentation algorithm amalgamates local features into spatially disjoint regions. The representations of these regions are used as predictors to amalgamate objects consisting of disjoint segments. Classification of 1000 images from 11 categories yields 55.1% accuracy.

In all, this generic framework is successfully applied to three different problems. Numerous useful tools are developed which are applicable for a wide variety of image classification and recognition problems. Most importantly the adopted philosophy in association with these tools provides a powerful means for investigating new problems.

Contents

Abstract iii

Acknowledgements xix

1 Introduction 1

1.1 Hierarchical decomposition of images into sub-images . . . 1
1.1.1 The structure of words in English text . . . 1
1.1.2 Stroke detection in binary images . . . 3
1.1.3 The question being addressed . . . 4
1.2 Image definitions and scope . . . 5
1.2.1 Definition of an image and sub-image . . . 5
1.2.2 The class of images. What data are we dealing with? . . . 7
1.2.3 Examples of legitimate sub-images . . . 8
Text strings as images . . . 8
The structure of cubes . . . 8
The detection of anomalies . . . 9
Newspaper print . . . 9
Natural scenes . . . 9
1.3 Reasons for thinking images often do decompose hierarchically . . . 10
1.4 The objectives . . . 11
1.4.1 Generic feature extraction . . . 12
1.4.2 Automatic extraction of structure . . . 12
1.4.3 Classification and recognition . . . 12
1.4.4 Image reconstruction . . . 13
1.5 General uses and applications . . . 13
1.5.1 Anomaly detection and classification . . . 13
1.5.2 Segmentation of newspaper or mixed-mode images . . . 14
1.5.3 Content based image retrieval (CBIR) . . . 15
1.5.4 Scene analysis and robot applications . . . 15
1.5.5 Image compression . . . 15
1.6 Thesis outline . . . 16

2 Background Work . . . **19**

2.1 Traditional approaches to and fundamental issues of image recognition and classification . . . 20

2.1.1 Ad Hoc and specific examples . . . 20

Object segmentation and counting . . . 21

2.1.2 Template matching . . . 22

The Hough transform . . . 23

2.1.3 Shape representation . . . 24

Region descriptions . . . 24

Boundary and one dimensional descriptions . . . 28

Measurements, robustness and closing comments . . . 28

2.1.4 Features and feature vectors . . . 29

2.1.5 Clustering . . . 30

Exploratory data analysis . . . 30

How many clusters? . . . 31

Metrics and normalisation . . . 32

Dendrograms and cluster trees . . . 33

The k-means and CLUSTER algorithm . . . 33

Gaussian mixture models using the EM algorithm . . . 34

2.2 Locating defects . . . 36

2.3 Frequency methods . . . 39

2.3.1 The Fourier transform and its useful properties . . . 39

The discrete Fourier transform (DFT) . . . 44

2.3.2 Gabor filters . . . 45

Definition . . . 45

Properties and rationale for use . . . 46

2.3.3 Frequency issues, filter design and related characteristics . . . 48

2.3.4 Summary . . . 53

2.4 Texture and texture discrimination . . . 54

2.4.1 Introduction, texture theories and properties . . . 54

2.4.2 The Brodatz textures . . . 58

2.4.3 Statistical and simple texture measures . . . 58

2.4.4 Co-occurrence matrices . . . 61

2.4.5 The frequency spectra of textures . . . 62

2.4.6 Gabor filters . . . 64

2.4.7 Markov random fields . . . 67

2.4.8 Closing comments . . . 70

2.5 Segmentation . . . 71

2.5.1 Scope and types of images . . . 72

2.5.2 Histogramming and thresholding . . . 73

Standard techniques . . . 73
Specific examples . . . 76
2.5.3 Edge detection . . . 77
2.5.4 Regions growing and split merge techniques . . . 81
2.5.5 Segmentation using Gabor filters . . . 83
Dunn's optimal filter design . . . 87
2.5.6 Markov random fields . . . 88
Final comments . . . 89
2.5.7 Mixed-mode segmentation . . . 90
2.5.8 Summary . . . 93
2.6 Colour image representation and analysis . . . 93
2.6.1 Red, green and blue (RGB) and normalised RGB . . . 93
2.6.2 Hue, saturation and intensity: HSI Colour cones and cylinders . . . 95
2.6.3 Perception initiated schemes . . . 96
2.6.4 The CIE system of colourimetry . . . 99
2.6.5 Example of colour image analysis . . . 100
RGB . . . 100
HSI . . . 101
Natural colour system (NCS) . . . 102
CIE representations . . . 103
2.6.6 Conclusion . . . 104
2.7 Content based image retrieval (CBIR) . . . 104
2.7.1 Overview . . . 105
2.7.2 Query by image content: IBM's QBIC . . . 108
2.7.3 VisualSEEk . . . 110
2.7.4 Berkeley digital library project . . . 113
2.8 Syntactic pattern recognition . . . 115
2.8.1 Overview . . . 115
2.8.2 Grammatical inference of natural language . . . 118
2.8.3 Stroke detection in binary images . . . 120

3 Approach: The *UpWrite* **125**
3.1 The UpWrite . . . 125
3.1.1 Localisation . . . 125
3.1.2 Amalgamation . . . 126
3.1.3 Representation for further UpWrites or classification . . . 127
3.2 The DownWrite . . . 128
3.3 Discussion . . . 129
3.3.1 Representation problems . . . 129
3.4 Summary of approach . . . 130

4 Detection, representation and classification of anomalies in grey scale images 133
4.1 Context and simplifications . . . 133
4.2 Example data and applications . . . 133
4.3 Local feature extraction . . . 138
4.3.1 Parameter choices, variations and generalisations . . . 139
4.3.2 Computational and memory considerations . . . 140
4.4 The feature space . . . 141
4.5 Identification of anomalies . . . 143
4.5.1 Method . . . 143
4.5.2 Results . . . 145
4.5.3 Implementation details . . . 147
4.5.4 Consideration of Mahalanobis distance . . . 149
4.5.5 Isolating anomalies . . . 151
4.6 Object representation . . . 152
4.7 Representation of anomalies and higher level processing . . . 153
4.7.1 Local Karhunen-Loeve transformation (KLT) algorithm . . . 153
4.7.2 An illustrative example in a low dimension . . . 154
4.7.3 Choosing the resolution radius . . . 155
4.7.4 Determining dominant eigenvectors based on eigenvalues . . . 155
4.7.5 Application to anomalous feature vectors . . . 157
4.8 Classification of painted steel defects . . . 158
Processing time . . . 163
4.9 Anomaly removal: An illustration of a partial DownWrite . . . 165
4.9.1 The method . . . 166
4.9.2 Results of image restoration . . . 169
4.10 Summary and concluding remarks . . . 171

5 Segmentation of newspaper images 175
5.1 Context and simplifications . . . 175
5.2 Example data and applications . . . 175
5.3 Local feature extraction . . . 176
5.4 The feature space . . . 177
5.4.1 General characteristics . . . 177
5.4.2 Clustering . . . 179
5.4.3 Rule based segmentation . . . 180
5.4.4 Feature vector parameterisation . . . 182
5.4.5 Summary and discussion . . . 184
5.5 Identification of text, picture, background and other regions . . . 186
5.6 Object description, representation and initial classification . . . 187
5.7 Spatial neighbourhood considerations . . . 188

5.7.1 Some other results . . . 189
5.8 Post processing and presentation of results . . . 190
5.8.1 Simple region identification algorithm . . . 191
5.8.2 Final results . . . 193
5.8.3 Timing considerations . . . 198
5.9 The DownWrite . . . 198
5.10 A comparison and conclusions . . . 199
5.10.1 Optimal Gabor filter segmentation . . . 199
5.10.2 Summary and additional comments . . . 202

6 Segmentation of natural images **205**
6.1 Context and problem specification . . . 205
6.2 Example data and applications . . . 205
Applications . . . 207
6.3 Local feature extraction . . . 207
6.3.1 Colour measurements . . . 207
6.3.2 Colour deviation . . . 210
6.3.3 Measurement normalisation . . . 211
6.3.4 Parameter values and implementation summary . . . 211
6.4 The feature space . . . 211
6.5 Identification of spatially disjoint homogeneous regions . . . 215
6.5.1 Determining local change . . . 215
6.5.2 Algorithm details . . . 216
6.5.3 Region growing conditions . . . 217
6.5.4 Implementation details . . . 218
6.5.5 Segmentation results . . . 221
6.5.6 Computational requirements . . . 222
Possible algorithm optimisations . . . 226
6.5.7 Closing comments . . . 228
6.6 Representation of disjoint regions . . . 228
6.6.1 Measurement representation . . . 228
6.6.2 Shape representation . . . 229
6.6.3 The **upwrite** program . . . 229
6.7 Amalgamation of represented disjoint regions . . . 230
6.7.1 Using the Fisher distance . . . 230
6.7.2 Mutual prediction method . . . 231
6.7.3 Dendrogram clustering . . . 232
6.7.4 Conclusions . . . 232
6.8 Classification . . . 232
6.8.1 Appropriate images . . . 232
6.8.2 Human classification . . . 234

6.8.3 Simple single region based classification 235
6.9 Image region reconstruction via a partial DownWrite 236
6.10 Summary and concluding remarks 240

7 Conclusions and future work **243**
7.1 Summary and closing discussion 243
7.2 Original contribution 245
7.3 Future research 246

Bibliography **249**

List of Figures

1.1 An example of English text with different word delimiters. 2
1.2 Decomposition of the character 7 into strokes. 3
1.3 Decomposition of the character 8 into strokes. 4
1.4 The *checkerboard* structure. An illustration of sub-images with non-trivial spatial relationships. 6
1.5 An ultrasound image of the prostate. 7
1.6 The *space symbol* dimension of a binary image representing a string. 8
1.7 The hierarchical structure of cubes. 8
1.8 The decomposition into sub-images corresponding to image background and an anomaly or object. 9
1.9 A newspaper image containing picture, text and background sub-images. . . 10
1.10 A natural scene and its decomposition into sub-images. 10
1.11 The hierarchical structure of a scene into meaningful regions. 11

2.1 Simple segmentation of light microscope image of red blood cells. 21
2.2 Pixel connectedness with paradox. 24
2.3 Shape reconstruction of a region using central geometric moments. 25
2.4 Shape reconstruction of a digit using Zernike moments. 26
2.5 Shape reconstruction of a region using low frequency FFT coefficients. . . . 27
2.6 The lack of robustness of skeletonisation for representing shape. 28
2.7 An ambiguous number of clusters. 31
2.8 The dependency of scale on the number of clusters. 32
2.9 Dendrogram for five patterns. 33
2.10 The inadequacies of square-error clustering. 34
2.11 Detection of a polished surface. 37
2.12 The linearity property of the Fourier transform. 40
2.13 An example FFT with low and high pass filter responses. 41
2.14 Functions and their Fourier transforms illustrating the convolution theorem. 42
2.15 The result of rotating the FFT by 45 degrees and taking the inverse. 43
2.16 The co-efficients of the DFT for a simple discrete signal. 44
2.17 An example of a 2D Gabor function in time and frequency. 47
2.18 A series of 2D Gabor functions with FFTs. 48

2.19 Variations of Gabor functions with FFTs. 49
2.20 Self similar Gabor *wavelets* and Gabor filter design. 50
2.21 DC bias exhibited by Gabor filters at low frequencies. 51
2.22 Characteristics of the Gabor and *log Gabor* frequency spectra. 51
2.23 An example of the *log Gabor* filter. 52
2.24 The result of swapping the phase and magnitude components of two images. 53
2.25 Images each containing two visually disjoint textures. 54
2.26 Two textures immediately identified by preattentive vision. 55
2.27 Preattentive texture discrimination example contradicting Julesz's original conjecture 56
2.28 Texture pairs used in psychophysical studies of texture perception. 57
2.29 A taxonomy of textured and non-textured images. 57
2.30 A co-occurrence matrix for a synthetic image with 4 grey scales. 61
2.31 The frequency spectrum of textured images. 63
2.32 Texture features extracted from the power spectral density. 64
2.33 Neighbourhood configurations used to define cliques for MRFs. 68
2.34 Texture synthesised using an MRF model. 70
2.35 A mixture of two Gaussians used to model image histograms. 75
2.36 The ideal scattergram allows for basic image segmentation of tone images. . 76
2.37 The Laplacian operator and two convolution mask approximations. 79
2.38 An example of segmentation using edge detection. 80
2.39 Example result from a simple region growing method. 82
2.40 Gabor filter response on a textured image with nonlinear post-processing. . 85
2.41 An illustration of the cross-fitting problem by performing line fitting in a transformed co-ordinate system. 86
2.42 Heuristic dimensionless weight and *alar* functions. 91
2.43 Segmentation of a newspaper image using clustering of Gabor filters. 91
2.44 A projection of normalised RGB colour and the *triangle colour* space. . . . 94
2.45 The two standard HSI colour spaces. 97
2.46 Natural colour or Ostwald system for representing and ordering colours. . . 97
2.47 Arrangement of the Munsell colour system. 98
2.48 The international standard colour specification system proposed by the CIE. 99
2.49 The variation of perceptual deviation in the CIE chromaticity diagram. . . 100
2.50 An example query result from QBIC. 109
2.51 An example query result from VisualSEEk's WebSEEk interface. 111
2.52 An example query result from VisualSEEk's SaFe interface. 112
2.53 The representation of patterns using primitives. 115
2.54 Overview of a simple traditional syntactic pattern recognition system. . . . 116
2.55 The hierarchical decomposition of a simple scene. 117
2.56 Entropy across a text segment. 119

2.57 The Spot Algorithm in action. . . . 121
2.58 The concept of a smooth triple used to determine whether line segments belong to a stroke. . . . 122
2.59 Projections of the feature space obtained by modelling ideal and hand-drawn strokes. . . . 123

3.1 The structure of two artificial datasets containing embedded manifolds. . . 126
3.2 An example of the DownWrite from a geometric moment representation. . . 128
3.3 Data for which robust and non-robust representations are available. . . . 130

4.1 Painted steel image containing defects. . . . 134
4.2 Images after histogram equalisation. . . . 135
4.3 A painted steel image supposedly containing a *dross* type defect. . . . 136
4.4 The screening of production line fabric at National Textiles Australia. . . . 137
4.5 A good quality example of a fabric image containing a flaw. . . . 138
4.6 The structure of the mask used to obtain local image characteristics. . . . 139
4.7 Example masks with different parameters. . . . 140
4.8 Projections of the feature space from an image containing a *dirtspot* defect. 142
4.9 The same projection for feature spaces for different images. . . . 143
4.10 Standard sized one and two dimensional Gaussian distributions. . . . 144
4.11 Detection of *resin* and *dross* type flaws. . . . 146
4.12 The detection of defects from the remaining images. . . . 146
4.13 The detection of anomalies in a section of fabric. . . . 147
4.14 Location of anomalous sections in a miscellaneous image. . . . 148
4.15 Mahalanobis histograms for *dirtspot* image 2012. . . . 150
4.16 Mahalanobis distance from original and *refined* background model. . . . 150
4.17 Mahalanobis histograms from original and *refined* background model. . . . 151
4.18 Mahalanobis histograms for the crate images in Figure 4.14. . . . 151
4.19 The modelling of a image containing whales in the ocean. . . . 152
4.20 Effect of performing morphological *closing* operation. . . . 152
4.21 A simple illustration of the local Karhunen-Loeve Transformation. . . . 154
4.22 Determining the number of principal eigenvalues. . . . 156
4.23 Representations of anomalies based on the Karhunen-Loeve Transformation. . . . 157
4.24 Spatial projections of the feature space corresponding to images with *dirtspot* and *resin* type defects. . . . 158
4.25 The percentage of points of each local dimension. . . . 159
4.26 Images used for classification. (part 1) . . . 160
4.26 Images used for classification. (part 2) . . . 161
4.27 Measurements from *dross*, *resin* and *dirtspot* images. . . . 162
4.28 Two dimensional data modelled by a Gaussian. . . . 166

4.29 Restoration of painted steel image number 2127 using the global image model to remove the anomaly. . . . 170
4.30 Restoration from last Figure with histogram equalisation. . . . 170
4.31 A section from Figure 4.30 outlining the limitations of the Gaussian model. 171
4.32 Restoration of a larger defect. . . . 171
4.33 Applying a high pass filter to painted steel images. . . . 172

5.1 A newspaper image appropriate for mixed-mode segmentation. . . . 176
5.2 Value using a standard deviation window measurement. . . . 177
5.3 Projection of the feature space for a single image. . . . 178
5.4 Projection with points manually coloured and re-projected onto $(\mathbf{x},\mathbf{y})$ offsets. 178
5.5 Results of applying the EM algorithm on the feature space. . . . 179
5.6 Segmentation using rule based method $(\mathbf{n} = 7)$. . . . 182
5.7 Rule based segmentation at a finer scale $(\mathbf{n} = 4)$. . . . 182
5.8 The relationship between $\mathbf{l}$ and $\mathbf{d}$ for an extreme point in two dimensions. . 183
5.9 Feature space parameterisation using varying blocksizes 185
5.10 Initial classification of the parameter space. . . . 187
5.11 Initial classification using different feature extraction parameters. . . . 187
5.12 Final classification after neighbourhood smoothing. . . . 188
5.13 Initial and final classification illustrating the effects of scale. . . . 189
5.14 An illustration of rotational invariance. . . . 190
5.15 The undesirable effects of the background assumptions. . . . 190
5.16 An example in which the method fails. . . . 191
5.17 Large disjoint picture and text regions. . . . 192
5.18 A full example which demonstrates invariance to image brightness. . . . 193
5.19 An image not suited to post-processing. . . . 194
5.20 Segmentation of an image with very smooth picture regions. . . . 194
5.21 Final segmentation results. (part 1) 195
5.21 Final segmentation results. (part 2) 196
5.22 Original grey scale regions classified as picture. . . . 197
5.23 The classification of a reconstructed image provides insight into the algorithm. 199
5.24 Texture prototypes considered for designing the optimal Gabor filter. . . . 200
5.25 Histograms of the optimal Gabor filter's magnitude response. . . . 201
5.26 A comparison with the optimal Gabor filter approach. . . . 202
5.27 Results using the same models applied in Figure 5.26. . . . 203
5.28 The sensitivity of the optimal Gabor filter to rotation. . . . 204

6.1 Example images from the Corel photo CD collection. . . . 206
6.2 Proposed tradeoff between the HSI conical and cylindrical spaces. . . . 208
6.3 The distribution of the hue-saturation plane for natural images. . . . 208
6.4 Histograms of chroma values for the considered Corel photo CD images. . . 209

6.5 An image with well defined regions and its window measurements. 212
6.6 Projections of the feature space. 213
6.7 Window measurements along a horizontal line through an aeroplane. 214
6.8 The location of homogeneous regions for an aeroplane image. 221
6.9 Segmentation results. (part 1) 223
6.9 Segmentation results. (part 2) 224
6.10 The number of remodel iterations required to locate regions. 227
6.11 An illustration of the common pattern recognition problem of occlusion. . . 230
6.12 Regions which inappropriately have a Fisher distance of 0. 231
6.13 Isolated regions corresponding to the same object separated due to occlusion. 231
6.14 Prediction results. 233
6.15 Example images which are possibly inappropriate for classification. 234
6.16 DownWrite examples with segmentation and region amalgamation results. . 237
6.17 The DownWrite may be useful for computer generated art. 239

List of Tables

2.1 Fourier transform characteristics for functions with special characteristics. . 42
2.2 A sample of the different treatments of the Brodatz textures. 59
2.3 Common features extracted from grey level co-occurrence matrices. 62
2.4 Convolution masks used for differential gradient edge detection. 78
2.5 Convolution masks used for template matching edge detection. 79
2.6 Eigenvalues from applying the KLT to the RGB space of 8 different images. 95
2.7 Principal eigenvector using the KLT on the RGB space with 4 different images. 95
2.8 Class symbols found in natural language text. 119

4.1 The time taken to calculate window measurements for an image. 141
4.2 The proportion of points outside N standard deviations of the center of a Gaussian distribution for various dimensions. 145
4.3 Dimension key for results shown in Figures 4.23 and 4.24. 157
4.4 Classification results for painted steel images. 164

5.1 Colour assignments for rule based classification. 181

6.1 The images used from the Corel photo CD collection. 207
6.2 Breakdown of time taken to segment the image from Figure 6.5. 225
6.3 Total time to segment the images. 226
6.4 The classification of the Corel photo CD images by a human. 235
6.5 Classification results for 1000 Corel photo CD images. 236

Acknowledgements

Most of all I would like to thank Yianni Attikiouzel and Mike Alder. Yianni continues to look after members of CIIPS in the sense of providing appropriate facilities, advice and general encouragement. Over many years he has done a great job and I wish him all the best in the future. Mike has provided brilliant supervision, inspiring students to thoughtful and creative research. I have really enjoyed our numerous and often long and fascinating talks, gaining greatly from Mike's knowledge in mathematics, physics, history and philosophy. I look forward to future conversations and sincerely hope that he returns to good health. I would also like to thank Chris deSilva for his supervision in Mike's absence.

I would like to acknowledge numerous friends and colleagues who have been helpful over the last few years. Robert McLaughlin perhaps stands out most. He is always friendly, constructive and willing to take the time to help other students and share his knowledge and experiences. He has a heart of gold and I wish him all the best at Oxford. Thanks to Gek Lim for her encouragement and for ensuring that we were always busy. Her character, company and professional nature will be an asset wherever she goes. Thanks to Jason Hutchens and Bruce Cooper for their companionship and friendship along the same undergraduate and postgraduate paths. Jason's pursuit of novel ideas and puzzles has led to some interesting pastimes. I would also like to thank Jason for joining me with our initial direction after PhD research. Thanks to the 1995 honours students for their participation in the coke wall, 'haki' and a generally enjoyable atmosphere. And to the 1997 CIIPS honours students for their friendship, 'bd4' participation and the soft drink challenge! I would also like to acknowledge Sonny Tham, Adrian Yau, Phil Dunstan and Daniel Harvey.

Finally, thanks go to my family for their continual support and encouragement in all aspects of life. Their encouragement and interest in my thesis completion, and the time spent proof reading it, will always be remembered. Special thanks to Claire and her family for their various forms of encouragement along the way. Lastly, I would like to thank my close friends who continue to gain the letters 'Dr' in front of their names. Lots of love go to you all.

Paul Stefan Williams, 1999.