
Machine learning methods for discriminating natural targets in seabed imagery



Richard John Patrick Harrison

A thesis submitted for the degree of
Doctor of Philosophy (PhD)

University of East Anglia
School of Computing Sciences

December 2012

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

The research in this thesis concerns feature-based machine learning processes and methods for discriminating qualitative natural targets in seabed imagery. The applications considered, typically involve time-consuming manual processing stages in an industrial setting. An aim of the research is to facilitate a means of assisting human analysts by expediting the tedious interpretative tasks, using machine methods. Some novel approaches are devised and investigated for solving the application problems. These investigations are compartmentalised in four coherent case studies linked by common underlying technical themes and methods. The first study addresses pockmark discrimination in a digital bathymetry model. Manual identification and mapping of even a relatively small number of these landform objects is an expensive process. A novel, supervised machine learning approach to automating the task is presented. The process maps the boundaries of ≈ 2000 pockmarks in seconds - a task that would take days for a human analyst to complete. The second case study investigates different feature creation methods for automatically discriminating sidescan sonar image textures characteristic of *Sabellaria spinulosa* colonisation. Results from a comparison of several textural feature creation methods on sonar waterfall imagery show that Gabor filter banks yield some of the best results. A further empirical investigation into the filter bank features created on sonar mosaic imagery leads to the identification of a useful configuration and filter parameter ranges for discriminating the target textures in the imagery. Feature saliency estimation is a vital stage in the machine process. Case study three concerns distance measures for the evaluation and ranking of features on sonar imagery. Two novel consensus methods for creating a more robust ranking are proposed. Experimental results show that the consensus methods can improve robustness over a range of feature parameterisations and various seabed texture classification tasks. The final case study is more qualitative in nature and brings together a number of ideas, applied to the classification of target regions in real-world sonar mosaic imagery. A number of technical challenges arose and these were surmounted by devising a novel, hybrid unsupervised method. This fully automated machine approach was compared with a supervised approach in an application to the problem of image-based sediment type discrimination. The hybrid unsupervised method produces a plausible class map in a few minutes of processing time. It is concluded that the versatile, novel process should be generalisable to the discrimination of other subjective natural targets in real-world seabed imagery, such as *Sabellaria* textures and pockmarks (with appropriate features and feature tuning.) Further, the full automation of pockmark and *Sabellaria* discrimination is feasible within this framework.

Statement of originality

Unless otherwise noted or referenced in the text, the work described in this thesis is, to the best of my knowledge and belief, original and my own work. It has not been submitted, either in whole or in part, for any degree at this or any other academic or professional institution.

R J P Harrison, December 2012

Publications

Journal (peer reviewed)

Harrison, R., Birchall, R., Mann, D. and Wang, W. (2012) Novel consensus approaches to the reliable ranking of features for seabed imagery classification, *International Journal of neural Systems*, 22 (6), DOI: 10.1142/S0129065712500268.

Conference (peer reviewed)

Harrison, R., Bellec, V., Mann, D. and Wang, W. (2011) A new approach to the automated mapping of pockmarks in multibeam bathymetry, *IEEE International Conference on Image Processing (ICIP)*.

Harrison, R., Bianconi, F., Harvey, R. and Wang, W. (2011) A texture analysis approach to identifying *Sabellaria spinulosa* colonies in sidescan sonar imagery, *Proceedings of the Irish Machine Vision and Image Processing (IMVIP) conference*, Dublin.

Harrison, R., Birchall, R., Mann, D. and Wang, W. (2011) A novel ensemble of distance measures for feature evaluation: Application to sonar imagery, *IDEAL 12th Int. Conf. Proc., Lecture Notes in Computer Science*, 6936, 327–336.

Harrison, R. (2009), A spatial model of a towed cable system and USBL acoustic beacon, in *Società Geologica Italiana, proceedings of the International conference on seafloor mapping for geohazard assessment*, Eds., Chiocci, F., Ridente, D., Casalbore, D. and Bosman, A., 7, 59-62.

Contents

List of figures	xiii
List of tables	xvi
Selected abbreviations	xviii
Selected glossary	xxii
Acknowledgements	xxxv
Chapter 1 Introduction	1
1.1 Background to the thesis and research project	2
<i>1.1.1 General background</i>	2
<i>1.1.2 Industrial collaboration</i>	3
1.2 General aim, objectives and scope of the research	3
1.3 Importance and relevance of the work	4
1.4 Summary of collective novelty and contributions of the research	5
1.5 Summary of novelty, main contributions and published work by case study	6
<i>1.5.1 Case study 1: Pockmark discrimination</i>	
<i>1.5.2 Case study 2: Features for Sabellaria discrimination</i>	
<i>1.5.3 Case study 3: Feature evaluation and ranking methods</i>	
<i>1.5.4 Case study 4: Unsupervised approach on real-world mosaic imagery</i>	
<i>1.5.5 Related research (not included in the thesis) published during the project</i>	
1.6 Thesis structure	9

Chapter 2 Research context	11
2.1 Introduction	11
2.2 Why we need information about the seabed	14
2.3 Data acquisition for seabed classification: Methods and issues	16
<i>2.3.1 Overview</i>	16
<i>2.3.2 Backscatter imagery</i>	17
<i>2.3.3 Bathymetry</i>	27
<i>2.3.4 Ground truth</i>	28
2.4 Other pertinent issues	30
2.5 Summary	32
Chapter 3 Related research	34
3.1 Introduction	34
3.2 Backscatter image classification methods	36
<i>3.2.1 Feature creation</i>	36
<i>3.2.2 Sidescan sonar backscatter image formation considerations</i>	43
<i>3.2.3 Classification and clustering methods</i>	47
3.3 Bathymetry features	50
3.4 Combining bathymetry and backscatter information	54

3.5 Ensemble methods	56
3.5.1 <i>Supervised ensemble</i>	56
3.5.2 <i>Unsupervised (cluster) ensemble</i>	57
3.6 Outline of machine learning methods used in the experimental work in this thesis	
3.6.1 <i>Computational kernels</i>	61
3.6.2 <i>Feature evaluation and selection</i>	63
3.6.3 <i>Supervised learning</i>	64
3.6.4 <i>Unsupervised learning</i>	66
3.7 Summary	67
Chapter 4 Research analysis and design	69
4.1 Introduction	69
4.2 Research analysis and design	70
4.2.1 <i>Background</i>	70
4.2.2 <i>Aim, objectives and scope of the project and case studies</i>	70
4.2.3 <i>Outline of research type and general methodology</i>	76
4.2.4 <i>Generic case study design</i>	78
4.2.5 <i>Timeline of general project activities</i>	80
4.2.6 <i>Risk awareness</i>	81
4.3 Discussion of evaluation issues	81

Chapter 5	A new approach to the automated mapping of pockmarks in multi-beam bathymetry	91
5.1	Introduction	92
5.1.1	<i>Pockmarks</i>	92
5.2	Problem outline and previous work	94
5.3	Methods and process design	95
5.3.1	<i>Overview of process and methods</i>	95
5.3.2	<i>Test data</i>	97
5.3.3	<i>Boundary definition</i>	98
5.4	Feature creation	99
5.4.1	<i>Terrain features</i>	99
5.4.2	<i>Texture features</i>	104
5.5	Feature evaluation	107
5.5.1	<i>Qualitative feature evaluation</i>	107
5.5.2	<i>Quantitative feature evaluation</i>	111
5.6	Real -World object discrimination	120
5.7	Evaluation and discussion of results	123
5.8	Evaluation, conclusions, recommendations and scope for further work	129
5.8.1	<i>Evaluation</i>	129
5.8.2	<i>Conclusions</i>	130
5.8.3	<i>Recommendations</i>	131
5.8.4	<i>Scope for further work</i>	132

Chapter 6	Feature based discrimination of Sabellaria spinulosa textures in sidescan sonar imagery	134
6.1	Introduction	135
6.2	Sabellaria texture discrimination in waterfall imagery	136
6.2.1	<i>Methods</i>	138
6.2.2	<i>Experimental work</i>	142
6.2.3	<i>Results and discussion</i>	144
6.3	Sabellaria texture discrimination in mosaic imagery	147
6.3.1	<i>The Gabor filter bank</i>	148
6.3.2	<i>Experimental work</i>	156
6.3.3	<i>Results and discussion</i>	160
6.3.4	<i>Further results and considerations</i>	168
6.3.5	<i>Summary</i>	175
6.4	Other useful features	176
6.5	Evaluation, conclusions, recommendations and scope for further work	184
6.5.1	<i>Evaluation</i>	184
6.5.2	<i>Conclusions</i>	185
6.5.3	<i>Recommendations</i>	186
6.5.4	<i>Scope for further work</i>	187
Chapter 7	Novel consensus approaches to the reliable ranking of features for seabed imagery classification	188
7.1	Introduction	189
7.2	Related work	192

<i>7.2.1 Domain specific feature selection</i>	192
<i>7.2.2 Robustness of feature selection methods</i>	194
7.3 Consensus feature ranking methods	196
<i>7.3.1 Nonparametric distribution estimation</i>	197
<i>7.3.2 Individual distance measures</i>	197
<i>7.3.3 Consensus ranking of features</i>	200
<i>7.3.4 Reliability weighting of permutations</i>	202
7.4 Application context	205
<i>7.4.1 GLCM feature creation</i>	205
<i>7.4.2 Experimental data</i>	207
7.5 Experimental results and discussion	208
<i>7.5.1 Investigating saliency and classification accuracy</i>	209
<i>7.5.2 Defining a baseline rank</i>	214
<i>7.5.3 Investigating task robustness</i>	215
7.6 Evaluation, conclusions, recommendations and scope for further work	222
<i>7.6.1 Evaluation</i>	222
<i>7.6.2 Conclusions</i>	223
<i>7.6.3 Recommendations</i>	224
<i>7.6.4 Scope for further work</i>	225
Chapter 8 Unsupervised classification of sonar imagery	226
8.1 Introduction	227
<i>8.1.1 Important note</i>	228

8.2 Previous work	228
8.2.1 <i>Problem background and solution outline</i>	228
8.2.2 <i>Sidescan sonar image classification</i>	233
8.2.3 <i>Ensemble/fusion approaches</i>	234
8.3 Proposed feature-distributed process and methods	235
8.3.1 <i>Sampling strategies</i>	237
8.3.2 <i>Image block processing scheme</i>	237
8.3.3 <i>Feature creation</i>	238
8.3.4 <i>Model induction</i>	240
8.3.5 <i>Independent classification channels</i>	242
8.3.6 <i>Post processing of output blocks</i>	243
8.3.7 <i>Fusion of classifications</i>	243
8.3.8 <i>Assigning class labels</i>	245
8.4 Experimental work	246
8.4.1 <i>Experimental aim</i>	246
8.4.2 <i>Data</i>	247
8.5 Results and discussion	248
8.5.1 <i>Inspection of the specimen classification</i>	248
8.5.2 <i>Single channel supervised and unsupervised approaches</i>	250
8.5.3 <i>Unsupervised fusion approach</i>	257
8.5.4 <i>Overall discussion</i>	262
8.6 Evaluation, conclusions, recommendations and scope for further work	264
8.6.1 <i>Evaluation</i>	264
8.6.2 <i>Conclusions</i>	266
8.6.3 <i>Recommendations</i>	266
8.6.4 <i>Scope for further work</i>	267

Chapter 9 Evaluation	268
9.1 Overall evaluation	268
9.2 Research project management	272
<i>9.2.1 Stakeholder influence on research design</i>	272
<i>9.2.2 Division of project time</i>	273
<i>9.2.3 Publication output and schedule</i>	274
<i>9.2.4 Execution of project case studies</i>	275
<i>9.2.5 Risk revisited</i>	278
Chapter 10 Conclusions and scope for further work	280
10.1 Conclusions	281
<i>10.1.1 Case study specific conclusions</i>	281
<i>10.1.2 General conclusions</i>	283
<i>10.1.3 General recommendation</i>	284
10.2 Scope for further work	285
<i>10.2.1 Mini proposal 1: Pockmark classification</i>	285
<i>10.2.2 Mini proposal 2: Unsupervised/supervised Sabellaria discrimination</i>	286
Appendix 1	
ICIP poster presentation	287
References	288

List of figures

Figure	Description	Page
G.1	Block processing layout illustration	xxiv
G.2	Tiling of an image block	xxix
2.1	Example of a seabed habitat map	12
2.2	Class map production processing flow	13
2.3	Examples of sidescan sonar image textures	18
2.4	Examples of sidescan sonar waterfall and mosaic imagery	19
2.5	Typical towed sonar instrument configuration	23
2.6	Rendered section of multibeam bathymetry raster	27
2.7	Modified (BGS) Folk Trigon	30
3.1	Machine learning ensemble (supervised)	56
3.2	Machine learning ensemble (unsupervised – cluster ensemble)	59
3.3	Mean filter kernel	62
4.1	Research design	77
4.2	Research timeline	80
4.3	Inter-rater variability of manual classifications	83
4.4	Class prescriptions for example classification	84
4.5	Machine vs. human segmentation	86
4.6	Uncertainty in ground truth location	88
5.1	Pockmark shapes	93
5.2	Pockmark mapping process	97
5.3	Pockmark nodes and boundaries	98
5.4	PCM kernel	100
5.5	Asymmetric pockmark	107
5.6	Feature response values	109
5.7	Feature response values at a pockmark control node	110
5.8	Feature response values at a non-pockmark control node	110
5.9	Feature values according to Bhattacharyya's distance measure	112
5.10	Feature ranking results using leave-one-out method	114
5.11	Correlation matrix	117
5.12	Training sample point selection user interface	121
5.13	Pockmark boundaries display	122
5.14	Evaluation regions	123

5.15	Comparison of regions	126
5.16	Comparison of raw segmentations	128
6.1	Operation of a sidescan sonar	136
6.2	Waterfall image segment	137
6.3	Examples of five textural classes	143
6.4	Intra-class variability of sandwave imagery	144
6.5	Variation in accuracy with sub-image size	146
6.6	Gabor filter kernels	151
6.7	Gabor filter banks at different resolutions	153
6.8	Gabor filter bank redundancy	154
6.9	Downsampling of image	155
6.10	Test image patches	159
6.11	Sabellaria-sand discrimination accuracy	161
6.12	Sabellaria-sand discrimination, sd of accuracy	162
6.13	Relative frequency of kernel sizes	163
6.14	Task dependent discrimination accuracy	165
6.15	Classification accuracy and feature cost	166
6.16	Composite image example	168
6.17	Classification accuracy variation	173
6.18	Visual classification results	174
6.19	Mosaic regions	178
6.20	Intensity histograms	178
6.21	Density of edge pixels	181
6.22	Variation in ratio of edge pixels	182
6.23	Variation in edge pixel ratio with kernel size	182
7.1	Process for estimating feature saliency and ranking reliability	204
7.2	Examples of textural classes	208
7.3	Classification accuracy vs. estimated saliency	210
7.4	Mean and sd of R^2 statistic	213
7.5	Baseline ranking of individual features	214
7.6	Mean and sd of correlation coefficient	217
8.1	Unsupervised model induction process	241
8.2	Sonar mosaic and class map	247
8.3	Relative frequency of class occurrences	249
8.4	Examples of class maps from different feature channels	250

8.5	Specimen machine class maps	252
8.6	Results from statistical feature channel	253
8.7	Results from filter bank feature channel	254
8.8	Relabelled, unsupervised class maps	255
8.9	Similarity of machine and human produced maps	256
8.10	UKSeaMap classification scheme and class maps	257
8.11	Unsupervised fusion strategy results	260
8.12	Manually produced map and machine produced map	262
9.1	Division of time on main research tasks	274
9.2	Publication timeline	275

List of tables

Table	Description	Page
G.1	Contingency table for binary classification outcomes	xxii
2.1	Factors influencing sidescan sonar imaging quality	20
3.1	Features derived from DEM and DBM data	52
4.1	Case study aims, objectives and scope	72
4.2	Summary of case study objectives according to SMART criteria	79
5.1	Summary of features and resolutions	106
5.2	Top feature combination subsets	116
5.3	Summary of agreements and disagreements	127
5.4	Summary of total agreements	127
6.1	Summary of textural classes	143
6.2	Overall classification accuracy	145
6.3	Total scores and rank	145
6.4	Output channel configuration	152
6.5	Summary of rotational configurations	155
6.6	Summary of feature space dimensions	158
6.7	Summary of parameter ranges	158
6.8	Mean and sd of classification accuracies	166
6.9	Summary of parameters and configurations	167
6.10	Class descriptions	169
6.11	Classification accuracy	171
6.12	Generic class labelling	172
6.13	Overall accuracy metrics	174
7.1	Candidate distance measures	198
7.2	Co-occurrence matrix parameters	206
7.3	Summary of classes	208
7.4	R^2 values	211
7.5	Satisfaction of properties (i) and (ii)	213
7.6	Rank score permutation evaluation	219
7.7	Mean and sd of Spearman's correlation coefficient	220
7.8	Satisfaction of properties (iii) and (iv)	221
8.1	Summary of main processing stages	236

8.2	Summary of feature channel parameters	240
8.3	Sediment types (and legends)	248
8.4	Pairwise NMI	251
8.5	Similarity to manual classification	256
8.6	Ground truth type conditional probabilities	259
9.1	Research themes compared with SMART	275
9.2	Further comparison of research themes	276

Abbreviations and acronyms

ACE: Automated Classification Engine

AIMD: Angular Inverse Difference
Moment

ANN: Artificial Neural Network

ASM: Angular Second Moment

BGC: Binary Gradient Contours

BGS: British Geological Survey

Bh: Bhattacharyya

BPI: Bathymetric Position Index

BVM: Ball Vector Machine

Ca: Canberra

CCA: Curvilinear Components
Analysis

CCD: Charged Coupled Device

CEFAS: Centre for Environment,
Fisheries and Aquaculture Science

CCR: Coordinated Cluster
Representation

CMOS: Complementary Metal Oxide
Semiconductor

CNN: Cellular Neural Network

Co: Consensus

Co+: Weighted consensus

CON: Contrast

COOC: Co-occurrence

COR: Correlation

Cs: Chi-squared

CSa: Coarse sand

CSt: Crushed stones

CWT: Complex Wavelet Transform

DBM: Digital Bathymetry Model

DDM: Digital Depth Model

DEM: Digital Elevation Model

DeO: Decision output

DFT: Discrete Fourier Transform	GA: Genetic Algorithm
DMP: Deviation from a Mean plane	GEF: Gabor Elementary Function
DsI: Soft decision input	GFB: Gabor Filter Bank
DT: Decision Tree	GIS: Geographic Information System
DT-CWT: Dual Tree – Complex Wavelet Transform	GLCIA: Grey Level Co-occurrence Iteration Algorithm
DWT: Discrete Wavelet Transform	GLCM: Grey Level Co-occurrence Matrix
EDF: Estimated Distribution Function	GMSL: Gardline Marine Sciences Limited
ENT: Entropy	GPS: Global Positioning System
EPSRC: Engineering and Physical Science Research Council	GT: Ground Truth
EU: European Union	GUI: Graphical User Interface
Eu: Euclidean	HABMAP: Habitat Mapping for Conservation and Management of the Southern Irish Sea
EUNIS: European Nature information system	HOM: Homogeneity
FFD: Fuzzy Fractal Dimension	HVS: Human Visual System
FFV: Full Feature Vector	ICES: International Council for the Exploration of the Sea
Fl: Flour	
FSa: Fine sand	

ICIP: International Conference on Image Processing

IDEAL: Intelligent Data Engineering and Automated Learning

IDM: Inverse Difference Moment

IFREMER: Institut Français de Recherche pour l'Exploitation de la Mer

IHO: International Hydrographic Organisation

ILBP: Improved Local Binary Pattern

INFOMAR: Integrated Mapping for the Sustainable Development of Ireland's Marine Resources

INS: Inertial Navigation System

KL: Kullback-Leibler divergence

k-NN: k – Nearest Neighbours

KS: Kolmogorov-Smirnov

LBP: Local Binary Pattern

LFH: Local Fourier Histogram

LGA: Laplacian of a Gaussian

MAGIC: Marine Geohazards along the Italian Coasts

MB: Megabyte

MBES: Multi Beam Echo Sounder

MDAC: Methane Derived Authigenic Carbonate

MESH: Mapping European Seabed Habitats

MFNN: Multi-layer Feed-forward Neural Network

Mix: Mixed sediments

MLP: Multi-Layer Perceptron.

Mnl: Mineral

Mu: Mussels

NGU: Geological Survey of Norway

NMI: Normalised Mutual Information

NN: Nearest Neighbour

NOC: National Oceanography Centre

OUTEX: Outex Texture Database (University of Oulu)

PCA: Principal Components Analysis	SNR: Signal to Noise Ratio
PCM: Prototype Crater Matching	SOM: Self Organising Map
PCNN: Pulse Coupled Neural Network	Sonar: Sound Navigation and Ranging
pdf: Probability density (mass) function	SVM: Support Vector Machine
PLVQ: Proportional Learning Vector Quantisation	Sw (Sw#): Sandwaves (Sandwaves1)
PRN: Pseudo-Random Noise	TexAn: Proprietary textural analysis software (University of Bath)
PSA: Particle Size analysis	TIFF (TIF): Tagged Image File Format
QTC: Quester Tangent Corporation	TPI: Topographic Position Index
RBF: Radial Basis Function	TRI: Terrain ruggedness Index
RF: Random Forest	UEA: University of East Anglia
Ro: Rock	US: Unsupervised
Sab: Sabellaria spinulosa	USBL: Ultra Short Base Line
SAC: Special Area of Conservation	USGS: United States Geological Survey
SAR: Synthetic Aperture Radar	VTP: Virtual Terrain Project
SD (sd): Standard Deviation	Wo: Wood
SMART: Specific, Measurable, Agreed, Realistic and Time constrained.	

Selected glossary

Accuracy (binary classification accuracy.) Binary classification accuracy can be defined as a function of the outcomes from a specimen machine classification task when the ground truth of the test instances is known. The four outcomes commonly specified are; (1) true positive (*TP*), (2) true negative (*TN*), (3) false positive (*FP*) and (4) false negative (*FN*). The number of instances of each outcome, e.g., $n(TP)$ can be summed and entered in a contingency table, as shown in table G.1.

		Machine predicted class	
		+	-
Actual (Ground truth) class	+	True positive (<i>TP</i>)	False Negative (<i>FN</i>)
	-	False positive (<i>FP</i>)	True negative (<i>TN</i>)

Table G.1 A contingency table for binary classification outcomes.

The outcomes are defined as followed:

TP A target ground truth instance is correctly identified by the machine as belonging to the target class.

FP A non-target ground truth instance is incorrectly identified by the machine as belonging to the target class.

FN A non-target ground truth instance is correctly identified by the machine as belonging to the non-target class.

TN A target ground truth instance is incorrectly identified by the machine as belonging to the non-target class.

The metrics of overall classification accuracy, sensitivity and specificity are then defined as:

$$\text{Overall accuracy} = \frac{n(TP) + n(TN)}{n(TP) + n(TN) + n(FP) + n(FN)}$$

$$\text{Sensitivity (true positive rate)} = \frac{n(TP)}{n(TP) + n(FN)}$$

$$\text{Specificity (true negative rate)} = \frac{n(TN)}{n(TN) + n(FP)}$$

Acoustic impedance. The acoustic impedance, Z , of a medium (material) is defined as $Z = \rho V$, where ρ is the density of the medium and V , the velocity of the acoustic wave within the medium.

Along-track. A direction parallel to the direction of motion of the longitudinal axis of the sonar transducer.

Across-track. Orthogonal to the along track direction.

Anisotropic. Orientation (direction) dependency of a measurement or property. As an example, smooth sand has an isotropic texture whereas rippled sand is more anisotropic.

Annex 1 Habitat. A habitat that is protected by the European Union (EU) under the Habitats Directive: 2006/105/EC, Annex 1.

Attribute. Synonymous with feature or descriptor. In this research all attributes are numerical.

Backscatter. A diffuse (non-specular) reflection of signal components back toward the point of origin, i.e. the transmitter/receiver.

Backscatter imagery. Imagery formed by transforming transducer output of quantised signal intensities to a grey scale image.

Ball Vector Machine (BVM.) The Ball Vector Machine (BVM) is a type of SVM that uses Minimum Enclosing Balls rather than Quadratic Programming or other numerical techniques, to solve the convex optimisation problem in the model induction stage.

Base learner. The core learning algorithms(s) used in a machine learning application, for instance, a Naïve Bayes Classifier.

Bathymetry. Relating to depth of the seabed beneath the water surface. An individual depth measurement from an acoustic device such as a Multibeam Echosounder is called a sounding. Multiple soundings can be used to create a Digital Bathymetry Model (DBM). Bathymetry is the underwater equivalent of land surface relief or terrain.

Bedform. A surface pattern induced by fluid flow, such as ripples, in the mobile sediments on the seabed.

Benthic. Relating to the bottom of the sea, for example, benthic organism.

Biogenic reef. A reef structure formed from an aggregation of sea creatures.

Block processing. Block processing breaks a P -pixel \times Q -pixel image into M contiguous, identically sized (n -pixel \times n -pixel) blocks prior to processing, as illustrated in figure G.1. Zero padding is used on the right and at the bottom of the image to enforce this. Multiple computational processes can then be applied to each block sequentially.

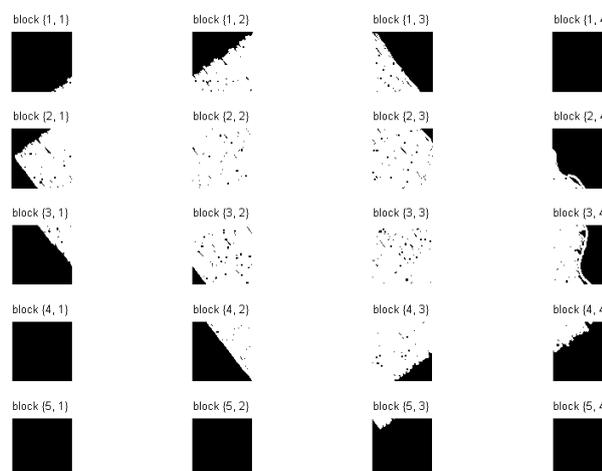


Figure G.1 Block processing illustration

Classification objective. The desired (prescribed) outcomes of the classification process in relation to a combination of input data, ground truth, machine learning processes, target classes and any established conventions for the specific target classification. E.g. Sabellaria discrimination is a different objective to sediment classification. Sediment classification is a vague, generic description, since in practice there are many different methods, with sediment classification objectives dependent on the classification schema, target class(es) and intended purpose/use of the classification results.

Cluster ensemble. Is a collection of at least two unsupervised (clustering) algorithms. The ensemble provides a consensus on the labelling of a pattern instance using label alignment and decision fusion strategies.

Consensus. The collective decision of a group of non-communicating machines, concerning the properties or classification of pattern instances.

Decision fusion. A process and methods for combining the independent machine decisions relating to the properties or classification of a pattern instance.

Digital Bathymetry Model (DBM). A representation of seabed relief (bathymetry). In its simplest form the DBM comprises a grid (raster) of individual depth values at numerous georeferenced locations. It is the underwater equivalent of a Digital Elevation Model (DEM).

Downsampling. Reduces the size of the image input space by transforming neighbourhoods of pixels into single values by application of an operator kernel (filter) to the pixel neighbourhoods.

Ensemble. Often called a committee, it is a collection of at least two independent base learners that are classifiers in the case of a classification ensemble. The ensemble provides a consensus on the class label of a pattern instance using a decision fusion strategy.

Feature. Attribute or descriptor. An individual feature is an element of a feature vector – a collection of features. The numerical value of the feature forms an element in a pattern instance, representing a data sample.

Filter (1.) A method for evaluating features independently of the machine learning algorithm (c.f. wrapper.) A filter approach evaluates the intrinsic saliency of features. There are various kinds of filters. Distance or similarity measures can be used for evaluating individual features. Feature extraction such as PCA is a different type of filter that reduces dimensionality by constructing a feature space that is a linear recombination of the original (higher dimensional) feature space.

Filter (2.) A kernel process that changes the data values over a given neighbourhood or transforms many data values into a single value, such as a down sampling filter. Filters are often applied to remove noise by smoothing the data or for creating features from the data.

Full feature vector (FFV.) When several features are created from a data set, the Full Feature Vector (FFV) is the vector containing all of the constituent feature components (elements.)

Geomorphometry. The extraction of land surface parameters and objects from a DBM or DEM.

Georeferenced. Data or attribute values located at known points on the surface of the Earth, within a coordinate framework, are georeferenced.

Grain (particle) size. The actual size (diameter) of a granular particle of material, such as a sand grain or boulder.

Grain (particle) size distribution. A mathematical model or an empirical distribution relating to the sizes of many particles. The distribution may be parameterised by metrics such as estimates of the average size and spread, skewness and kurtosis.

Ground truth. Ground truth is a type of evidence of some “imaged” physical target and certain characteristics of that target. Ground truth description is a semantic label representing some known characteristic of the target in a seabed region. “Rock” and “Sand” are examples of ground truth labels. Ground truth labels must be available if the image regions are to be classified, regardless of whether a supervised or unsupervised process is used.

Habitat. An environment where an organism lives.

Heterogeneous. Diverse or non-uniform properties.

High resolution. A vague term that can be applied to distinguish the resolution of different sensor types or settings. In general, sidescan sonar can capture higher resolution backscatter than a MBES and is therefore referred to as a “high resolution” instrument.

Homogeneous. Having relatively uniform properties of some description.

Human (manual visual) interpretation. A.k.a interpretative classification, is the process of manually delineating (segmenting) textural regions in sidescan sonar imagery or multibeam bathymetry by visual analysis, using acquired knowledge of the visual properties of the imagery, considered together with ground truth.

Intensity: The pixel intensity value (grey level) is rendered as image brightness, usually on an 8-bit scale of [0 255]. White and black are the end points of the intensity scale and may correspond to either 0 or 255 or some other value range, depending on tiff format settings.

Interpretative classification. See Human (manual visual) interpretation.

Isotropic. Independent of orientation (direction) of measurement.

Kernel (computational kernel.) The square region of coverage of the filter (or other) computational operator applied to the raster node or pixel neighbourhood. E.g. a 5×5 kernel covers 25 grid nodes (or pixels). The functional process implemented by the kernel, e.g. convolution, depends on the filtering or feature creation algorithm.

k-Means clustering. Is a prototype-based clustering method using the cluster centroid as the prototype. K-means groups instances in a feature space by minimising the sum of squares of distances between the instances and the cluster centroids. It is an iterative process, converging when there is no longer any movement of instances between the clusters.

k-NN classifier. The *k*-Nearest Neighbours (*k-NN*) classifier is a non-parametric, lazy learner, as there is no explicit training or generalisation phase. The learning doesn't begin until the query example is presented to the "model" which is defined implicitly by stored data instances and classification rules. In the unweighted distance case, the class label is assigned according to the modal label of the *k*-instances surrounding the test instance.

Lacunarity: A feature kernel capturing information about the "gapiness" of binarised pixel neighbourhoods.

Land surface parameter. A quantitative attribute, or feature such as curvature, the degree of which characterises the surface.

Landform object. A naturally occurring, discrete spatial feature in the land (or seabed) surface. Pockmarks are examples of landform objects that can be rendered in a DBM.

Local (property). A characteristic or feature value that is valid for a small region of the sonar image or seabed, as defined under a kernel of a specified size. For example, local slope of the seabed can be estimated from a raster, over many, say 5m by 5m sub regions, whereas global slope is a single value representing the larger scale trend over the entire region rendered in the raster.

Mask. A binary (or logical) mask is a template, accompanying the image data identifying pixel neighbourhoods that will and will not be processed.

Median filtering. A non-linear filtering method in which the value at the centre of the filter kernel is replaced by the median value of the $k \times k$ kernel neighbourhood.

Morphological filtering. Binary morphological filters use a structuring element (shape) of a prescribed size, to carry out dilation, erosion, opening and closing operations.

Multibeam Echo Sounder (MBES.) An acoustic swath imaging platform which may be mounted on the hull of a survey ship or on a remote underwater vehicle. The MBES captures calibrated (quantitative), georeferenced backscatter and bathymetry data.

Nadir. A direction, pointing vertically downwards beneath the sonar transducer to an imaginary location on the seabed. The seabed in the proximity of the nadir location is an imaging “blind spot.”

Naïve Bayes classifier. Is a multi-class, probabilistic classifier. In the learning stage, training patterns are used to induce a model of the data. In the prediction stage, the model is used to assign class labels and posterior support (certainty) to the test instances.

Neighbourhood (tile.) Image blocks can be sub-divided into a number of contiguous equally sized, square collections of pixels, called neighbourhoods or tiles, as illustrated in figure G.2. The dimension of the tile (in pixel units) is identical to the size of the feature kernel. E.g. a 5×5 computational kernel covers a neighbourhood of 5×5 pixels.

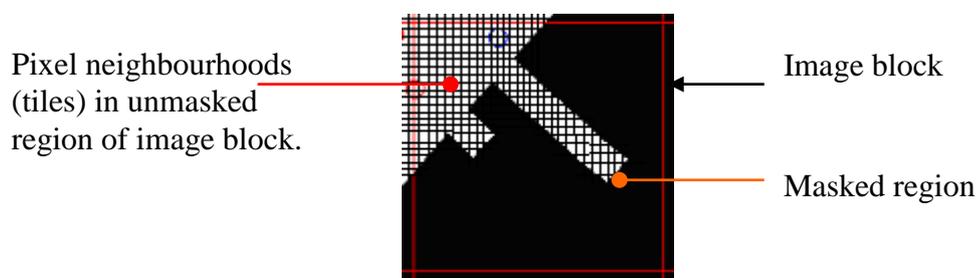


Figure G.2 Tiling of an image block.

Neritic. A zone extending from the low tide mark to a water depth of about 200 m.

Node. A 2-D raster can be imagined to have a local orthogonal grid with each data point uniquely located on a crossing point (node) of the lines in the imaginary grid.

Particle size analysis (PSA.) PSA is a technical process used to determine the size and distribution of particles (grains) in physical ground truth samples.

Pattern. The word “pattern” is often used interchangeably with feature vector. A pattern is a numerical vector of feature values, created by applying different feature kernels to the same image neighbourhood.

Phi (ϕ) Scale. A quantitative measure of unconsolidated grain sizes using the base 2 logarithm of the grain diameter. It is a modification of the Wentworth scale.

Pockmark. A crater-like, specific landform object formed in the seabed by fluid venting through soft sediments.

Qualitative imagery. Backscatter imagery - mosaic or waterfall imagery which has not been calibrated to represent an accurate measure of the seafloor backscatter response at georeferenced points. Most sidescan sonar imagery is qualitative.

Quantitative imagery. Grey levels correspond to accurately calibrated values of backscatter intensity at georeferenced pixel locations. Backscatter from a Multi beam echosounder may be regarded as quantitative, compared to that from a sidescan sonar.

Random sampling locations. In an unsupervised approach, no assumptions are made about the relationship of any image regions to known properties of the seabed (ground truth.) The image can only be segmented (not classified) when ground truth labels are unavailable. Many virtual ground truth points of unknown class label are therefore created in random locations in the mosaic, for sampling, pattern generation, clustering and subsequent model creation using cluster labels as tentative class labels.

Raster. A 2-D array (grid) of data values at regularly spaced grid node locations. All image data used in this work are rasterised.

Resolution (1.) The relationship of the spatial (sampling) distances between pixels or grid points to the true (ground) distances. High resolution imagery shows detail in relatively small ground objects. At a lower resolution the objects may not be imaged or rendered. Image resolution may be defined as the number of pixels per unit ground length (the reciprocal value of ground length per-pixel can also be used.) Image resolution may be anamorphic, i.e., different resolutions in mutually perpendicular directions.

Resolution (2.) Kernel resolution is the number of pixels covered by the width of the (square) computational kernel. E.g., a kernel covering a 5×5 pixels neighbourhood has a resolution of 5 pixels. Kernel resolution defines the resolution of the output class map. When the kernels are applied to contiguous tiles, inputting a mosaic image with an image (ground) resolution of 2 metres per pixel, using a 5×5 kernel resolution, will generate a class map resolution of 10 metres per pixel.

Rugosity. Is a measure that can be used to characterise the variability in height or “ruggedness” of the seabed surface.

Sabellaria Spinulosa. A tube building worm that can form biogenic reefs. Such reefs are designated as Annex 1 habitats.

Saliency. A measure of the importance of a particular feature or subset of features in the context of specific data and classification objective(s.) When saliency is measured in a wrapper, using classification accuracy as a metric, an individual feature yielding a relatively high accuracy compared to the others could be described as the most salient individual.

Scan line. A one-dimensional array of quantised signal values output from a sidescan sonar receiver, representing returns from a single acoustic pulse in the across-track direction. A waterfall image comprises of many such scan lines collected sequentially in the along-track direction.

Seabed classification. A process of delineating different regions of the seabed (segmentation) and assigning class labels to the regions based on a variety of real samples (ground truth), expert knowledge and proxies (acoustic imagery, for instance).

Seabed class map. A (usually colour coded) map of the seabed delineating different regions according to prescribed or perceived properties of the region.

Segmentation. The partitioning of an image into different, relatively homogenous regions without assigning any class labels to the regions.

Self organising map (SOM.) The Self Organising Map (SOM) is a type of unsupervised Artificial Neural Network (ANN.) It is a prototype based clustering method using topographically prescribed centroids. In a 2-D map, cell centroids that are relatively close together represent data points that are similar in some way, in the input space.

Sensitivity. See Accuracy.

Sidescan sonar. An acoustic, swath imaging platform usually towed behind a survey ship on a long cable. A simple towed sidescan sonar captures acoustic backscatter only. Imagery from the sidescan sonar is qualitative and can be interpreted in mosaic and waterfall forms.

Sonar mosaic image. An acoustic image of a surveyed region of the seabed formed by merging image and ancillary data from several individual survey lines using a processing pipeline and dedicated software.

Sonar waterfall image. Backscatter imagery that has not been georeferenced (and may not have had any corrections applied). Each across-track scan line forms a single row in the image data.

Specific geomorphometric analysis. The analysis of a specific landform object in terrestrial or seabed terrain.

Specificity. See Accuracy.

Supervised process: In a supervised learning process, ground truth points (or textures identified by humans) are used as sample locations (training regions), to generate the classification model. The output of a supervised classification is a class map.

Support Vector Machine (SVM.) The Support Vector Machine (SVM) is a binary classifier with a classification hypothesis (model) whose decision boundary is a discriminatory hyperplane. It is a maximal margin classifier. An optimal model with the greatest separation between the decision boundary and the n -dimensional patterns in the training data set is induced. The optimal hyperplane is represented using a subset of training patterns known as Support Vectors. In the prediction stage, unseen pattern instances are discriminated according to the side of the hyperplane they lie on.

Swath sonar. An acoustic imaging sonar (Sidescan sonar, Multibeam Echosounder) capable of capturing a wide swath of data either side of the Survey Vessel track.

Target recognition. In this context, the recognition and delineation of specific distributed natural targets in a DBM or backscatter imagery. A colony of Sabellaria Spinulosa is an example of a distributed natural target that imparts a distinctive textural signature in backscatter imagery.

Texture: There is no universally accepted definition of what texture is. Qualitatively, image texture can be described as “rough”, “smooth”, “isotropic”, “anisotropic”, “regular”, “irregular” and so on.

Unsupervised process: In an unsupervised process, the image is randomly sampled and a classification model is induced (see, Randomly sampled locations.) The output of the unsupervised process is a segmentation of the mosaic. The segmentation can be transformed into a classification by manually (or automatically) assigning semantic ground truth labels to the different segmented regions.

Well-sorted. Sediments comprise mixtures of different particles. Well-sorted refers to a grain size distribution with a relatively low standard deviation, i.e., the particles sizes are more uniform.

Wentworth scale. A scale for describing the actual size of sedimentary particles in millimetres.

Wrapper. A wrapper evaluates features by using a harness containing the base learning algorithm(s) and an objective function, commonly classification accuracy. Each individual feature or feature subset can be used to induce a different model from the learning algorithm using a set of training instances.

Acknowledgements

Funding

The research for this thesis was supported by the Grant (R17336) jointly funded by the Engineering and Physical Sciences Research Council (EPSRC), the University of East Anglia (UEA) and Gardline Marine Sciences Limited (Gardline Geosurvey) of Great Yarmouth, UK.

UEA supervisory team

I would like to express my gratitude to the academic supervisory team at the UEA. In particular I would like to thank my primary supervisor, Dr. Wenjia Wang for his invaluable guidance and encouragement throughout the duration of the research programme. I would also like to thank Dr. Richard Harvey for his input and for arranging for me to work with visiting Professor, Francesco Bianconi. Thanks also to Dr. Ben Milner for his helpful advice.

Industrial stakeholders (Gardline Marine Sciences Limited)

Thank you to Gardline Marine Sciences Limited (Gardline Geosurvey) of Great Yarmouth, for co-funding the research, in collaboration with the UEA. Particular thanks to Roger Birchall, Ken Games and Dave Mann for their input to the project and for arranging the provision of some of the data used.

Gardline staff

Thank you to the following staff at Gardline for assisting with certain aspects of the preparation and interpretation of some of the data used in this thesis (ordered alphabetically by surname); Chris Buckmaster, Rosie Harrison, Kelly Johns, Andrew Kerr, Ben Sansome, John Sperry, Stephane Theurich.

Other individuals (ordered alphabetically by surname)

Thank you to Dr. Tim le Bas, of the National Oceanography Centre (NOC), Southampton for his encouragement in submitting my first conference paper that was subsequently presented at the MAGIC conference in Italy, May 2009. Thank you to Dr. Valérie Bellec of The Geological Survey of Norway (NGU) for arranging the provision of data for the case study on the pockmarks and for helpful comments on the preparation of the resulting paper and poster presentation. Thank you to Prof. Francesco Bianconi, whom I worked with for several weeks whilst he was on a Sabbatical at the UEA. There were many stimulating discussions on textures and texture classification and our work resulted in a conference paper, presented by Prof. Bianconi at IMVIP, Dublin, 2011. Thank you to Dr. Chris Collins of the High Performance Computing Centre at the UEA for his help on using the computer cluster. Thank you to Dr. Marcus Diesing, from the Centre for Environment, Fisheries and Aquaculture Science (CEFAS), of Lowestoft for helping arrange for me to deliver a presentation at the CEFAS centre in Lowestoft and for his interesting and informative discussions on a variety of seabed/habitat classification topics. Thank you to Dr. Margaret Dolan, also of the NGU who worked on the discrimination of pockmarks (in the data provided by Dr. Valérie Bellec) using a GIS package. Shape files and other useful information had been generated that were useful during the course of my own research into pockmark discrimination. An apology and a thank you is extended to any others who may have assisted in some way whose names are not mentioned here.

Introduction

Chapter 1

Contents

1.1 Background to the thesis and research project

1.1.1 General background

1.1.2 Industrial collaboration

1.2 General aim, objectives and scope of the research

1.3 Importance and relevance of the work

1.4 Summary of collective novelty and contributions of the research

1.5 Summary of novelty, main contributions and published work by case study

1.5.1 Case study 1: Pockmark discrimination

1.5.2 Case study 2: Features for Sabellaria discrimination

1.5.3 Case study 3: Feature evaluation and ranking methods

1.5.4 Case study 4: Unsupervised approach on real-world mosaic imagery

1.5.5 Related research (not included in the thesis) published during the project

1.6 Thesis structure

The purpose of this introductory chapter is to briefly present the background to the thesis and the research project, to highlight the aims, importance, novelty and contributions and to describe the document structure.

In section 1.1, the general background and industrial collaboration is outlined. Section 1.2 states the aims, objectives and scope of the research. The importance and relevance of the work is outlined in section 1.3. Section 1.4 summarises the collective novelty and main contributions of the research. Section 1.5 summarises the novelty, main contributions and publications by case study. The structure of the thesis by chapter is stated in section 1.6.

1.1 Background to the thesis and research project

1.1.1 General background

This thesis concerns feature-based machine learning approaches to the discrimination and mapping of qualitative natural targets (or regions) in seabed imagery. A natural target can be considered a relatively homogeneous and visually distinct region (or regions) of interest on the seabed. Seabed imagery, for the purposes of the thesis, is rasterised sidescan sonar backscatter imagery or rasterised bathymetry i.e., a digital bathymetry model (DBM) from multibeam echosounder (MBES) depth soundings.

As the thesis is situated in the Computing Science domain, the underlying theme is the investigation of ideas for algorithms and processes in virtual experimental spaces, with a bias towards feature creation and evaluation for the target-specific application contexts. The study is not about the science of acoustic seabed classification or of backscatter image formation. It focuses on computational methods and processes that could potentially be used to assist a human with certain interpretative, qualitative data driven tasks, through automation or semi-automation using machine learning systems. For instance, manually (visually) identifying and mapping pockmarks in a DBM is a slow, labour intensive and therefore expensive undertaking. Furthermore, the results of a human inspection and mapping are inconsistent due to intra- and inter-rater variability. It would be very useful and profitable then, to engineer new tools and processes using machine learning technology to carry out these types of task reliably and expediently.

As an applied, industry-based project, it is predominantly about “problem solving research.” Established computing science research methodologies are used in the thesis, such as case study, proof of concept and empirical investigation. Sidescan sonar imagery and ground truth are qualitative and uncertain. Therefore, the research, although empirical, uses mixed methods and is exploratory, descriptive and evaluative in nature.

Within the immense domain of machine learning and the limitless scope of application scenarios it is only possible to consider a very small subset of ideas and applications, given the available time and resources. The novel ideas and applications are explored

through four coherent and progressive case studies, linked by common technical methods and forming the core of the experimental work presented in the thesis.

1.1.2 Industrial collaboration

This research project was undertaken in a collaborative arrangement between the University of East Anglia (UEA), School of Computing Sciences and Gardline Geosurvey, a division of Gardline Marine Sciences Ltd. (GMSL) of Great Yarmouth, UK. The project was supported by the Grant (R17336) jointly funded by the Engineering and Physical Sciences Research Council (EPSRC), UEA and Gardline Geosurvey.

Potentially, some of the algorithms, processes and ideas proposed and investigated in the thesis may later be developed into commercial/industrial applications by the UEA and GMSL project stakeholders.

1.2 General aim, objectives and scope of the research

Aim

The general aim of the research is to devise and investigate novel processes and methods that could potentially be used as part of a machine learning system for assisting humans with the subjective (interpretative) discrimination and classification of natural targets in seabed imagery. The general objectives reflect the aims of the case studies.

Comprehensive details of the aims, objectives and scopes of the individual case studies are provided in chapter 4 (section 4.2.2.)

General objectives, corresponding to the case studies 1-4

1. Propose and investigate a novel machine learning process for pockmark object discrimination and boundary mapping in a DBM.

2. Investigate if the machine discrimination of Sabellaria textures in sidescan waterfall and mosaic imagery is a tractable task. Determine suitable features, feature configurations and parameters for this task.
3. Design a framework for evaluating the robustness of distance measures for feature evaluation and ranking. Investigate properties of a novel committee of distance measures for feature evaluation and ranking on sonar imagery.
4. Investigate how to apply an unsupervised machine learning processes to larger real-world sonar mosaic imagery. Devise novel methods and processes to meet the challenges of this task. Qualitatively evaluate the merits of a fully automated supervised and novel, hybrid unsupervised approach (including an unsupervised ensemble approach) to a target discrimination task in a real-world mosaic image.

Scope

Technical methods scope is limited to feature based machine learning processes and methods summarised at the end of chapter 3. The problems are data driven and the data scope is confined to sidescan sonar mosaic imagery, sidescan sonar waterfall imagery and MBES bathymetry in the form of a DBM. All data are rasterised and processed. The problems and data are considered at a phenomenological, human interpretative level. The application scope concerns two identified topics for investigation in chapter 2; pockmark discrimination and Sabellaria discrimination. The sediment type discrimination task in case study 4 is a contingent application.

1.3 Importance and relevance of the work

Manual, interpretative discrimination and mapping of natural targets in seabed imagery is a frequent and necessary task in the marine surveying industry. The work is important as it addresses some of the current issues and presents solutions to some of the challenging technical problems associated with industry-based, machine-assisted interpretative discrimination of specific targets in seabed imagery. It is complementary to the quantitative modeling (remote characterisation) approaches that are dependent on

empirical and physics based constructs of acoustic backscattering from seabed sediments.

Despite the advances in automated acoustic seabed classification, there remain many discrimination and classification problems that still depend to a large degree on expensive and time consuming office-based human interpretation and are likely to do so for the foreseeable future. Given the continuing demand for seabed surveys and for the types of class maps derived from the human-centred processing of data captured in these surveys, the research is highly relevant as it applies directly to identified “bottle-neck” stages in the commercial, interpretative processing flows. The importance and relevance are placed in a wider context in chapter 2, section 2.

1.4 Summary of collective novelty and contribution of the research

Although the experimental work in the thesis comprises four relatively self-contained case studies, they are not independent. In fact, the case studies are coherent and progressive. So, in addition to contributing individually, as outlined in the next section (1.5) there is also a collective contribution and novelty.

The overarching technical issues linking the case studies primarily concern the features that can be used to represent the targets and the reliable measurement of their saliency for the specific data driven task. It presents a significant conceptual and technical challenge. All other things being equal, effective discrimination of natural targets in seabed imagery depends primarily on an appropriate and reliable choice of feature generators and their parameterisation for the target class. This issue is addressed in each case study and exemplified in particular, in case study 3. The collective contribution of the work is therefore to the machine learning domain of feature evaluation, as applied to the specific industrial problems and data described herein. Novelty arises not only in the methods but in the niche focus of the work, i.e., machine assisted discrimination of subjective natural targets, for which there are currently no established approaches other than manual human interpretation.

1.5 Summary of novelty, main contributions and published work by case study

The research carried out in the course of project has resulted in five peer reviewed publications. Three of these publications have arisen from the case studies in chapters 5-8. This section outlines the novelty and contributions of the case studies and states the published work.

1.5.1 Case study 1: Pockmark discrimination

This work contributed to the wider domain of DBM analysis and particularly to the feature based discrimination of specific landform objects in a DBM. A novel feature based machine learning approach to automating the identification of pockmarks and mapping their boundaries was presented. The process maps the boundaries of ≈ 2000 pockmarks in seconds - a task which would take days for a human to complete.

Some of the research work carried out in this case study resulted in the publication of a conference paper:

Harrison, R., Bellec, V., Mann, D. and Wang, W. (2011) A new approach to the automated mapping of pockmarks in multibeam bathymetry, IEEE International Conference on Image Processing (ICIP).

The accompanying poster presentation is included in Appendix 1.

Proportional contributions to the paper are: R. Harrison 85%, V. Bellec 5%, D. Mann 5%, W. Wang 5%.

1.5.2 Case study 2: Features for Sabellaria discrimination

The main contribution of this work was to the novel task of machine discrimination of Sabellaria textures in sidescan sonar imagery. It was the first study of its kind on this specific application and directly addresses another industrial problem. It was not known whether this task would be tractable or what features should be used. The study

demonstrated the tractability of the problem and identified the Gabor filter bank as a useful feature creation method for discriminating Sabellaria in waterfall imagery. Filter bank configurations and filter parameters were investigated in depth on mosaic imagery. Suitable configurations and parameter ranges were identified.

The research on texture feature creation methods for Sabellaria discrimination in waterfall imagery led to the publication of a conference paper:

Harrison, R., Bianconi, F., Harvey, R. and Wang, W. (2011) A texture analysis approach to identifying Sabellaria spinulosa colonies in sidescan sonar imagery, Proceedings of the Irish Machine Vision and Image Processing (IMVIP) conference, Dublin.

Proportional contributions: R. Harrison 65%, F. Bianconi 25%, R. Harvey 5%, W. Wang 5%.

1.5.3 Case study 3: Feature evaluation and ranking methods

This work contributed to the application domain of sonar imagery feature evaluation and to the machine learning domain of feature evaluation and ranking methods, specifically, the measurement of the robustness of feature evaluation methods. A novel framework for measuring the robustness of distance measure feature evaluation methods was devised. Two novel consensus approaches to feature ranking with multiple distance measures were proposed. Experimental results showed that the consensus approaches could improve robustness over a range of feature parameterisations and various seabed texture classification tasks in sidescan sonar mosaic imagery.

The research on the ranking reliability of distance measures resulted in a journal publication:

Harrison, R., Birchall, R., Mann, D. and Wang, W. (2012) Novel consensus approaches to the reliable ranking of features for seabed imagery classification, International Journal of neural Systems, 22 (6), DOI: 10.1142/S0129065712500268.

Proportional contributions: R. Harrison 80%, R. Birchall 5%, D. Mann 5%, W. Wang 10%.

1.5.4 Case study 4: Unsupervised approach on real-world mosaic imagery

The work in this case study brings together a number of earlier ideas and contributes to the application domain of machine discrimination of qualitative targets in sonar imagery. A novel, hybrid unsupervised process and methods were proposed for the discrimination of textural targets in qualitative, real-world sidescan sonar mosaic imagery. The novel methods included a pre-clustering stage for inducing a sub-optimal probabilistic unsupervised model and a method for combining unsupervised classifications from independent feature channels. On the data used in the case study, a single-channel unsupervised approach using Gabor filter bank features produces a plausible partitioning of the sediment classes in the imagery over a range of resolutions. Although originally intended for Sabellaria discrimination, the success of the process and methods in the application context demonstrates the flexibility of the approach and the potential versatility for application to a range of target discrimination problems.

1.5.5 Related research (not included in the thesis) published during the project

[1] *Distance measure committee for feature evaluation: Application to sonar imagery*

A novel committee of distance measures was proposed as an objective function for feature evaluation. The idea was motivated by the fact that different distance measures yield conflicting feature ranking results. In the committee, distances can be combined to produce a consensus rank score. As a test case, parameterisation of a grey level co-occurrence matrix, for identifying textures peculiar to Sabellaria colonies was considered. A strong correlation ($r = 0.94$) is found between saliency scores and classification accuracies using the committee.

Conference publication details:

Harrison, R., Birchall, R., Mann, D. and Wang, W. (2011) A novel ensemble of distance measures for feature evaluation: Application to sonar imagery, IDEAL 12th Int. Conf. Proc., Lecture Notes in Computer Science, 6936, 327–336.

Proportional contributions: R. Harrison 80%, R. Birchall 5%, D. Mann 5%, W. Wang 10%.

[2] *Towed cable, sonar and acoustic beacon model*

A novel hybrid model, comprising a static towed instrument-cable system combined with an acoustic model for the attached ultra-short baseline (USBL) positioning beacon was proposed and was believed to be the first of its kind. Initial numerical experiments indicated the model results were consistent with observed field data.

Conference publication details:

Harrison, R. (2009), A spatial model of a towed cable system and USBL acoustic beacon, in Società Geologica Italiana, proceedings of the International conference on seafloor mapping for geohazard assessment, Eds., Chiocci, F., Ridente, D., Casalbore, D. and Bosman, A., 7, 59-62.

1.6 Thesis structure

- Chapter 2 establishes the wider academic, industrial and practical context of the research and examines some of the motivating issues and applications. The application scope is narrowed to two interpretative discrimination tasks, pockmarks and Sabellaria.
- Chapter 3 reviews some of the technical approaches (i.e., methods and processes) related to the thesis and identifies some underexplored gaps.

-
- Chapter 4 describes and discusses the research analysis and design. It also includes a consideration of evaluation issues related to qualitative tasks in the seabed classification domain.
 - Chapter 5 is the first of four case studies. A novel approach to the discrimination of pockmark objects in multibeam bathymetry is proposed and evaluated.
 - Chapter 6 presents an investigation into the tractability of the novel task of automated Sabellaria texture discrimination in sidescan sonar imagery.
 - Chapter 7 considers the more abstract problem of the robustness of distance measures for feature evaluation and ranking. A framework is created for evaluating the distance measures and a novel committee proposed and evaluated.
 - Chapter 8 brings together several ideas from the earlier work and concerns the discrimination of qualitative target regions in real-world mosaic imagery. A novel hybrid unsupervised process is proposed and evaluated.
 - Chapter 9 evaluates the research overall, assessing the extent to which the aims and objectives have been successfully achieved and how the research was managed.
 - Chapter 10 summarises the conclusions from the individual case studies. General conclusions are also drawn, a general recommendation is given and two mini-proposals for further work are presented.

Research context

Chapter 2

Contents

2.1 Introduction

2.2 Why we need information about the seabed

2.3 Data acquisition for seabed classification: Methods and issues

2.3.1 Overview

2.3.2 Backscatter imagery

2.3.3 Bathymetry

2.3.4 Ground truth

2.4 Other pertinent issues

2.5 Summary

2.1 Introduction

The research presented in this thesis concerns the design of novel feature based computational methods and processes for assisting humans in expediting certain qualitative, manual (visual) seabed classification tasks on processed data. As a Computing Science thesis, it is not about the discipline of seabed classification *per se*. However, since the research is applied to solving specific problems within this domain, it is necessary to have a wider understanding of the relevant background to fully appreciate some of the issues involved and to provide some justification and focus for the research. Hence, at the end of this chapter, the context limited application scope for two of the case studies can be defined. The purpose of the chapter then, is to;

- (1) elucidate the problem areas, identifying some important issues for further investigation,
- (2) establish the need for the research, the commercial motivation and define the application scope,
- (3) position the research within the application domain and the wider technical and environmental context.

The chapter further serves to highlight the immense scope and cross-disciplinary nature of seabed classification tasks and processes and introduces some of the terminology associated with the application domain. The word “*classification*” is used in this context to mean the production of a seabed class map of some description, *i.e.* the depiction of targets and their properties on a two-dimensional map, over a region of the seabed at a particular scale. Figure 2.1¹ is an example of an end product of the processing pipeline, a class map.

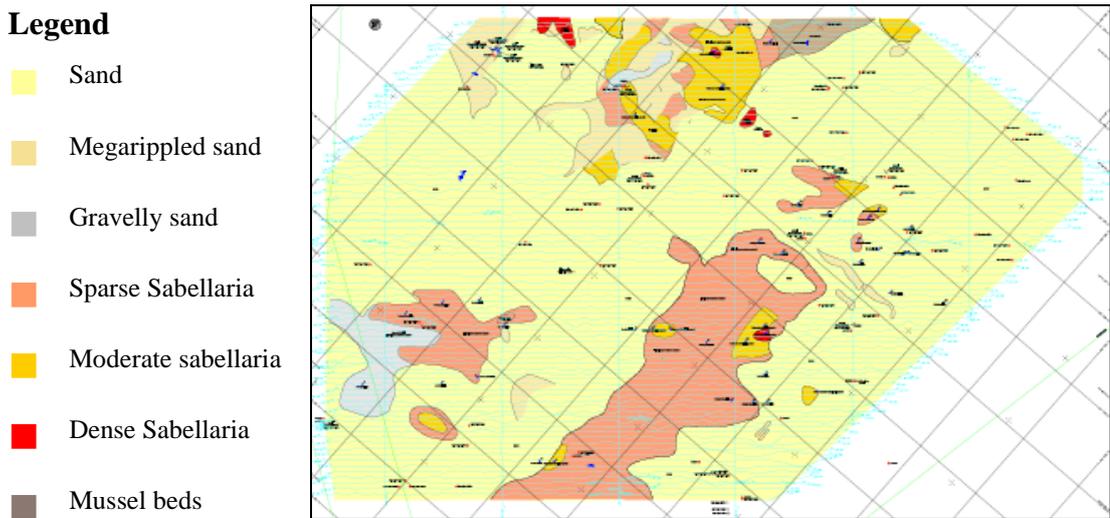


Figure 2.1. A class map of a neritic habitat, indicating different sediment types, species assemblages and geomorphic objects as homogeneous colour-coded regions in a georeferenced framework.

The map represents a neritic habitat and indicates different sediment substrates (sand, gravelly sands), species assemblages (*Sabellaria spinulosa*, mussel beds) and geomorphic objects (dunes or megarippled sand) as homogeneous, colour-coded areas within a georeferenced framework. The graticule squares are 1 km × 1 km and very small sized (10-100 m) to small sized (100 m – 1 km) mesohabitats are resolved, according to the scales in Greene *et al.* (1999). “Habitat” can be defined as “a place where a microorganism, plant, or animal lives” (Begon *et al.*, 1990).

Seabed classification is a broad based, challenging and developing field of study. Hamilton’s (2005) bibliography lists a non-exhaustive selection of numerous academic

¹ Gardline Geosurvey, 2007. Client and location details have been removed.

and technical publications on the dozens of disciplines involved. Since 2005, several other related studies have been published on for instance; acoustic remote sensing techniques, different geographic regions of the World, specific habitat types, shallow and deepwater environments along with many approaches to data processing, remote characterisation and the automation of seabed target recognition and classification tasks.

Class map production can involve a combination of human and machine functions in the processing pipeline. A five-stage sequence, from data acquisition through to the final mapping product is summarised in figure 2.2.

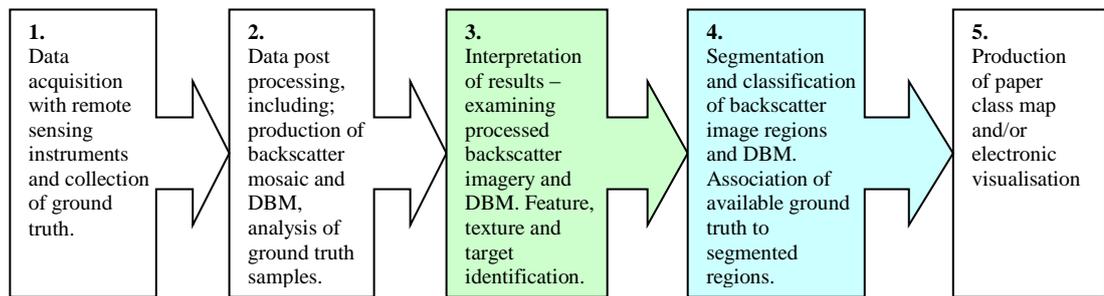


Figure 2.2. A simplified processing flow for class map production. The thesis focuses on aspects of feature based machine learning approaches applied at stages three and four.

In this simplified sequence, data acquisition (1) is followed by a data processing stage to convert the raw data into meaningful results or representations. Stage (2) typically includes tasks such as estimating distributions of sediment grain sizes by sieving sediments retrieved during the physical sampling campaign. Computationally intensive work, such as processing depth soundings to produce a rasterised Digital Bathymetry Model (DBM) is also carried out at this stage.

Interpretation of the DBM and sonar imagery may be performed manually by a trained human interpreter or in some cases, automatically by a machine learning system. As an example, the human visual system (HVS) texture discrimination in the sonar imagery can be replaced by an automated or semi-automated feature based segmentation/classification system at stages (3) and (4). Target regions of the bathymetry and imagery are identified and delineated, then semantic labels assigned at stage (4), according to class information derived from the ground truth. Finally, the end product, a paper class map or electronic, virtual representation is produced at stage (5)

in readiness for delivery to the client. The thesis focuses on feature based machine learning and other computational methods that can be used in novel ways to support the interpretative process at stages three and four of the above sequence. The problems are approached through a set of context-limited case studies. Aspects of the five stages in the process of figure 2.2 are considered further in the remainder of this chapter, which is organised as follows.

Section 2.2 explains the importance of the seabed and why we need information about its nature, in the form of mapping products (class maps). Section 2.3 discusses commonly used techniques for acquiring seabed data through remote sensing and ground truth campaigns. Some important issues in the interpretation, mapping and classification process are outlined in section 2.4. The main points of the chapter and the application scope of the thesis case studies are summarised in section 2.5.

2.2 Why we need information about the seabed

There are many reasons for wanting to find out more about the bottom of the seas and oceans. Mostly hidden from view, the complex seascape supports a diverse range of biotopes and a rich variety of marine life. “*Mapping European Seabed Habitats*” (MESH)² and “*Habitat Mapping for Conservation and Management of the Southern Irish Sea*” (HABMAP)³ are just two examples of recent large-scale European projects, dedicated to gaining a better understanding of the types, extents and distributions of seabed habitats. This mapping is essential, if we are to develop an insight into the health and dynamics of marine ecosystems and how they can be conserved or protected from the detrimental affects of human activities.

Anthropogenic activities and natural processes impact on seabed habitats, yet conversely, distant processes taking place at the seafloor can have disastrous effects on human populated coastal regions. For instance, submerged mass failures (Billi *et al.*, 2010) can trigger destructive tsunamis, posing a significant threat to vast stretches of coastline. A national project currently being undertaken in Italy “Marine Geohazards

² <http://www.searchmesh.net/default.aspx> [accessed 14-02-2012]

³ <http://habmap.org/> [accessed 14-02-2012]

along the Italian Coasts” (MAGIC)⁴ aims to map the hazardous geological regions along submarine continental margins of Italy. Knowledge from this project can be used to aid the development of more effective tsunami mitigation strategies in civil defence programmes.

Perhaps surprisingly, habitats at the bottom of the seas around the UK and Europe are intimately connected to our quest for renewable energy. Situated within complex Europe-wide and national legislative frameworks, the European Union (EU) member states are mandated to generating a proportion of their energy from renewable sources by 2021⁵. For the UK, this target is set at 15 %⁶. Offshore wind farms can contribute significantly to meeting this target but site selection needs careful consideration and must be sympathetic to the marine ecosystem. Biogenic reefs, formed by the tube-building worm *Sabellaria Spinulosa* (Sabellaria) are protected by the EU under their habitats directive⁷. Destruction of these protected areas must be avoided and site investigation therefore includes identifying and mapping any Sabellaria colonies, especially reef-like structures. It is essential that site planners and contractors are aware of the locations, extents and nature of any protected (or potentially protectable) habitats in their operational areas.

The offshore mining industry and in particular the oil and gas exploration and production sector is another important consumer of seabed mapping products. During the exploration phase for oil and gas, it is useful to identify geomorphic objects such as pockmarks, as they can be surface indicators of hydrocarbon reserves, deeper in the underlying geology. They are also associated with shallower trapped gases and seabed instability, potentially serious hazards to drilling operations and seabed infrastructure, respectively (Judd and Hovland, 2007, p 362-371). Pockmarks and numerous other geomorphic objects such as canyons, dunes and outcrops are rendered in a DBM, the submarine equivalent of a Digital Elevation Model (DEM).

An up-to-date knowledge of the seafloor is therefore critical to the success of all offshore projects where mechanical or other forms of seabed interaction are involved. It

⁴ <http://www.magicproject.it/> [accessed 14-02-2012]

⁵ Directive 2009/28/EC, Article 3.

⁶ Directive 2009/28/EC, Annex 1.

⁷ Directive 2006/105/EC, Annex 1.

is also important to organisations engaged in conservation, management and monitoring activities. From a commercial perspective, mapping products provide persistent revenue streams for marine geophysical and hydrographic surveying companies such as Gardline Geosurvey⁸. Surveys and maps are commissioned and purchased largely by clients in the offshore industrial sector. In addition to surveying for new projects, repeat mapping of a region is often required, since the high-energy marine environment creates transient seabed regimes with mobile sediments and ephemeral habitats and landform objects. Thus, any class map is an uncertain, grossly simplified static representation of dynamic, interchanging biotic and abiotic elements in the seascape.

2.3 Data acquisition for seabed classification: Methods and issues

2.3.1 Overview

Inaccessibility of the seafloor makes it much harder to map and characterise, compared to our terrestrial environment. The conductivity and turbidity of seawater rapidly attenuates and scatters electromagnetic signals, particularly in the visible spectrum, see, for example, Apel (1987), p 588-591. Excluding relatively short imaging ranges, under low turbidity conditions, the practical usage of optical devices for remote sensing applications in the underwater environment can be quite limited. On the other hand, acoustic signals exhibit excellent propagation characteristics in seawater, surface sediments and the deeper geology of the seafloor. Acoustic remote sensing has therefore become the principal technique for capturing an assortment of data about the seafloor and its properties at various spatial scales and resolutions. An acoustic image of the seascape, rather like an aerial photograph of a terrestrial landscape, is a surrogate for the real, physical surface and all that is contained within it. Backscatter imagery is formed by rendering the diffuse acoustic radiation scattered from the seabed back toward the point of origin, i.e., the transducer that transmits and receives the radiation. Grey-levels of the image pixels correspond to the received intensity (radiosity) values from sonified patches of seabed. Acoustic echoes or soundings are commonly used to determine depth. Rasterised depth soundings can be rendered as a georeferenced DBM showing accurately positioned relief of the seabed surface and depths, relative to a vertical datum.

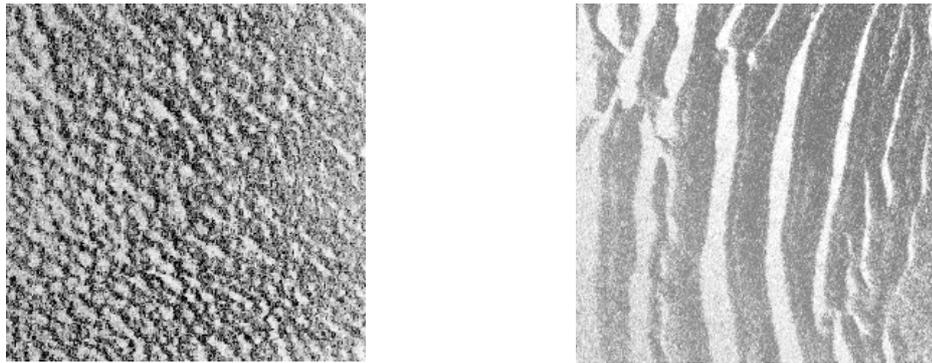
⁸ <http://www.gardlinemarinesciences.com/page/gardline-geosurvey/> [accessed 14-02-2012]

Two commonly used acoustic remote sensing devices are the simple towed sidescan sonar (sidescan) and the more sophisticated, keel-mounted multi beam echosounder (MBES). The sidescan and MBES are sometimes referred to as “imaging sonars” as they are both capable of producing a greyscale backscatter “image” of the seafloor. Sidescan and MBES can map a wide area of the seabed called a swath. The sidescan records acoustic backscatter only, whereas the MBES records accurate depth soundings and backscatter. Data used in this thesis are primarily rasterised (square-gridded) backscatter imagery from the sidescan and MBES and bathymetry from the MBES.

Ground truth captured during the survey forms an important component of the data used in the process, as it provides evidence of what is actually on or near to the surface of the seabed. Therefore, it is a means of validating a speculative interpretation of the acoustic facies in the backscatter imagery. Ground truth data often takes the form of physical samples or near-field video footage. The following two subsections outline and discuss acoustic imaging and ground truth methods, pertinent to this thesis.

2.3.2 Backscatter imagery

Seabed backscatter is a diffuse, highly scattered form of acoustic radiation containing noisy information about the interaction of a transmitted acoustic pulse with its propagation environment. There is no general theory of backscattering from seabed sediments, although several empirical and mathematical models have been developed which may comprise; (1) roughness scattering, due to the roughness of the seabed surface and (2) volume scattering, due to penetration and re-radiation of acoustic energy in the heterogeneous sediments. See, for instance, Urick (1967), Novarini and Caruther (1998), Chotiros (2006), Holliday (2007), Jackson and Richardson (2007) and Kloser (2007).



(a) Characteristic “lumpy” Sabellaria textures. (b) Dunes (anisotropic texture). Image scale is the same as in figure 2.3 (a). The wave number for each pixel represents a ground area of approximately 20×20 cm. The image area covers about 2500 m^2 of seabed. The wave number for the widest spaced dunes is about 0.15.

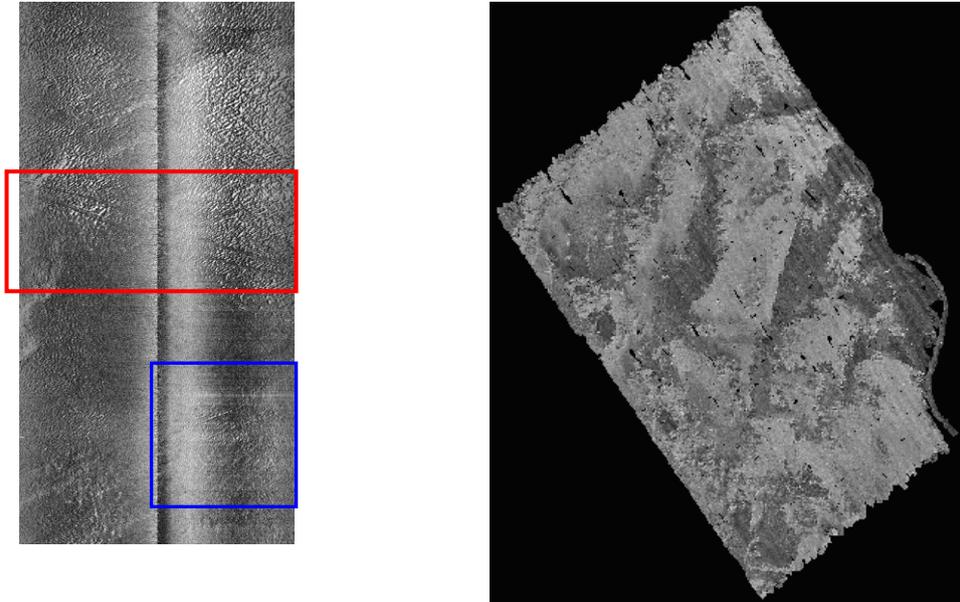
Figure 2.3. Examples of two different textures rendered in backscatter imagery from a sidescan.

A sidescan captures qualitative backscatter, since the recorded signal comprises relative amplitudes. Sidescan imagery contains useful information about the seabed surface and texture, providing a means for discriminating acoustic facies based on textural similarities and contrasts and regions of different intensity. The different image regions can be correlated to the morphology and in some cases the material properties of real targets on the seabed. For instance, Sabellaria colonies are often rendered with a distinctive “lumpy” texture, recognisable to a human interpreter, as shown in figure 2.3(a). In this image region, the ground resolution of each pixel is approximately $20 \text{ cm} \times 20 \text{ cm}$ or 5 pixels per metre. The image covers a seabed area of about 2500 m^2 . Figure 2.3 (b) shows an acoustic image of submarine dunes (a.k.a sand waves), at the same scale. The texture, unlike that of Sabellaria, is strongly anisotropic, with dark and light bands orientated from top to bottom of the image. The wave number (spatial frequency) for the dunes with the widest spacing is approximately 0.15.

Sidescan imagery is initially captured by the transducer as a “waterfall” image (figure 2.4 (a)). Waterfall sections may be processed and merged together, to create a sonar “mosaic” of a larger survey area. Figure 2.4 (b)⁹ shows a backscatter mosaic of seabed off the coast of Florida, mapped by the United States Geological Survey (USGS). A

⁹ <http://pubs.usgs.gov/of/1999/of99-396/> [accessed 14-02-2012]

waterfall is built up from consecutive individual scan lines as the instrument platform translates in the along-track direction. The pixels are not georeferenced and may not have had any geometric or radiometric corrections applied to them at this stage. Waterfall imagery is sometimes used for human interpretation as the resolution is better compared to a down sampled mosaic image.



(a) Waterfall segment – part of a survey line. Unlike the mosaic it is not georeferenced and aside from water column removal and slant range corrections, it is not or radiometrically and geometrically corrected.

(b) Mosaic of an entire survey region off the coast of Sarasota, Florida, comprising multiple, survey lines (transects).

Figure 2.4. Waterfall (a) and mosaic (b) backscatter imagery.

In figure 2.4 (a), a central vertical region of pixels, corresponding to the water column offset, has been removed and the pixel positions corrected for slant range. Processing was carried out with the commercial software package, Coda GeoSurvey¹⁰. The dark central vertical band highlights where backscatter returns from the acoustic beams on either side of the instrument have been joined together, i.e., where the water column has been removed. Different textural regions can be clearly discerned, including the “lumpy” textures characteristic of Sabellaria. The two darker, diagonal lines in the red box are probably trawling marks. The brighter horizontal banding in the blue box is

¹⁰ <http://www.codaoctopus.com/coda-geosurvey/> [accessed 14-02-2012]

most likely due to “cable snatch”, a sudden jerking of the tow cable due to wave induced motion of the surface vessel.

In figure 2.4 (b) the lighter and darker regions in the mosaic represent areas with different backscatter intensities. These intensities are correlated to different grain size (granular particle diameter) distributions¹¹, as in for instance, Collier and Brown (2005) and Goff *et al.* (2000). Rough and smooth textural regions are also evident indicating differing surface rugosities (surface height variability) and morphologies. Small black patches on the image are “holes” where data is missing between transects. Producing a good mosaic is a challenging task and the development of methods to achieve this is an active research area. See, for instance, Chailloux *et al.* (2011) and Coiras *et al.* (2004).

Pixels are geometrically and radiometrically corrected for spatial position and intensity (grey value), respectively in the mosaicing process. Geometric corrections include compensating for water column offset, slant-range to ground-range correction, anamorphic distortion and changes in ships velocity (Chavez *et al.*, 2002). The corrective stages applied to the sidescan data and whether it is mosaiced or not will depend on the purpose for which the imagery is being acquired, the type (manufacturer) of instrument, operational parameters and so on. A non-exhaustive list of factors influencing the image quality is given in table 2.1.

Noise	Acoustic signal	Instrument motion	Miscellaneous
Coherent	Beam radiation pattern	Heave	Line drop-out
Incoherent	Source level	Sway	Pixel drop-out
Speckle (clutter)	Angular responses	Surge	Missing data
Striping (banding)	Sea surface reflections	Pitch	Water column
	Absorption	Roll	
	Spherical spreading	Yaw	
	Refraction		

Table 2.1. A non-exhaustive list of factors influencing sidescan imaging quality.

In neritic environments typical of Sabellaria colonisation, where the water may be very shallow, noise and instrument motion can be particularly problematic. Weather noise

¹¹ Grain size distribution is a mathematical or empirical distribution of the mixture of sediment particle sizes in the analysed sample.

such as wind and rain can severely degrade the imagery and the degree of degradation is extremely difficult to quantify (Kieser *et al.*, 2007). Transmission of mechanical forces to the instrument along the tow cable due to surface waves causes erratic platform motion, compounding the problem, as is clearly illustrated in figure 2.4 (a).

Since sidescan systems have a longer history than MBES, dating back to the early 1960's there is more published literature available. Blondel (2007) covers much of the theory and practical applications of sidescan devices and the processing and interpretation of backscatter imagery. See also; SeaBeam (2000), IHO, chapter 4, (2011) and Hughes-Clarke *et al.* (2009). For instrument specific information, see the GeoAcoustics (Kongsberg)¹², Klein¹³ and Edge-Tech¹⁴ websites. There are dozens of publications providing details of the numerous methods and strategies for processing the sidescan data. A small selection of these includes; Reed and Hussong (1989), Johnson and Helferty (1990), Muller *et al.* (2007), Cobra *et al.* (1992), Cervenka and de Moustier, (1993), Kenny *et al.* (2001), Chavez *et al.* (2002) and Capus *et al.* (2008). Backscatter physics and processing are not of central importance to the thesis, since the research concerns machine assistance of qualitative visual tasks. All backscatter imagery used for experiments in this thesis has been sourced post-processed and no additional corrections are applied.

There are numerous differences between MBES and sidescan instruments, their supporting systems, operational procedures, practical applications and processing pipelines for the production of backscatter imagery. See for instance, Le Bas and Huvenne (2009) who compare sidescan and MBES systems in the context of habitat mapping. A comprehensive coverage of many aspects of MBES systems is given by; Hughes-Clarke *et al.* (2009), SeaBeam (2000), and the International Hydrographic Organisation (IHO), chapter 4 (2011). Instrument specific technical information on Kongsberg Simrad devices (used by Gardline Geosurvey and many other marine survey contractors) can be found at the Kongsberg website¹⁵.

¹² <http://www.km.kongsberg.com/geoacoustics> [accessed 14-02-2012]

¹³ <http://www.l-3klein.com/> [accessed 14-02-2012]

¹⁴ <http://www.l-3klein.com/> [accessed 14-02-2012]

¹⁵ <http://www.km.kongsberg.com> [accessed 14-02-2012]

A major distinction between MBES and sidescan backscatter imagery is the accuracy and precision with which the image acquisition geometry can be determined. Compared to sidescan, the position in space of a recorded backscatter intensity value from the seabed is fixed more accurately and precisely with a MBES system. The keel-mounted transducer location is established using translational offsets relative to a Global Positioning System (GPS) receiver at a fixed reference point on board the survey vessel. Motion sensors record pitch, roll and yaw of the vessel so the attitude of the transducer at signal transmit and receive times is known. Thus, with MBES instruments, measures of backscatter intensity are associated with georeferenced locations and the geometric relationship of the transducer with the seabed is deterministic.

Seafloor slopes and aspects are derived from bathymetric measurements collocated with the backscattering footprint region. Corrections can then be applied to recover an estimate of the signal grazing angle, as in Hou and Hough (2004). This is useful if a backscatter modelling approach is applied to characterise the seabed sediments, such as the composite roughness model, described in Jackson *et al.* (1986). The goal of such geoacoustic modelling is to estimate, through inversion, the geoacoustic parameters of the seabed from an angular dependent backscatter response. Hou and Hough (2004) found that their process of grazing angle estimation and application of corrections to the backscatter signal may provide an improved correlation of the backscatter with the sediment ground truth, particularly grain size distributions. Another remote characterisation method proposed by Fonseca and Mayer (2007) and Fonseca *et al.* (2009) divides the grazing angles into angular ranges from which slope and intercept parameters are extracted. An iterative inversion technique is applied to a modified form the effective density fluid model of backscattering (Williams 2001, Fonseca *et al.* 2002). It is then possible to estimate the acoustic impedance¹⁶ of the seabed, which can be used to predict sediment grain size and other properties. Approaches using backscatter models and angular dependencies are but one paradigm for characterising the seabed. In order to work within a reasonable scope and considering the interpretative nature of the problems addressed in the case studies, such physics based techniques are not considered in the thesis.

¹⁶ The acoustic impedance, Z , of a medium (material) is defined as $Z = \rho V$, where ρ is the density of the medium and V , the velocity of the acoustic wave within the medium.

A considerable amount of research has been carried out on processing and correcting (“calibrating”) MBES backscatter for the many effects imparted on the signal by the environment. See, for example, Hellequin *et al.* (2003), Mitchell, (1996), Augustin and Lurton (2005), Beaudoin *et al.* (2002), and Hammerstadt (2000).

Although some keel-mounted studies are reported in the literature (e.g. Hughes Clarke, 2004), the sidescan is usually towed behind a survey vessel on a cable, as shown in figure 2.5. Acoustic positioning estimates, if used at all, are typically provided by an ultra-short baseline (USBL) beacon system, such as Sonardyne¹⁷ or Applied Acoustics¹⁸. For a discussion of the features of a USBL system, see Philip (2003).

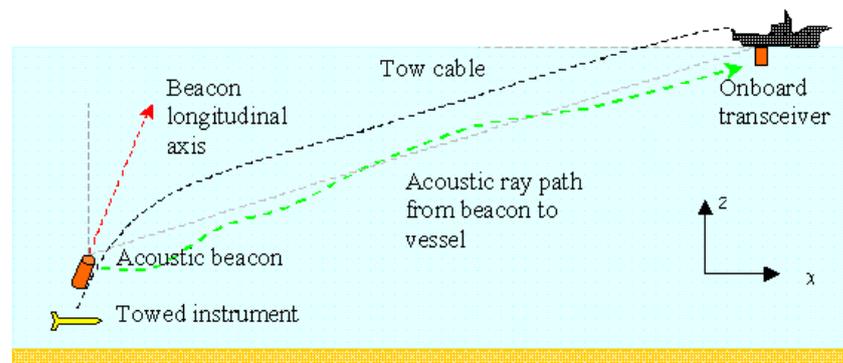


Figure 2.5. A typical towed system arrangement, showing an acoustic beacon attached to a tow cable. The maximum response axis of the beacon is parallel to the cable at the attachment point (Harrison, 2009).

A typical set-up comprises an acoustic beacon (transponder) attached on the cable or tethered to it, near the towed instrument. The beacon, when interrogated, communicates with a transceiver onboard the survey vessel, at a fixed vessel reference position. This type of positioning system can lack accuracy and precision, compared to keel-mounted positioning with GPS, especially with long cable deployments, approaching the beacons operational range limitations. Beacon axis alignment, power, frequency and dispersion pattern as well as prevailing seawater, weather, oceanographic conditions and depth significantly influence the system performance. Fundamentally, sidescan positioning is a hardware issue. USBL positioning can be improved upon by deploying a high-accuracy inertial navigation system (INS) with the instrument platform or by using an acoustic beacon network. Initialisation of an INS system involves position fixing, using

¹⁷ <http://www.sonardyne.com/products/positioning/usbl-all-systems.html> [accessed 14-02-2012]

¹⁸ http://www.appliedacoustics.com/Positioning_Products.aspx [accessed 14-02-2012]

a GPS and an alignment process to determine the initial attitude, as described in Panish and Taylor (2011). INS hardware can be bulky and expensive though, so despite the much better navigational accuracy attained, its usage for simple sidescan surveys is not commonplace. Likewise, acoustic beacon networks require deployment, calibration and recovery. This is a time consuming and expensive component of the survey, so beacon networks are not routinely used for sidescan imaging operations.

Another potential means of estimating the position of a towed sidescan, improving on precision without using expensive additional hardware is by computational modelling. For instance, a hydrodynamic model of the tow cable and towed instrument can be designed and used to estimate the position by numerical integration. The hydrodynamic model can further be combined with a model of the acoustic beacon, as in related but separate work of the thesis author, Harrison (2009). Such a hybrid model can be parameterised adaptively, to fit the cable shape so that the estimated position from numerical integration coincides with known beacon positions. It could help to prevent gross positioning errors and facilitates interpolation or extrapolation in cases where the acoustic positioning from the USBL system is subsequently missing or unreliable. However, whilst this model can potentially provide more precise estimates of spatial positioning under relatively steady conditions, the attitude of the instrument is indeterminate.

Unknown attitude of the sidescan transducer is again, fundamentally, a hardware issue and can be resolved if pitch, roll and yaw motion sensors are attached to the instrument platform. As with INS positioning though, motion sensors are typically not fitted to the platform due to costs. Usually, the only directional information recorded is from a simple heading sensor. Hence, due to the uncertain position and unknown geometry of the transducer in relationship to the bathymetry, the backscatter signals from a simple sidescan survey cannot be geometrically and radiometrically compensated with the same accuracy and precision as the MBES. Capus *et al.* (2008) also claim sidescan platform stability is an ongoing issue and monitoring sensor roll, pitch and yaw remains a topic for investigation. They point out that these challenges could be met with accurate and synchronised attitude data. These data are rarely available though. It is still possible to use backscatter recorded by a sidescan to remotely characterise the seabed sediments under certain, idealised conditions. For well-sorted (relatively uniformly

sized) particles, both Collier and Brown (2005) and Goff *et al.* (2000) found a strong linear correlation between mean intensity, over a backscatter image neighbourhood and mean sediment grain size, as sampled from the seabed in the corresponding image region.

Despite the difficulties in positioning and attitude determination, sidescan is capable of acquiring very high-resolution (about 10-20 cm ground resolution) backscatter data showing intricate details of textured structures at smaller scales. Since the sidescan is usually towed a few metres above the seafloor, it can operate at high (~500 kHz) frequencies in deeper water. This would not be possible for a keel mounted MBES at the sea surface, due to rapid attenuation of the high frequency signal with depth.

Sidescan has proven to be particularly useful for qualitative imaging of biogenic reef structures (Limpenny *et al.*, 2010, Birchall, 2007, Degraer *et al.* 2008). Imaging of Sabellaria colonies using different acquisition parameters was investigated by Limpenny *et al.*, 2010. In a visual inspection of the imagery they found the quality of the textures was influenced greatly by towing speed and the towing direction. Higher towing speeds reduced the detectability of Sabellaria reefs in the backscatter image, as did imaging in a direction perpendicular to the long axes of the small-scale reef features. Whilst beam range alone did not seem to greatly affect the imaging results, the ratio of instrument height above the seabed to the range of the acoustic beam is an important consideration. They found no discernible differences in the imaging of the reef textures at low (100 kHz) and high (400 kHz) frequencies. Interestingly, Degraer *et al.* (2008) imaged another type of biogenic reef formed by the tube-building worm, *Lanice conchilega*. Imaging results were strongly frequency dependent, when they compared reef textures at 132 kHz and 445 kHz. The higher frequency imagery resolved finer details much more clearly, compared to the lower. However, in comparison to the study of Limpenny *et al.* (2010) a different sidescan instrument was used and the towing altitude was at least 2 m lower (3-4 m above the seabed). Collier and Brown (2005) also established that a higher frequency (410 kHz) produced qualitatively better images of an artificial reef, made from concrete blocks. Further, Birchall (2007) found that higher (500 kHz) frequencies were more useful for imaging details of Sabellaria reefs and suggests that further investigations are required in order to establish the best frequency to use.

Under controlled laboratory conditions, backscatter responses from smooth homogeneous sediments exhibit frequency and angular dependent characteristics that can be modelled empirically or mathematically. In fact, there are many such models, some of which, as mentioned earlier, can be inverted to estimate seabed sediment properties in MBES remote characterisation applications. However, there are no proven, reliable models for higher frequencies such as 500 kHz, used in sidescan imaging. There is very little evidence in the literature to suggest that any frequency or angular dependent relationship exists for imaging more rugged, heterogeneous structures, e.g Jackson and Richardson (2007, p.376). Beyond a certain grain size (larger than the wavelength of the resonating acoustic signal) factors other than grain size dominate the scattering process, such as surface roughness and orientation of the imaged object. According to Chotiros (2006) in this scattering regime, “the concept of grain size dependence breaks down.” He goes on to point out that large coral reefs have facets with preferential orientations and this gives rise to wide variability in backscatter intensity with aspect. Due to the heterogeneity of the real seabed, the backscatter models can fail and produce unexpected or unreliable results. A surface mixture of larger sized particle grains or shell fragments with an otherwise well-sorted sediment can produce a poor, highly variable correlation of backscatter with the dominant grain size of the substrate, as reported in Goff *et al.* (2000) and Bellec *et al.* (2010).

At higher acoustic frequencies, in general, the scattering is mostly due to surface roughness (rather than volume heterogeneities), due to the rapid attenuation of the signal as it penetrates the sub-surface. To complicate matters further, there are instances where the surface and volume components cannot be separated. Worm tubes are just such a case (Jackson and Richardson, 2007, p 406). A Sabellaria reef is a rugged structure composed of a heterogeneous matrix of worm tubes, sediments and other organisms. Even if it were imaged by a sidescan, under ideal conditions, with deterministic, precise imaging geometry, its backscatter response cannot currently be modelled. Therefore, there is no reliable, objective means of remote, global characterisation of the Sabellaria through a process of geoacoustic inversion. Despite the ambiguity, human identification of the textural surrogates for Sabellaria colonies in sidescan imagery, verified by ground truth remains the predominant method in the process of identifying and mapping the species. It is a subjective, laborious and

expensive task with potentially unreliable results. One of the aims of the thesis is to investigate the tractability of automating this manual Sabellaria texture discrimination task, with feature based machine learning. In addition to the novelty of the texture identification and discrimination task, there would be clear commercial gains in expediting this process in an automated or semi automated processing pipeline.

2.3.3 Bathymetry

Bathymetry data are nowadays routinely captured with a MBES. The MBES can record hundreds of depth soundings simultaneously enabling rapid, accurate coverage of large areas. By processing the raw soundings together with position, motion, tidal and water column data, an accurate DBM is produced which represents sampled depths of the seafloor terrain. This can also be thought of as another type of seabed image.

A small region of a bathymetry raster from the Geological Survey of Norway (NGU) is rendered in figure 2.6. The crater-shaped depressions are pockmarks in soft sediments, approximately 30 m across and 2 m deep. The horizontal resolution of this DBM is 5 m between grid nodes. Vertical resolution is 1 cm.

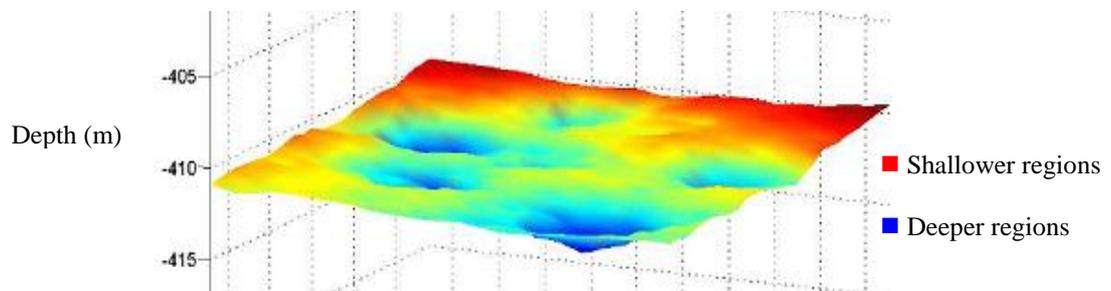


Figure 2.6. Rendered section of a bathymetry raster, supplied by the NGU, showing pockmarks.

In addition to rendering geomorphic objects over a wide range of spatial scales, such as pockmarks, dunes and canyons, the DBM surface has a texture. Rocky or reef-like surface regions are relatively rough, compared to say, a sedimentary plain. Often, objects or regimes that are visible in the bathymetry have a corresponding signature in the backscatter imagery, for instance, a prominent and rugged reef-structure. Targets can be recognised and delineated in the bathymetry by considering their size, morphology and texture.

As discussed further in chapter 3, there have been relatively few detailed investigations into the machine discrimination of specific submarine landform objects. Due to the importance of landform objects such as pockmarks, as pointed out earlier, it is usually necessary to identify and map these objects in the DBM. The discrimination and mapping of these objects in a machine learning process is thus considered in another application case study in the thesis. In common with the Sabellaria discrimination, there is novelty in and commercial motivation for investigating the potential for automation of this process and devising methods to accomplish the task.

Many aspects of MBES depth determination and data processing can be found in Hughes-Clarke *et al.* (2009), SeaBeam (2000), and IHO, chapter 3 (2011).

2.3.4 Ground truth

Acoustic imagery contains information about the expected real morphology and properties of the seabed. Even to the non-expert, it should be clear that the distinctive textures in figures 2.3(a) and (b) represent different natural targets. However, the imagery is merely a surrogate - a distorted, acoustic representation of the seabed. Sabellaria colonies can present a distinctive texture but there is a possibility something different is actually on the seabed, say, a shellfish bed or a mixture of shellfish and clumps of Sabellaria interspersed with a sandy substrate. The colony may be in decline or growth but this and many other properties of the target cannot be established from imagery alone.

The speculative interpretation of the imagery requires validation and this is partially fulfilled by using ground truth – collecting evidence of what is actually at the seabed. It typically involves sparse grab sampling at a few point locations to collect unconsolidated sediment samples. The sampling pattern may be regularly spaced, random or directed toward specific targets already provisionally identified in the imagery as in Galloway (2008). Subject to prevailing water turbidity conditions, video footage is also routinely captured to non-invasively provide information about the biota and consolidated structures such as rocky or biogenic reefs where physical sampling may not be practical (or legal.) Photographs and video footage can be taken of different

areas of the seabed, before the camera needs to be brought up to the surface, thus, economically covering a wide area. Collection of video data is described in, for instance, Ierodiaconou *et al.* (2007) and recent advances in underwater imaging, in Kocak *et al.* (2008).

In contrast to video, physical samples are invasive and a relatively slow, expensive undertaking as the survey vessel has to stop, lower and retrieve the sampling apparatus at a number of locations. The position at which a grab sample is taken is uncertain and may deviate greatly from the position of the vessel, depending on the water depth and local currents. Collier and Brown (2005) report an estimated positional accuracy of ± 15 m in 30-52 m water depth (approx. 35-50% of depth), based on the wire length, entry angle and sea conditions. However, wire angle at entry to the water is not always a reliable indicator of the likely cable trajectory, since the speed and direction of currents varies with depth. Acoustic positioning systems such as USBL can be deployed with the grab, to improve the estimated location of the grab on the seabed, relative to the vessel. In general though, an accurate position of the point on the ground from which the sample was taken is not known. This can create problems in supervised classification processes when the ground truth locations are used to collect seed samples from the mosaic imagery, for training a classifier. This matter is considered again later in chapters 4 and 8.

A multitude of devices are in routine use for seabed sampling and many of these are described in McIntyre (2005). For sampling Sabellaria, Limpenney *et al.* (2010) recommend four different types of mechanical grab; Hamon, Day, Shipek and Van Veen as well as a small epibenthic trawling method.

Quantitative information, such as a sediment grain size distribution can be obtained from Particle Size Analysis (PSA) of a grab sample. The distribution can be parameterised by the geometric mean, standard deviation, skewness and kurtosis, as in Folk and Ward (1957). Semantic labels are applied using descriptive terminology for grain size scales, e.g., Friedman and Sanders (1978) and for mixtures of sediment, as in the British Geological Survey (BGS) modified Folk sediment trigon, illustrated in figure

2.7¹⁹. The trigon indicates 11 sediment classes, defined according the region occupied by the sediment mixture. In producing a class map for offshore industrial use though, a simplified representation of a few classes is often preferred. The UK SeaMap classification scheme (Long, 2006) is an example, with four sediment classes; mixed sediment, coarse sediment, mud and sandy mud and sand and muddy sand, as shown by the four coloured borders in figure 2.7.

Legend

G	Gravel
mG	Muddy gravel
msG	Muddy sandy gravel
sG	Sandy gravel
gM	Gravelly mud
gmS	Gravelly muddy sand
gS	Gravelly sand
M	Mud
sM	Sandy mud
mS	Muddy sand
S	Sand

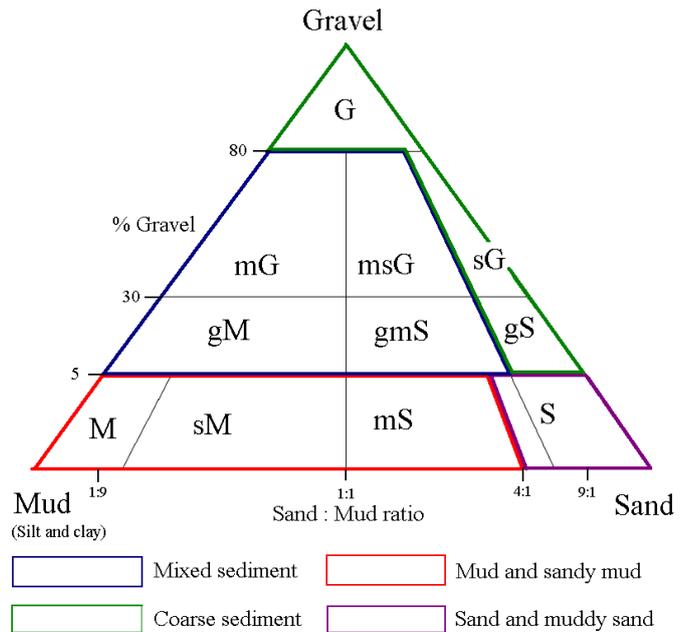


Figure 2.7. BGS modified Folk sediment trigon showing the UK SeaMap classes (not to scale.)

Properties such as sediment permeability and hydraulic conductivity can also be derived from the mean grain size.

2.4 Other pertinent issues

Processing, interpreting and integrating the various sources of data and information to produce a class map are reliant on a host of technology, scientific principles and human factors. As the discipline of automated seabed classification is emergent, there are few standards and benchmarks with which to evaluate and compare the work of others quantitatively. Perhaps for this reason, the objective statistical methods for acoustic characterisation of the seabed is favoured by many researchers (Anderson *et al.*, 2008).

¹⁹ Figure is not to scale and is based on fig.3, page 5 in Long (2006).

When the substrate characterisation derived from the backscatter is considered together with topographic descriptors of the bathymetry, such as slope and aspect, it becomes possible to remotely map a physical habitat, in terms of these surrogates, using the instrument platform and a dedicated software processing pipeline.

However, seabed substrates and topographic surrogates alone do not capture the richness and diversity of the seabed environment. Remote characterisation is one representation of this environment at a particular scale of mapping using a specific type of instrument. Sidescan sonar affords another view, usually on a smaller scale and at a higher resolution than the MBES. It further captures details of important seabed targets such as Sabellaria colonies, a task which is unproven with current MBES technology. Textures in qualitative sidescan imagery are usually interpreted by humans. There are drawbacks to a human approach though, as Blondel (2007, p 249) points out, human interpreters with different experiences, will produce different interpretations. Also, validation of tentative interpretations/classifications with ground truth is always uncertain. It is not practical (or possible even) to ground truth every single point on the seabed - this is the whole idea behind extrapolating the classification of acoustic facies over a large homogeneous region based on sparse ground truth associated with the properties (texture, intensity) of the corresponding sonar image region. A machine process can expedite the manual tasks and improve their consistency. Yet the problem of validating the results and evaluating the efficacy of the processes in relation to the work of others still remains a formidable challenge, irrespective of whether the task is performed by a machine or human. This important issue is given further consideration in chapter 4.

Binary class maps showing the distribution of a target class can be particularly useful such as “Sabellaria” and “non-Sabellaria”. The number of classes that can be usefully represented on a single map is open to debate and will depend on the classification objective and client specifications. Also, meaningful, standardised unambiguous and consistent labeling of classes is very difficult to achieve. As Fraschetti *et al.* (2008) point out, there is “a lack of common vocabulary on habitat types.” This creates a further complication in trying to evaluate seabed classification results in relationship to the work of other researchers (not to mention their use of differing instruments, environmental conditions, operational procedures, data processing pipelines and so on.)

There are many different classification schemes for habitats that can help with the assignment of standardised class labels, thus facilitating comparisons between different surveys and habitat maps, when identical schemes are used. A physical habitat approach, applied in Canada was defined by Roff *et al.* (2003) and in Europe, the European Nature Information System (EUNIS)²⁰ classification scheme is sometimes applied, as in the MESH project (MESH, 2009). However, Harris and Baker (2012, p 877) in a survey of recent habitat mapping research published in the Geohab Atlas, found that most researchers did not use any published classification scheme.

2.5 Summary

This chapter has served to elucidate some of the techniques and problem areas, in the wider seabed classification domain. The research has been positioned within a technical, environmental and commercial context. Two important tasks for case study investigations have been identified (1) pockmark discrimination in multi beam bathymetry and (2) Sabellaria texture discrimination in sidescan sonar imagery. The need for the research and commercial motivation has been established. The main points arising from the chapter are:

- Class maps of natural targets in seabed imagery are vitally important for the scientific study and management of marine habitats and ecosystems. They are business critical for companies engaged in resources exploration and production, site placement and infrastructures. Production of seabed class maps from marine surveys provides a strong revenue stream for organisations involved in the surveying and mapping process.
- Under certain conditions and assumptions, some types of seabed sediments can be remotely characterised with MBES using backscatter models. However, important natural targets at smaller scales such as Sabellaria colonies and pockmarks cannot presently be directly identified and characterised in this way.

²⁰ <http://eunis.eea.europa.eu/> [accessed 05-12-2012]

- Identifying, classifying and mapping such subjective natural targets involves manual human interpretation and processing (an interpretative approach). This is inconsistent, difficult, time consuming and therefore expensive. Results may be unreliable and the certainty of the classification and the class boundaries is usually unknown or not quantified. Noisy and ambiguous labeling is another issue compounding the reliability problem.

- The purpose of this thesis is to present an investigation into some of the issues and applications identified and design bespoke computational methods and processes for effectively representing, identifying and delineating some of these seabed targets, with a goal of assisting the human interpretation and classification process.

The next chapter advances the position incrementally by considering related research with an emphasis on the computational methods and processes, particularly the feature based machine approaches used in various seabed classification tasks.

Related research

Chapter 3

Contents

3.1 Introduction

3.2 Backscatter image classification methods

3.2.1 Feature creation

3.2.2 Sidescan sonar backscatter image formation considerations

3.2.3 Classification and clustering methods

3.3 Bathymetry features

3.4 Combining bathymetry and backscatter information

3.5 Ensemble methods

3.5.1 Supervised ensemble

3.5.2 Unsupervised (cluster) ensemble

3.6 Outline of machine learning methods used in the experimental work in this thesis

3.6.1 Computational kernels

3.6.2 Feature evaluation and selection

3.6.3 Supervised learning

3.6.4 Unsupervised learning

3.7 Summary

3.1 Introduction

In the previous chapter, the application context of the work was established and two niche areas for further investigation were identified; pockmark discrimination and Sabellaria discrimination. In this chapter, the emphasis shifts towards machine methods and processes that could be used to help solve these problems and potentially, other related problems. The aim of this chapter is to narrow the scope of the computational methods to be considered in the thesis, by identifying underexplored areas for further investigation within the application confines.

To recapitulate: research in the thesis focuses on feature based machine learning methods and processes that could be useful for assisting humans in performing qualitative manual, target discrimination tasks. The means by which the targets are imaged is not a critical factor in this work. Since, if there were a universal standard, whereby the target could be discriminated and characterised automatically and reliably at a hardware level, by the acoustic remote sensing platform (and an associated standardised processing pipeline), it would not be necessary to carry out a tedious and expensive manual discrimination and classification process. The problem is exemplified in the case of legacy data, where nothing is known about how the data were captured, yet it is still clear to a human interpreter that the data contain valuable visual information. It is an important distinction to make, between the computing science and engineering research in this thesis and the science of acoustic seabed classification.

Several researchers have recently remarked on the need for advancement in the area of automated seabed classification tasks. Parnum (2008) states that improvements could be made by applying more sophisticated classification algorithms. Marsh and Brown (2009) claim there is a need to develop automated computational methods to transform high dimensional bathymetric and backscatter data into easily visualised, low-dimensional maps, representing the seafloor. According to Simard and Stepnowski (2007), emphasis should be placed on the extraction and selection of appropriate features, for the seabed classification to be successful. Blondel (2007, p. 275) claims there is rapid growth in the disciplines of computer assisted interpretation and these will be developed in industrial and academic applications over the next decade.

Some of the most relevant related research on feature-based machine learning methods as applied to seabed classification problems on rasterised backscatter and bathymetry data is introduced in this chapter. Dozens of books and thousands of research papers have been written on methods in the associated domains of data mining, machine learning and image processing. Some specific methods are considered in greater detail and additional works discussed within the case study chapters later on, as the need to further contextualise the authors' work arises.

The remainder of this chapter is organised as follows. Section 3.2 considers methods used on sidescan and MBES backscatter imagery. Approaches to MBES bathymetry are

described in section 3.3 and examples of combining bathymetry and backscatter considered in section 3.4. The paradigm of ensemble classification is outlined in section 3.5, covering supervised and unsupervised (cluster ensemble) approaches. An outline of some of the fundamental machine learning and computational methods applied in the experimental work in this thesis is provided in section 3.6. Finally, the chapter is summarised in section 3.7.

3.2 Backscatter image classification methods

The basic backscatter image classification problem involves dividing a sonar image into a number of relatively homogeneous regions with similar properties, then assigning class labels to those regions. Generically, in a feature based machine approach to the problem, numerical feature creation kernels need to be parameterised, evaluated and applied to the imagery. The features must be capable of capturing salient information, representative of the target class properties and it is usually necessary to measure this in some way on the data specific classification objective. If the features are ineffective or inappropriately parameterised, the classification accuracy may be compromised. When the feature vector is large, dimensionality is sometimes reduced by feature extraction or selection. Patterns generated from the data in the feature creation process are then used to partition the regions in a machine learning process. The process may be supervised, if some data where target class labels are known in advance are available for training a classifier or unsupervised when no *a priori* assumptions can be made about the classes present. The following two subsections consider some of the feature creation and machine learning methods that have been used in the domain.

3.2.1 Feature creation

Simple, effective and efficient features for discriminating regions are statistics based on backscatter intensity. A basic assumption is that lighter and darker regions of the imagery correspond to areas with different backscatter intensities and thus different material properties or sediment types. In a machine learning approach, the mean backscatter intensity calculated on an image neighbourhood or “tile” under a computational kernel can be used as a discriminatory feature. Although the relationship between say sediment grain size and backscatter intensity is complex, using mean

intensity as a feature can be useful for mapping the distribution of estimated mean grain sizes of sediments in sidescan imagery. Combinations of statistical features, such as mean, median and standard deviation may be used together for a meaningful partitioning of backscatter imagery in relationship to the substrate and for habitat prediction, as in Brown and Collier (2008).

For identifying and segmenting particular morphologies, texture is more important than local mean intensities, i.e., some image regions may have similar mean intensity values but disparate textures. Texture is a manifestation of the pattern of spatial arrangements of the intensities of individual image pixels. Using mean intensity alone, would not reliably detect the textural differences characteristic of say the morphology of smooth sand and rippled sand. According to Blondel and Gómez Sichi (2009), countless studies have shown that textures capture most of the useful information contained in sonar imagery. However, the features used should be dictated by the purpose of the classification task. For instance, if it is desired to partition the imagery into regions of “high”, “intermediate” and “low” backscatter levels, and relate these to mean grain size distributions, then descriptive statistics would suffice, it would not be necessary to use textural features.

Approaches to texture segmentation fall into two broad categories,

- (1) **Region based methods**, identify all pixels as members of different homogeneous regions or textural classes in the image. Boundaries are established implicitly as a by-product of identifying the spatial extent of the regions.
- (2) **Boundary based methods**, directly identify the boundaries between different textural types according to prescribed criteria.

Boundary based methods depend on successful boundary detection, to segment the regions of interest. The classification is then achieved by assigning class labels to the segmented regions. Boundary methods are suitable for tasks where the boundaries are reasonably distinct, for instance, between rock outcrops and sediment. A region-based approach is used throughout the thesis as boundaries are generally not well-defined. In

the region based approach used, the segmentation can be thought of as an outcome or by-product of the tentative, supervised classification of small image tiles or neighbourhoods.

Feature creation methods used to represent the textures can be divided into four main categories, see for instance, Sonka *et al.* (2008), Gonzalez and woods (2009);

- (1) **Statistical methods**; the texture is described by the statistics of a local neighbourhood of pixels, e.g. co-occurrence matrices, autocorrelation, edge-frequency, run-length and surface roughness statistics.
- (2) **Syntactic (structural) methods**; the texture is modelled as a composition of small base elements (texture primitives) forming a defined spatial pattern; e.g. shape-chain grammars, graph grammars.
- (3) **Model based methods**; the texture is represented by a probabilistic or physics based model, representing the underlying stochastic process generating the image texture, e.g. Markov Random Field, fractal dimension, autoregressive-models.
- (4) **Signal processing methods**; texture features are derived from the response of a filter kernel applied in the frequency or spatial domain, e.g., localised (windowed or short-time) multi-resolution spatial filters such as Gabor filter banks and wavelets.

For a comprehensive review of texture analysis methods, see Xie and Mirmehdi (2008).

Of these methods, statistical and signal processing are probably the most widely used in the sonar image classification domain. The reasons for this are not entirely clear. Methods such as the Grey Level Co-occurrence Matrix (GLCM), described in further detail below have been in existence for 40 years and are still in widespread use for generic classification tasks in commercial sonar image processing software packages. So, it is perhaps this “pedigree” and their wide applicability that has led the community to maintain the *Status Quo* with such feature creation techniques. However, for specific,

subjective target discrimination problems, towards which, this thesis is biased (as opposed to more generic classification) there are many other methods available that have not been widely used or explored in the sonar domain.

Grey Level Co-occurrence Matrix Features (GLCM)

Penrose *et al.* (2005) and Parnum *et al.* (2004) claim the most common type of statistical analysis used for sonar image textures is the Grey Level Co-occurrence Matrix (GLCM). Devised by Haralick *et al.* (1973), the GLCM captures the distribution of pixel intensities and their spatial relationship. Blondel and Gómez Sichi (2009) claim that other research has shown GLCM's are optimally adapted for the purpose of describing texture. Haralick *et al.* (1973) initially defined 14 textural features that can be computed from the co-occurrence matrix. Since the method was first introduced, several other features have been devised and used for specific purposes, such as cloud classification in Welch *et al.* (1988). However, it is well-known that many of the 25-30 GLCM features presently in use are intrinsically correlated due to their functional relationships. According to Reed and Hussong (1989) the five features; angular second moment (ASM), contrast (CON), correlation (COR), entropy (ENT) and angular inverse difference moment (AIDM) have low correlation, as determined by principal components analysis (PCA). Reed (1987) also points out, that ASM, CON, ENT and AIDM are insensitive to gain settings and look direction of the sonar. The correlation feature is independent of the other GLCM features and produces results very similar to autocorrelation and semi-variogram methods (Van der Sanden and Hoekman, 2005).

If more than a few different features are used, not only is computational complexity increased but it is likely that redundant information will be generated which could be detrimental to the classification accuracy. Blondel and Gómez Sichi (2009), Blondel *et al.* (1993) and Blondel (1996) found the two features, ENT and homogeneity (HOM) were able to capture most of the textural variability in sidescan images and they use these features in "TexAn" a proprietary software package developed at the University of Bath. Blondel (1996), points out that the definition of HOM is the same as inverse difference moment (IDM) in Welch *et al.* (1988). This is consistent with the original definition of IDM by Haralick *et al.* (1973). Reed and Hussong (1989) used the same definition of AIDM from Haralick *et al.* (1973) for IDM. In other words, AIDM, HOM

and IDM can be regarded as the same feature. Kalcic and Bibee (2004) used ENT, HOM, CON and COR features, averaged over four different orientations. Averaging over different directions has a two-fold advantage of (1) reducing dimensionality and (2) producing rotationally invariant features. Reed and Hussong (1989) and several other researchers have used these rotationally invariant features. A reason for this, as Kalcic and Bibee (2004) mention, is that using rotationally invariant features helps suppress variability in the patterns due to platform motion artefacts imparted in the image data. The GLCM has several parameters that can be tuned for the image data and classification objective.

Although GLCM derived features are ubiquitous in the domain and are used in leading commercial software packages, there are many other types of features and feature combinations that can be applied. Karoui *et al.* (2008) point out that a number of studies have indicated a fusion of features from the different methods could improve texture characterisation. They investigated a fusion of filter bank responses and co-occurrence distributions to characterise the texture of sidescan images. However, a potential drawback of their approach is the use of predefined filter bank parameters without explicitly evaluating the affect of different parameterisations. Many texture feature generation methods are dependent on multiple parameters. Evaluation and selection of feature parameters for the data dependent classification task is critical to its success. As Ruiz *et al.* (2004) point out, there are several parameters to optimise within a chosen method or when using a combination of methods to generate features. They considered four methods; GLCM, energy filters and edgeness, Gabor filters and a multi-resolution (discrete wavelet transform) analysis. Their results showed that a combination of different texture methods could improve classification accuracy. However, they concluded that a lack of widely accepted benchmark data means all results must be considered within the reported set up. Martin (2005) investigated GLCM, run-length matrix, wavelet transform and Gabor filters and found that a fusion method was more robust to badly selected features but in general, features from the run-length matrix had poor performance. Müller *et al.* (2007) compared four different approaches to sonar texture classification; The GeoTexture¹ commercial software package, a wavelet-domain approach, Artificial Neural Networks (ANN) together with grey level run length features and GLCM and ANN with the Grey Level Co-occurrence Iteration Algorithm

¹ <http://www.km.kongsberg.com/> [accessed 22-03-2012]

(GLCIA). They found that the wavelet approach performed well in identifying gravel but comparatively poorly with sand and mud. They used the supervised wavelet method of Fan and Xia (2003) and suggested it was not suitable for efficient sonar texture classification purposes at this stage. The GLCIA and ANN gave the best all-round results, but involved a subjective, manual feature extraction process. A different choice of GLCIA feature vectors can result in different ANN's based on the same training data set, leading to inconsistent results.

In one of the leading commercial software packages for seabed classification, QTC Swathview², Preston (2009) applies a variety of kernels to generate 29 features, including GLCM from the sonar imagery. Principal Components Analysis (PCA) is used to convert the high-dimensional feature space into an orthogonal, three-dimensional component space by forming weighted linear combinations of features. PCA was also used by Kalcic and Bibee (2004). However, they claim a more optimal approach would be to choose the individual features with the greatest discriminatory potential between the regions of interest, if sufficient ground truth is available to reliably evaluate class separability. Further, when using PCA, in the case of swathview, directional components are preserved and used in the recombination process, since the GLCM features are not rotationally invariant. It is therefore possible that local textural anisotropies and directionally dependent artefacts could adversely influence the feature components.

Using separate feature creation and selection (or extraction) stages followed by classification is the most common approach in the domain. It is though not necessary to have separate stages. By using a Kohonen Self-Organising Map (SOM) (Kohonen, 1990), Zhou and Chen (2005) were able to circumvent feature selection in their MBES backscatter image classification procedure. Pican *et al.* (1998) also used a SOM approach and compared co-occurrence matrix features with the SOM, for texture segmentation of colour seabed imagery. They found that the GLCM features appeared to capture better textural information than the SOM, although the optimum set of GLCM features has to be determined.

² Quester Tangent Corporation, <http://www.questertangent.com> [accessed 30-11- 2011]

Model based and other approaches

Mignotte and Collet (2000) published one of the few model-based approaches to sonar image classification. They claim their approach, using a Markovian segmentation model and fuzzy logic modelling, is robust to the instrument type and conditions of acquisition, as the method works on a shadow detection map as opposed to the grey levels of the input image in a texture based approach.

Liu and Zhang (2007) used a Pulse coupled Neural Network (PCNN) which imitates an animal retina. The PCNN produces a series of binary images from which a feature vector is formed. They claim subsequent classification accuracy better than spectral methods and GLCM on test cases of 12 sonar images.

Signal processing approaches

Studies using signal processing based methods for sonar texture analysis are not as widely reported as the statistical methods. Although relatively uncommon in sonar image processing, the popularity of Gabor Filter banks in many other domains is possibly due to certain similarities between the mechanism of textural decomposition of the filters and the inferred texture decomposition mechanisms of the Human Visual System (HVS.) In an empirical study involving twenty human subjects, Rao and Lohse (1993), identified three key dimensions of textural perception; repetition, direction and complexity. Multi-channel filters, such as Gabor filter banks can generate textural features at different orientations and frequencies replicating, in a very simple manner, the orientation (direction) and spatial frequency (repetition) mechanisms of the HVS. Atallah (2004) found, in a comparison between Log-Gabor filters and Discrete Wavelet Transform (DWT) features on bathymetric sidescan imagery, that generally, the best classification results from both methods were close. However, the performance of any method is dependent on the data, classification task and parameterisation. In a different case study using sonar imagery of an artificial coral reef, Atallah (2004) found that increasing the number of orientations and using non-dyadic frequency scales with the Gabor filters did not greatly influence the classification results. Two other recent studies where Gabor filter banks were applied to sonar imagery are those by Samiee and Rad

(2008) and Sun and Shim (2008). The feature creation process of Samiee and Rad (2008) used the filter banks to generate a sub-image from each channel followed by morphological closure of the sub-images. A modified version of the Chan and Vese (2001) Active Contour Model is then applied to segment the binary sub-images. Whilst their method is robust to noise and the data acquisition processes, their results were reported on small, clean image regions with distinct textural contrasts and relatively well-defined textural boundaries. Sun and Shim (2008) combined Gabor filter bank features with a model-based feature, the fuzzy fractal dimension (FFD) in a hybrid fusion method. Using a Multi-Layer Perceptron (MLP) classifier, they found overall, the classification accuracy was higher with the fusion approach, compared to using the Gabor filter bank or FFD features in isolation.

3.2.2 Sidescan sonar backscatter image formation considerations

Perhaps the two most widely developed paradigms for generic seabed backscatter classification are; (1) image based, using texture as a discriminatory subspace and (2) model based approaches utilising calibrated angular responses of backscatter signals measured in-situ, together with parametric models of the backscatter interaction with different (known) sediment types.

As has already been discussed in the previous chapter, there are some severe limitations and technical challenges relating to the calibration of backscatter signals captured by a towed sidescan sonar. From a practical perspective, the availability of reliable ancillary measurement data (or lack of it) that could facilitate signal corrections is also a concern. The key problems of positioning and imaging geometry are surmounted by MBES sensing platforms and systems, which are now routinely used for collecting calibrated backscatter. Subsequently, automated, software-centred, model-based, backscatter angular dependency classifications of the seabed sediment types can be carried out on these calibrated data. Some of the research in this thesis though, focuses on subjective targets in qualitative (uncalibrated) sidescan backscatter imagery, so by default, an angular response, model-based approach to classification is ruled out. This should also be apparent when there are no backscatter models of specific targets such as Sabellaria colonies.

An important purpose of a commercial sidescan sonar survey is to image the target regions of interest, within the operating parameters and budget. Hence the survey needs to be carefully designed, taking into consideration factors such as imaging direction, towing height and speed, acoustic frequency, pulse duration and so on, appropriate for capturing the target signatures, as in Limpenney *et al.* (2010.) Expensive, usually sparse ground truth, provides a validation of the presence of the target regions of interest that are subsequently related, through a human visual inspection to the sidescan sonar image textures. It is well known though, that due to the plethora of factors affecting the signal, the rendered textural information in an image presented to the human analyst is not necessarily a faithful representation of the real targets on the seabed. Target regions will be missed altogether if their signatures are not imaged and rendered recognisably. It should be clear too, that if the texture is severely corrupted or not visible to a human, the discrimination is not necessarily improved upon by applying a machine approach to the discrimination problem. Discrimination by the machine and human (ignoring internal processing factors) is always constrained by the amount and quality of information they are presented with.

Areas of the target regions in some of the data used in the thesis were imaged from different (approximately parallel) directions but there were no raw data or ancillary data available to potentially make use of this information (only the processed image raster data were available to work with.) Evidence from laboratory experiments indicates that directional illumination acts as a filter of three-dimensional textures in optical (Chantler *et al.*, 1994, Chantler and McGunnigle, 1995) and sidescan sonar (Bell *et al.*, 1999) image formation processes. The direction from which a target is sonified produces a different appearance to the image texture. Imaging from more than the two approximately parallel directions (e.g., from orthogonal directions) improves the likelihood of capturing the signatures of the targets. It also improves the reliability of the classification *if* a similar texture is identified in the same location in two independent data sets. In a binary discrimination problem of sonar imagery, if “correctly” imaged textures i.e., those rendered as being recognisable as a particular class to a human expert and also corresponding to the actual target regions on the seabed are correctly identified, then the region of interest is a true positive. Changing the direction of sonification could in some cases then potentially result in a different classification outcome if the target has strong anisotropic properties, such as sand

waves. Imaging from multiple directions is rarely done in commercial surveys though, due to cost considerations coupled with post processing complexities.

As Bell *et al.* (1999) point out, the majority of research on texture based classification methods ignores the physical processes by which the sonar image is formed. It is also the case in this thesis. A scientific treatment and modelling of the image formation process, say for Sabellaria textures would also require a dedicated, highly controlled campaign to capture and process the data needed for model verification and validation of the synthesised images. Such data were not available for this project. However, even if the data were available, there are still many further uncertainties involved, concerning the ground truth definition, sample location accuracy, sampling density and its relationship to what is discriminated in the backscatter image. So, any image formation process model devised to take factors such as directional illumination into consideration (and possibly even compensate for its affects) will contain limitations and uncertainties. Further, its validation would ultimately be based on data that are statistical and uncertain in nature.

The work on Sabellaria texture discrimination in this thesis considers high resolution imagery (up to 6 pixels/m) and the classifications are carried out on small patches, generally corresponding to ground areas of less than about 25 m². Even over these relatively small ground areas, the physical target textures can be heterogeneous and mixed with other texture classes with uncertain boundaries. This represents natural variability, especially in the isotropy, spatially heterogeneous size, shape, surface morphology and distribution of the imaged biogenic structures of interest. The natural variability is apparent over a variety of resolutions and is different to say the more homogeneous and isotropic texture of a smooth sandy surface comprising small, well-sorted (uniformly sized) particles to which some image formation modelling approaches have been successfully applied.

If there were a representative image formation model for the Sabellaria then it may be possible to synthesise the appearance of Sabellaria textures in sidescan sonar imagery under many different conditions. Any sidescan sonar image data captured in the real-world surveys could be compared texturally on a neighbourhood basis with the texture model generated, under the conditions specific to the survey. Similarity of the real-

world image texture to the model could then be used as an indicator of the likelihood that the imaged region is a Sabellaria colony. The complexities of devising such a model should not be underestimated though. Although it has not been done specifically for Sabellaria, Bell (1995) and Bell and Linnett (1997) modelled generic underlying physical image formation processes by integrating separate models for towfish motion, transducer directivity, ray path propagation and scattering from a fractal model of the seabed surface. The underwater scene can include other object models, above or within the surface, using primitives such as spheres or cylinders. Scattering of the acoustic ray at the intersection point with the scene is simulated using the bistatic model of Jackson (1994.)

As was pointed out previously though and also in Bell and Linnet (1997) problems arise in validating model simulations due to the lack of definitive data sets for characterising the experimental environment. Bell and Linnet (1997) also considered the problem of how to compare the simulated output from the model to the actual sidescan imagery and used mixed-methods, combining visual interpretation and statistical techniques to achieve this. Although identical visual matches between synthetic and real images could not be achieved, some qualitative similarities of the general characteristics, in the case of rippled sediments were observed between the synthetic and real image regions.

Statistically, backscattering intensity in homogeneous, isotropic sidescan sonar imagery can be approximated by Rayleigh distributions (Bell, 1995.)³ In Bell and Linnett (1997) the pdfs of the simulated images (for silt and sand) were strongly consistent with fitted Rayleigh pdf's. Further there were close visual similarities between synthesised isotropic images and a real image of an isotropic region of the seafloor composed of a silt sediment. The synthetic image pdf, real image pdf and the fitted Rayleigh distribution again indicated a strong statistical consistency between the real and synthesised image regions. Noise and volume reverberation were not considered in their model, as these are dependent on the specific survey conditions. They state that it may be possible to extend the model to include more complex synthetic topographies and make use of the K-distribution as a means for statistical verification rather than visual inspection. So, within this framework, it is therefore feasible that Sabellaria-like

³ There are several other parametric pdf models which can be used to represent sonar imagery, such as the K-distribution (Jakeman and Pusey, 1976) and the Rayleigh-Rice model, discussed in Thorsos and Jackson (1989.)

textures could be synthesised and compared statistically with real imaged Sabellaria textures. For mostly practical reasons (data availability, time, funding) this interesting avenue of exploration lies outside the scope of the thesis.

3.2.3 Classification and clustering methods

Artificial Neural Network (ANN) and Self-Organising Map (SOM) approaches

There are several machine learning algorithms that have been applied to classification and segmentation tasks on seabed imagery. One of the most widely reported methods is the Artificial Neural Network (ANN).

According to Marsh and Brown (2009) ANN's have been applied successfully to sonar classification problems by several researchers. They claim this research work demonstrates that ANN perform better than conventional methods which typically require knowledge of underlying probability distributions in the data. Their classification scheme is a SOM implementation in Matlab, using a combination of backscatter and bathymetric features. They concluded that SOM's provide one of the most accurate methods for clustering acoustic swath data. However, with so many different ANN available it is not easy to state definitively which type is better, as there are not any standardised, published, benchmarking data or results for ANN's applied to seabed imagery.

ANN's have been used for texture classification of sonar images since the early 1990's, for instance, Shang and Brown (1992), Shang and Brown (1993), Stewart *et al.* (1992) and Stewart *et al.* (1994). Shang and Brown (1993) used a supervised approach, comprising a cascade of two ANN's for texture classification of sidescan images. They trained two, Multilayer Feed-forward Neural Networks (MFNN) and used GLCM image features. The first of the MFNN performed a principal components transformation to produce an uncorrelated feature set. The second MFNN carried out the classification. A back propagation algorithm was used to learn the weights of the two networks. Stewart *et al.* (1992) describe the use of a supervised ANN classifier with different training patterns, network architectures and various combinations of features. They conducted experiments to evaluate classification performance and found that using

hybrid features (a combination of twenty, spectral and grey level features) was a promising approach. Stewart *et al.* (1994) in a continuation of their work published two years earlier found that it was more important to select appropriate elements of the feature set and to arrange a set of representative training data than to increase the quantity of training data. However they noted that since the classification was based purely on textural features there could be some drawbacks. These included the effects of grazing angle, other system artifacts and high frequency information loss in the gridding (mosaicing) process.

Zhou and Chen (2005) used a cascaded approach, applied to MBES data. In the first stage a SOM produced coarse clusters. Output from the SOM formed the input to the second stage, in which the results were refined by applying a supervised, Proportional Learning Vector Quantisation (PLVQ). They compared their results against those obtained with a statistical, Bayesian Decision Theory approach, involving a probability model of features created on the data. In all of their test cases, the classification accuracy of the SOM/PLVQ was significantly better than the statistical model and the use of a SOM in isolation. Liu *et al.* (2005) proposed a new method of sonar image segmentation using snake models based on Cellular Neural Networks (CNN). Their results showed the approach was efficient and immune to noise when compared to results from other snake based models.

Martin and Osswald (2008) considered the uncertainty in the seabed environment and the sometimes conflicting interpretation of the information by different human experts. They conjectured that in order to train an automatic classification algorithm it is necessary to take into account the differences and uncertainties associated with each human expert. A Bayesian Belief Model and a fusion of the experts opinions was used to address the uncertainty issues. They found that using a MLP with belief learning significantly improved on the results of a classical MLP. Müller and Eagles (2007) concluded the combination of GLCIA and ANN had a wide range of potential applications in pattern recognition in remotely sensed geophysical images. Currently, there is just one commercial software package using an ANN for seabed classification, the Triton Imaging SeaClass⁴. SeaClass uses a supervised MLP, trained on sonar image textural features, using training regions selected by a human operator.

⁴ <http://www.tritonimaginginc.com/site/content/products/modules/seaclass/> [accessed 06-03-2012]

Decision Trees

Another machine learning algorithm that has been applied to seabed classification problems, is the Decision Tree (DT) classifier. Ierodionou *et al* (2006) applied an automated decision tree classification technique to a fusion of features from MBES bathymetric and backscatter data. The classification procedure was supervised and they used ground truth from a video sledge to assist in parameterising the model and assessing its accuracy. Dartnell and Gardner (2004) used a fusion of four images derived from one data set. These images comprised the original acoustic backscatter image and three variance images derived from backscatter and bathymetry. Their classification was a two-stage process. The first, supervised stage required some *a priori* knowledge of the seafloor. The results from this stage were used to generate rules for the subsequent hierarchical DT classification. The high intensity, specular reflection along the nadir was treated as a separate “noise” class, to avoid any adverse impact it may have had on the sediment classification results.

Clustering

The commercial seabed classification package, QTC swathview, uses Simulated Annealing (SA) with a measure of Bayesian Information, as a cost function, to search for the best clustering solution within its Automated Classification Engine (ACE) (Preston, 2009).

Standard, *k*-Means clustering is a commonly used technique, with researchers in the domain. Whilst *k*-Means is quite straightforward, it does have some limitations. Tan *et al.* (2006) point out that *k*-Means cannot handle non-globular clusters or clusters with different sizes and densities and they state that performance is also significantly affected by outliers. Blondel and Gómez Sichi (2009) claim that more recent domain specific classification applications have tended to use the *k*-Means algorithm. Their tests indicated that a simple Euclidean metric was sufficient for the classification application in their research. However, the user needs to have an idea of how many different clusters are expected in the data, as this must be specified before running the *k*-Means

algorithm. Unlike a SOM, *k*-Means does not determine the number of clusters automatically.

In Brown and Collier (2008) a procedure called ISOCLUST, in the IdrisiTM Geographical Information System (GIS) and image processing package was applied to carry out unsupervised classification of the backscatter imagery. They computed mean, median and standard deviation of intensity as features on the sidescan backscatter imagery, in neighborhoods corresponding to a 50 m × 50 m ground dimension. Depth was used as another feature and three output classes were selected based on multivariate analyses of the ground truth sites. A depth threshold partitioned one of the classes into two sub-classes. They achieved 78% reliability in habitat predictions (as defined by underwater video footage) from the automated classification of the acoustic imagery.

Other, proprietary approaches

Lucieer (2008, 2007b) used the proprietary software package, eCognition to perform segmentation and classification of sidescan imagery. The segmentation process uses region growing, to break the image into numerous irregular but contiguous patches of unequal size. An object oriented fuzzy-rule based classifier was subsequently applied, to label the regions as “reef”, “low-reef” and “sand.” In this supervised approach, mean and standard deviation of intensity and GLCM mean were found to be the three most discriminatory features.

Further coverage of various approaches can be found in Schumann *et al.* (2010), Christensen (2007), Anderson *et al.* (2007) and Hughes Clarke *et al.* (2009).

3.3 Bathymetry features

Relative to sidescan and MBES backscatter imagery, until recently, there has been comparatively little research into machine methods for feature generation and classification of rasterised MBES bathymetry. However, research into the geomorphometric analyses of terrestrial land surfaces and land form objects is extensive and there are numerous features which have been devised and machine learning tools

which have been applied, to a wide range of segmentation and classification tasks. One reason for this lag between DEM and DBM (also called Digital Depth Model (DDM)) has been the lack of high-resolution remote sensing instruments for submerged landforms (Zeiger *et al.*, 2009). The advent of MBES has revolutionised seabed mapping, with its capacity to capture accurate, wide coverage, densely sampled depth soundings from the seabed.

According to Harris and Baker (2012), in a survey of recent research, water depth is the most useful surrogate for communities of benthic biota. Water depth is important to the benthos as it influences the amount of light reaching the seabed, temperature, current strength and the impact of physical disturbances originating at the sea surface. In fact, depth was used as a surrogate in all 57 habitat case studies carried out by the researchers questioned in their survey. Despite its usefulness for habitat classification, in this thesis, depth *per se* at any point on the seabed is not important – it is not used as a feature. For geomorphometric analyses and target identification, features derived from local neighbourhoods of depth values over a range of spatial scales are far more useful as they contain information about the morphological properties of the seabed and the specific objects of interest. This reinforces the point made earlier, that the features used must be relevant to the objectives of the segmentation or classification task and objectives. Certain features useful in habitat classification may though, also be useful in terrain analyses or *vice versa*. Two examples are rugosity (the ratio of surface area to projected area) and Topographic Position Index (TPI)/ Bathymetric Position Index (BPI). Some examples of the many features derived from DEM and DBM for terrestrial and bathymetric terrain analyses respectively, are listed in table 3.1. There can be some variability in the definitions of the feature types and in the computational implementation of the feature creation algorithm, such as for the computation of directional gradients from which slope is derived. Further details can be found in the referenced work of the researchers.

Feature	Sub features	Researchers
Curvature	Maximum, minimum, Plan, profile	Bogaart and Troch (2006), Dolan <i>et al.</i> (2008), Ehsani and Quiel (2008), Fowler <i>et al.</i> (2008), Galparsoro <i>et al.</i> (2009), Holmes <i>et al.</i> (2008), Wilson <i>et al.</i> (2007), Wood (1996), Zieger <i>et al.</i> (2009).
Convexity		Iwahashi and Pike (2007)
Aspect/Slope		Dartnell and Gardner (2009), Dolan <i>et al.</i> (2008), Ehsani and Quiel (2008), Fowler <i>et al.</i> (2008), Franklin (1987), Galparsoro <i>et al.</i> (2009), Guinan <i>et al.</i> (2008), Henry <i>et al.</i> (2010), Holmes <i>et al.</i> (2008), Iwahashi and Pike (2007), Lundblad <i>et al.</i> (2004), Wedding and Friedlander (2008), Wilson <i>et al.</i> (2007), Wood (1996), Zieger <i>et al.</i> (2009)
BPI/TPI		Buhl-Mortensen <i>et al.</i> (2009), Dartnell and Gardner (2009), Diesing <i>et al.</i> (2009), Dolan <i>et al.</i> (2008), Galparsoro <i>et al.</i> (2009), Guinan <i>et al.</i> (2008), Henry <i>et al.</i> (2010), Lundblad <i>et al.</i> (2004), Weiss (2001), Wilson <i>et al.</i> (2007)
Fractal dimension		Dolan <i>et al.</i> (2008), Fowler <i>et al.</i> (2008), Wilson <i>et al.</i> (2007)
Rugosity		Dolan <i>et al.</i> (2008), Guinan <i>et al.</i> (2008), Holmes <i>et al.</i> (2008), Marsh and Brown (2009), Wedding and Friedlander (2008), Wilson <i>et al.</i> (2007), Zieger <i>et al.</i> (2009)
Terrain Ruggedness Index (TRI)		Dolan <i>et al.</i> (2008); Marsh and Brown (2009); Wilson <i>et al.</i> (2007)
Co-occurrence Matrix	Correlation, Contrast, Entropy, e.t.c.	Franklin and Peddle (1987), Wood (1996)
Local Fourier Histograms (LFH)		Cutter <i>et al.</i> (2003)
Autocorrelation		Wood (1996)
Morphometric descriptors	Channel, pass, peak, pit, plane, ridge	Ehsani and Quiel (2008), Iwahashi and Pike (2007), Lucieer (2007a), Wood (1996), Zieger <i>et al.</i> (2009)

Table 3.1. Examples of features derived from DEM and DBM for terrestrial and bathymetric terrain analyses respectively.

A review of several geomorphometric features used in a bathymetric context and an assessment of their utility as predictor variables for species distribution models is given

by Wilson *et al.* (2007). There are several commercial, non-commercial and academic GIS packages available which are capable of deriving features, from DEM's to represent terrestrial topography, see for instance, the Virtual Terrain Project (VTP)⁵. The survey of Harris and Baker (2012) found the commercial GIS package, ArcGIS⁶, is the most popular tool for analysing bathymetry and backscatter data, in relation to the study of the spatial distribution and composition of benthic communities. However, even though features such as slope and aspect are routinely used for habitat classification studies, there have been relatively few investigations into the features that can be used reliably for quantitative geomorphometry and the morphological characterisation of DBM's and in particular, specific landform objects within the DBM. Examples of the use of GIS packages applied to DBM analyses are described in, for instance, Wilson *et al.* (2007), Lucieer (2007a), Galparsoro *et al.* (2009) and Gorini (2009). As these researchers found, some of the features generated by GIS packages are useful when applied in a bathymetric context. The terrestrial GIS package, Landserf (Wood, 1996) was used by Wilson *et al.* (2007), Lucieer (2007) and Galparsoro *et al.* (2009) to generate features in their bathymetric terrain analyses for habitat studies. In Landserf, the DEM/DBM is represented as a quadratic surface. Six morphometric descriptors can be generated (derived from local curvature and gradient functions), these are; peak, pit, channel, pass, ridge and planar, as listed in table 3.1. "Local" is used to mean a feature value that is valid for a small region of raster, as defined under a kernel of a specified size.

The rocky reef (Lucieer, 2007a,b, Diesing *et al.*, 2009) is an example of a specific, distributed landform object, unique to the underwater environment. Pockmarks are another type. Visually, dense pockmark swarms are perhaps more comparable to the cratered surface of Mimas, one of Saturn's moons than any known terrestrial features. Since, the size, shape and surface properties of objects in the submarine environment are different to terrestrial land surface objects, feature kernels designed for capturing information about terrestrial objects are not necessarily effective or optimal for seabed object discrimination. In cases where the landform object is unique to the seabed and no similar objects are observed on the terrestrial surface, it may be necessary to develop a bespoke approach to say, the automated identification, segmentation, characterisation

⁵ <http://vterrain.org/> [accessed 17-03-2012]

⁶ <http://www.esri.com/software/arcgis> [accessed 05-12-2012]

and mapping of the object(s). Some examples relating to specific seabed objects in a DBM include; Rocky reef (Lucieer, 2007a,b), debris flows (Micallef *et al.*, 2007) and pockmarks (Fowler *et al.*, 2008, Gafeira *et al.* 2012).

3.4 Combining bathymetry and backscatter information

As described in chapter 2, the DBM is in some cases used to apply corrections for topography and angle of incidence to the MBES backscatter imagery, in an attempt to calibrate the backscatter intensities, prior to the classification process. Also, features derived from the DBM can be integrated with features derived from the backscatter imagery, as part of the classification process. The features may be used together to induce a classification model containing mixed backscatter and DBM features or in a (more challenging) post-classification fusion process, merging together the separate DBM and backscatter classifications. The combination of different data layers is appealing as it is likely to provide a more robust seabed classification.

Ierodionou *et al.* (2007) used a video ground truth database to estimate the separability of three substrate and seven biota classes, using the Jeffries-Matusita and transformed divergence measures in the commercial software package, ENVI (version 4.2)⁷. They next applied a DT classifier, with features derived from the bathymetry and backscatter data, together with the video information. Although the precise details of their methods and evaluation are not given they found that combining bathymetry and backscatter features improves class separability in the substrate and biota classifications, compared to using the bathymetry or backscatter independently.

Marsh and Brown (2009) used normalised backscatter features derived at the MBES beam level, together with features such as slope and seabed roughness, derived from the bathymetry. Their classification approach employs a SOM and can combine features from the two data types, ranging from bathymetry only, to backscatter only. Thus, a multitude of different class maps can be produced with various contributions of backscatter and bathymetry features in combination. They claim that the SOM is the most accurate means of clustering swathe data. Additionally, they state that a classification based on beam level angular response is preferable to one based on

⁷ <http://www.exelisvis.com/language/en-us/products/services/envi.aspx> [accessed 19-03-2012]

backscatter strength mosaics, as the angular response can be useful for discriminating different seabed sediments. However, no accuracy figures or quantitative evaluation is provided in their work. Further, the provisional classes determined by the SOM are not compared against the ground truth (as these data were not available).

Dartnell and Gardener (2004) used a MBES backscatter mosaic image and three derived images, based on local neighbourhood variance of the original backscatter mosaic and a collocated bathymetry raster. A 3×3 variance kernel was applied to the backscatter mosaic and to the bathymetry raster. An additional 11×11 kernel was applied to the bathymetry raster, to provide the same feature at small and large-scale resolutions. The feature sub-images and the original backscatter intensity image were subsequently clustered into five groups. The four sets of clustered results were further analysed by rule-based DT classification, implemented in the proprietary package ERDAS Imagine⁸, to predict the seabed facies. Three separate data sets were available for validating the predicted class map; sediment samples, seabed photographs and high-resolution seismic reflection profiles. The seismic sections are used to qualitatively validate the discrimination between rock outcrops and sediments. Photographic validation of seabed classes was also qualitative but accuracy of sediment facies prediction for three classes could be quantified by comparison with the physical samples and 72 % accuracy was achieved in this case.

There are many classification algorithms and paradigms, routinely used in satellite remote sensing and other areas that have so far not been considered or have received little attention, in the context of seabed classification problems. An example of a specific type of discriminative classifier is the Support Vector Machine (SVM) (Vapnik, 1982, 1998). With exceptions of Martin *et al.* (2006) and Lanaaya *et al.* (2005b) few researchers have used the SVM on sonar image classification problems, despite its widely acknowledged excellent performance on many types of problems. Naïve Bayes is another well known and widely used probabilistic classifier that has not been considered in any published studies on seabed classification problems.

⁸ <http://www.erdas.com/products/ERDASIMAGINE/ERDASIMAGINE/Details.aspx> [accessed 19-03-2012]

3.5 Ensemble methods

An ensemble or committee comprises a collection of machine-learning models and decision fusion strategy for combining the model predictions (Dietterich, 2000). See Sinha *et al.* (2008) for a recent survey of various decision fusion methods. Multiple models are created from different samples of training data or using different features and can be supervised (classifier ensemble) or unsupervised (cluster ensemble). Although supervised ensembles are in common use and have many real-world applications (Pal, 2007), they have received very little attention within the seabed classification domain.

3.5.1 Supervised ensemble

Figure 3.1 illustrates the components in a simple ensemble system. The input channels to the ensemble are sets of pattern instances, P_j generated from the data, D_j . The features (or their parameters and analysis scales) used to generate the patterns may be different and data could be from multiple sources. Model induction and subsequent classification of the input channels is performed by N classifiers $C_1 \dots C_N$. The predictions of the individual classification hypotheses are combined using a decision fusion strategy to produce the overall, consensus result of the classification process as an output.

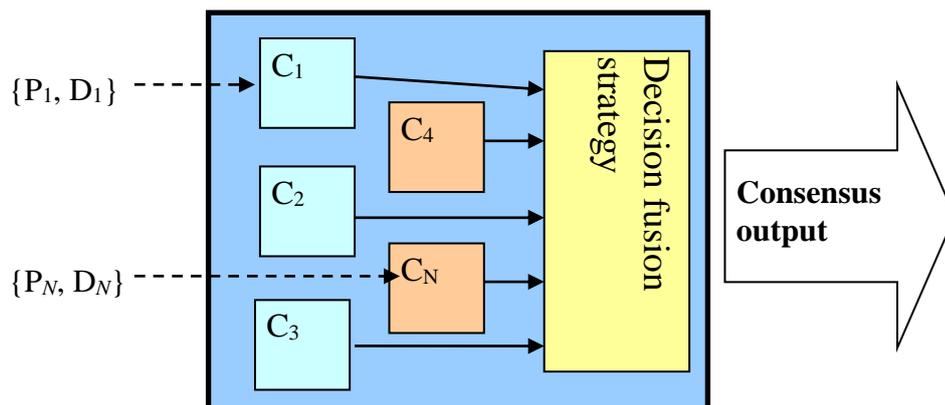


Figure 3.1. A classifier ensemble is made from a collection of classification algorithms and a decision fusion strategy.

Classification models can be induced by sampling different training instances, by using different feature groups, model parameters or data layers in the training phase. There are many possible variants of this approach, such as inducing multiple models on the same data layer, inducing different models for different dichotomies on the same data layer or even ensembles of models on multiple data layers such as backscatter and bathymetry.

Ensembles have found many commercial uses including, remote-sensing, e.g, Pal (2007), Steele and Patterson (2001), in predictive problems, such as Wang *et al.* (2005), control problems, e.g., Pardoe *et al.* (2005) and credit scoring, in Zhou *et al.* (2010).

One of the earliest applications of an ANN ensemble to a remote-sensing classification problem was by Cherkauer (1996). As a decision fusion strategy, he used a simple averaging of the output from individual ANN's to automatically detect and classify volcanoes in Radar images of Venus. Approaches using ensembles of DT's such as Random Forest (RF) (Breiman, 2001), have been used in land cover classification applications in Chan and Paelinckx (2008) and Gislason *et al.* (2006). Further applications of ANN and DT classifier ensembles to remotely sensed data are described in Oza and Tumer (2008). Applications of SVM ensembles to remote sensing classification problems are described in Chan *et al.* (2001), Zortea *et al.* (2007), Melgani and Bruzzone (2004), Bruzzone *et al.* (2006), Waske (2007).

Two studies on ensemble approaches have been carried out in the context of sidescan sonar imagery, by Martin *et al.* (2004) and Martin (2005). They investigated the fusion of classification results based on models created with different types of textural features. Generated patterns were input to a committee of MLP's, with each committee member associated to a discrete feature input channel. Channel outputs were combined in a decision fusion strategy based on evidence theory.

3.5.2 Unsupervised (cluster) ensemble

Clustering (unsupervised learning) concerns the problem of partitioning unlabelled patterns into groups so that similarity is maximised within the group and minimised between groups. Strehl and Ghosh (2002) call the problem of combining multiple partitions (different sets of clustered patterns), without accessing the original features or

the partitioning algorithms, the cluster ensemble problem. The problem is more challenging than creating a classifier ensemble, since the cluster labels are symbolic, there will be an issue of label correspondence. Further, the number of clusters in each partitioning may differ, depending on how the partitioning was created. According to Hadjitodorov *et al.* (2006), the different ways of building a cluster ensemble are;

- *Feature-distributed clustering*, where different feature subsets are used.
- *Heterogeneous or hybrid clustering*, using different clustering algorithms.
- *Homogeneous clustering*, where a random parameter is varied on the same algorithm.
- *Object distributed clustering*, where each ensemble member is induced with a different data sample.

Building the cluster ensemble is the first stage but the more difficult part is combining each clustering. Consider figure 3.2. Using a notation similar to that of Strehl and Ghosh (2002), let $X = \{x_1, x_2, \dots, x_n\}$ denote a set of pattern instances. A partitioning of the patterns into k clusters is $\{C_l \mid l = 1, \dots, k\}$, which can be represented by a label vector, λ . Each of the r partitioning algorithms, $\Phi^{(i)}$ takes the pattern instances as the input and outputs the label vector, $\lambda^{(i)}$. An algorithm dependent numerical support (posterior probability) vector $\rho^{(i)}$ may also be associated with each label instance in the output from the partitioning algorithm. Thus, the symbolic labels and their support may be used as inputs to a consensus labelling function or heuristic, Γ . The purpose of Γ is to combine the vectors, $\lambda^{(i)}$ (and $\rho^{(i)}$, if available) from the different partitioned sets, into a consensus label vector, L (with consensus support, if label support is input to the function.)

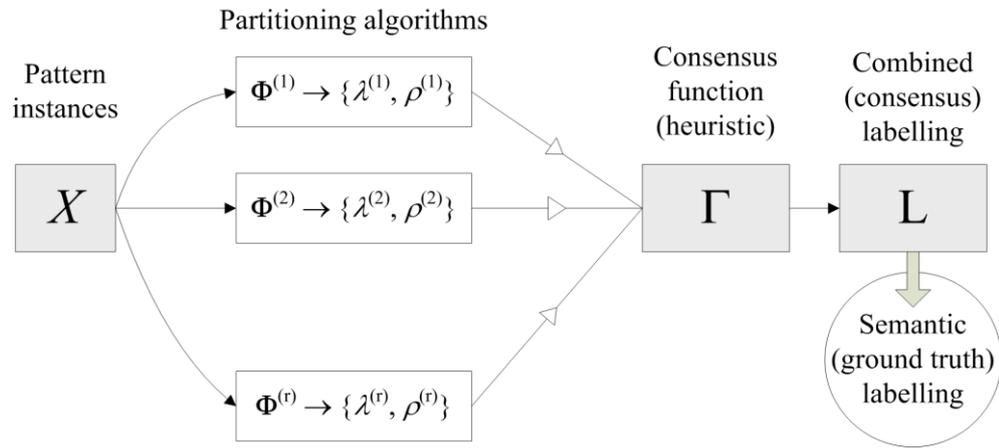


Figure 3.2. Basic components and process of a cluster ensemble.

A clustering selection process may also be used, which attempts to identify the best clusterings to combine in the consensus function. There are several ways of comparing different clusterings, such as the adjusted rand index (ARI) (Hubert and Arabie, 1985) and information theoretic measures, e.g. Normalised Mutual Information (NMI), of which there are a number of variants, as in, for instance, Vinh *et al.*(2010). Cluster comparison measures may also be used to measure the diversity of the clusterings. Kuncheva and Hadjitodorov (2004) showed that a diverse ensemble of clusterings can improve the clustering result.

The way in which the clusterings are combined depends on whether or not, aside from the cluster labels assigned to the pattern instances, any additional information is available, e.g., distances from cluster centres of the individual pattern instances. Hadjitodorov et al. (2006), outline four approaches to finding the consensus clustering (i.e. the resultant partitioning).

- *Direct approach*, requires relabeling of the individual clusterings so that the labels correspond, followed by a fusion of clusters with the same labels in for instance, Zhou and Tang (2006).
- *Feature-based approach*, the output of each clustering is treated as an “intermediate feature space” which is then re-clustered.

- *Hypergraph approach*, recasts the cluster ensemble as a hypergraph partitioning problem which may be solved using, say the Hypergraph Partitioning Algorithm (HGPA), as in Strehl and Ghosh (2002).
- *Pairwise (co-association)*, uses a coincidence (co-association) matrix from which the final clustering is derived.

Several of the different consensus function definitions and their relationships are described in Li *et al.* (2004).

Following the label combination process, the consensus vector L , may be replaced by associating the collections of symbolically labelled objects with a set of known semantic (ground truth) labels.

There is no general, universally accepted solution, methods or processes for the cluster ensemble problem. Strehl and Ghosh (2002) pointed out that very few approaches to combining clusterings had been proposed and the cluster ensemble approach was not widely used, in comparison to supervised ensembles. This same point was noted three years later by Zhou and Tang (2006) and again, more recently by Tumer and Agogino (2008).

3.6 Outline of machine learning methods used in the experimental work in this thesis

In machine learning systems, such as those engineered and investigated in the thesis, the machine learning algorithm (or base learner, for instance, a Naïve Bayes classifier) is but one component in a collection of computational methods comprising the learning system. The researcher uses the term “machine learning methods” to mean, in the broadest sense, methods that are collectively vital to the learning process and learning outcomes of the system as a whole.

Typically, the vital conceptual components and processes of the machine learning system comprise (1) an abstraction process, in the form of attribute (feature) creation to represent the concepts to be learnt. The attribute descriptions relate to global properties

of the objects they represent (Dietterich and Michalski, 1983, p.42). In this thesis, they are numerical feature vectors, i.e., patterns. (2) A means of evaluating the features to assess their relevance and subsequently, the provision of training examples to the learner. (3) In an inductive learning process (learning by example, as primarily used here) the classifier builds a hypothesis or model of what has been learnt from the training examples. The machine learning algorithm is treated as a “black box”. (4) There also needs to be a method for evaluating how the learning system has performed on the task, i.e., having learnt something about the concept, when class predictions about new (unseen) data instances are made, how does the prediction compare to what is already known about the instances. (5) In the case of ensemble or hybrid systems with more than one learning algorithm and therefore multiple models, a subsystem is needed for processing the results from the models within the learning system. This may create further design and engineering complexity in other areas such as the system evaluation and the system decision making process concerning model selection.

The focus of the methods in the thesis is biased towards feature creation and evaluation, since it is usually the case that the choice of features used to represent the targets is more important than the choice of machine learning algorithm. However, when we evaluate features and their parameters for a given classification task, the decision made about which features and parameter settings to use is not independent of the method used to measure the properties of interest. The researcher advocates a holistic approach to feature evaluation, using a mixture of methods in order to build up a clearer and more reliable picture of the saliency and reliability of the features on a specific data driven classification task. Many of these ideas will unfold in greater technical and analytical detail as the thesis progresses.

The remainder of this section outlines some of the key computational methods that are used in the experimental investigations in the thesis.

3.6.1 Computational kernels

The word “kernel” appears frequently in the thesis and refers to a computational sub-grid (or matrix) through which an algorithm is implemented. It can also be used in the mathematical sense to mean a matrix-like operation. For instance, non-linear SVM

kernel functions are implemented as a matrix of inner products between vectors representing points in the feature space. Generally then, a computational kernel implements an algorithm that transforms data values into a single or many-valued output.

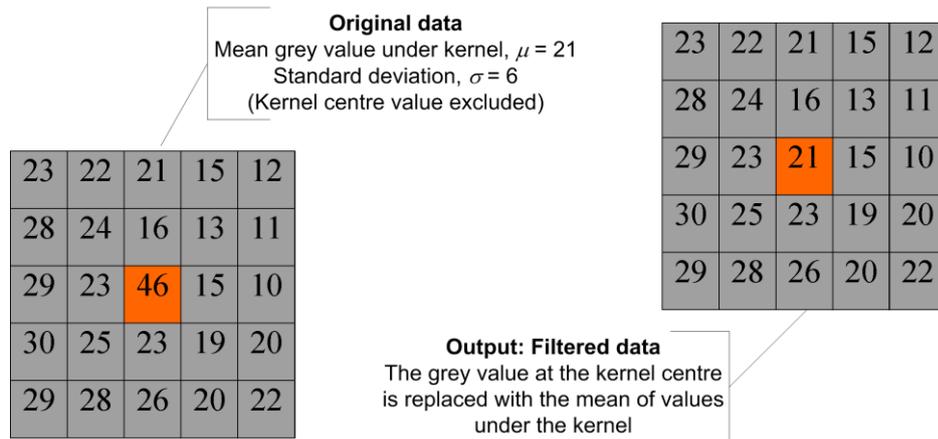


Figure 3.3. An example of a mean filter kernel.

Figure 3.3 illustrates a simple mean filter kernel. The kernel is an imaginary grid of a prescribed size (in this case 5×5 pixels or nodes) placed over a corresponding localised region of input data, e.g. a pixel neighbourhood or “tile” in a mosaic image. The algorithm replaces the pixel grey value at the centre of the kernel with the mean of the pixel neighbourhood (excluding the centre value) under the kernel. The output is single valued. The kernel may be used to downsample the data by reducing the number of pixels in the output image by up to a factor of 25 (in this example) if it is applied at contiguous locations, tessellating the entire input image. Alternatively, if the kernel is applied at each individual image pixel in the input image, the filtered output image contains the same number of pixels as the input. Some form of image edge padding is required in this case, e.g., mirroring the first and last two rows and columns of the input image.

The mean value filter is (excluding its size) a non-parametric kernel, since it contains no operator values. Certain parametric operations can be implemented through kernels comprising discretised values. The matrix of values is then applied to the corresponding input (image) neighbourhood values using some prescribed operation. For instance, in

the spatial domain implementation of Gabor filters, the 2-D filter kernel is discretised at regular intervals and these spatially ordered discretised values form the computational kernel, i.e., the matrix, A . The input matrix, B comprises a spatially ordered neighbourhood of (data) values, having the same dimension as A . The single-valued output from applying this kernel is equal to the sum of the Hadamard product, $\sum_{ij}(A \circ B)$ of the two matrices.

3.6.2 Feature evaluation and selection

As has already been pointed out, there are numerous different texture feature creation methods. Many of these are parametric. Depending on the type of features, changing the feature parameters changes the discretised values in the computational kernel for the feature, producing a different output result when the kernel is applied. Feature saliency is a measure of the importance of a particular feature and any specific parameterisations of the feature in the context of the input data and classification objectives. The features and their parameterisations often need to be evaluated in some way to determine their saliency for a prescribed task. By searching the parameter space, suitable parameter values can be selected so that the feature is tuned to perform well on the prescribed task.

Two established, widely used paradigms for feature (parameter) evaluation and selection are the filter and wrapper. Filters and wrappers may be used independently or combined in a more sophisticated hybrid evaluation procedure. The key difference between the two approaches is that filters do not require the induction of a machine learning model. A filter approach evaluates the intrinsic saliency of features independently of the machine learning algorithm. There are various kinds of filters. Distance or similarity measures are filters that are typically used for evaluating individual features.

A wrapper evaluates features by using a harness containing the base learning algorithm(s) and an objective function. Each individual feature or feature subset can be used to induce a different model from the learning algorithm using a set of training instances. When saliency is measured in a wrapper using say, classification accuracy as an objective function, an individual feature yielding a relatively high accuracy compared to the others could be described as the most salient individual. It is sometimes

desirable to have high values of sensitivity and specificity together with high accuracy. Thresholds for accuracy are occasionally prescribed together with thresholds for sensitivity and specificity in a conditional objective function, where all conditions need to be satisfied for the feature to be selected. For instance, finding feature parameterisations that result in $\{\text{accuracy} > a \text{ and sensitivity} > b \text{ and specificity} > c\}$ where a , b , and c are the threshold values. However, these accuracy metrics alone say nothing of the important concepts of stability of the feature in the presence of data variability or robustness of the feature to changes in parameters and classification objectives.

3.6.2 Supervised learning

(1) Support Vector Machine (SVM)

The Support Vector Machine (SVM) (Vapnik, 1995, 1998) is a binary classifier with a classification hypothesis (model) whose decision boundary is a discriminatory hyperplane. It is a maximal margin classifier. During the induction stage the optimal model with the greatest separation between the decision boundary and the n -dimensional patterns, \mathbf{x}_j in the training data set is found. The optimal hyperplane is represented using a subset of training patterns known as Support Vectors.

For a linear SVM with separable training patterns, the binary classification function, $\Omega(\mathbf{x}_i)$, for determining class membership of an unseen pattern, \mathbf{x}_i is, $\Omega(\mathbf{x}_i) = \text{sgn}(w \cdot \mathbf{x}_i - \beta)$ where w is the unit normal vector to the hyperplane and β is the (minimum) distance of the hyperplane from the origin. The class label $y_i \in \{\pm 1\}$ is assigned to a test instance according to whether $\Omega(\mathbf{x}_i)$ is positive or negative, which depends on the side of the hyperplane the test instance lies on in the feature space.

Maximising the margin during the induction stage is equivalent to minimising the Euclidean norm of the vector, w , subject to the condition, $y_j (w \cdot \mathbf{x}_j + \beta) \geq 1$. The problem can be reformulated as a Lagrangian dual that is solved using numerical techniques such as quadratic programming (QP) to find the optimal model parameters. When the training patterns are not separable, a so-called soft margin approach is used which permits small training errors, with an associated cost (penalty.) The objective

function still requires minimisation of the Euclidean norm of w but the inequality constraint, $y_j (w \cdot \mathbf{x}_j + \beta) \geq 1$ can no longer be satisfied. Thus, the optimisation problem now concerns maximising the soft margin subject to the trade-off of the cost of training errors.

Non-linear SVM's can be effectively implemented implicitly, using the kernel trick. The basic idea is to compute the similarity between pattern instances in the original feature space using a kernel function. It is a far more efficient approach, compared to transforming the pattern instances explicitly to a new (higher dimensional) feature space where the patterns are linearly separable.

Linear and non-linear SVM's models were induced in Active Learning experiments (not documented in the thesis.) The SVM is outlined here though due to its very close relationship to the Ball Vector Machine (BVM.)

(2) *Ball Vector Machine (BVM)*

The Ball Vector Machine (BVM) (Tsang *et al.*, 2007) is a type of SVM that uses Minimum Enclosing Balls (MEB) rather than QP or other numerical techniques, to solve the optimisation problem in the model induction stage. The C++ BVM implementation⁸ by Tsang *et al.* (2007) is an extension of the LibCVM/LibSVM toolkit.

(3) *Naïve Bayes Classifier*

Unlike the discriminatory SVM and BVM, the Naïve Bayes classifier is probabilistic. It is a multi-class classifier, based on Bayes' theorem. Only a small amount of training data is needed for the maximum likelihood estimates of the model parameters from the unknown probability distributions of the classes during induction. In the prediction stage, decision support, output with the corresponding class labels from the classifier, is the probability that the labelled pattern instance belongs to the assigned class.

⁸ LibCVM Toolkit Version: 2.2 (beta), How to use. <http://www.cs.ust.hk/ivor/cvm.html> [Accessed 23-07-2010]

The classifier is called “naïve” because of the naïve assumption that features used to train the model are conditionally independent, given the class, i.e., $P(\mathbf{x} | \omega) = \prod_{j=1}^n P(x_j | \omega)$, where, \mathbf{x} is a feature vector, $x_j \in \mathbf{x}$ and ω , a class. Although in practice it is known to work well even if the conditional independence assumption is not satisfied (Domingos and Pazzani, 1997) and also for functionally dependent features, as in Rish (2001.) The Naïve Bayes classifier from the Matlab library is used in the experiments in case studies 3 and 4.

(4) *K-Nearest Neighbours (k-NN)*

The *k*-Nearest Neighbours (*k-NN*) classifier (Fix and Hodges, 1951) differs again from the others outlined previously. Aside from setting the number of neighbours, *k*, there are no parameters. The *k-NN* is a lazy learner, as it has no explicit training or generalisation phase. The “model” is defined implicitly by stored data instances and classification rules.

The algorithm works by computing the distance (or similarity) between a query instance and its *k-NN* in the “model” data set. In the case where all *k* instances are equally weighted, the predicted label of the query instance is the modal class label of its *k-NN*. The contribution to the classification decision of the individuals in the *k-NN* can also be weighted according to their distances from the query instance, as in Dudani (1976.)

3.6.3 Unsupervised learning

(1) *k-Means*

The *k*-Means is a prototype-based clustering algorithm that uses the centroids of clusters as the prototypes. It groups pattern instances into clusters in a feature space by minimising the sum of squares of distances between the patterns and the cluster centroids. Initially, the number of clusters (*k*) needs to be specified or estimated in a cluster evaluation procedure. Random points may be assigned as the initial centroids. *K*-Means clustering is an iterative process and the key steps are (1) compute centroids (2) determine the distance of each instance from the centroids (3) group the instances into clusters based on the minimum distance from the centroids (4) update the centroid

locations. Instances may move from one cluster to another in the iteration and this continues until convergence, when there are no further changes in the instances belonging to each cluster.

(2) *Self-Organising Map (SOM)*

The Self Organising Map (SOM) is a type of unsupervised ANN. It can be thought of as a prototype clustering method, with the centroids topographically prearranged in a fixed sequence, i.e. a map. In a 2-D map, the centroids (cells) close together represent data points (pattern instances) that are similar in some way, in the input space. Therefore, the SOM can serve as a cluster analysis tool (as is the case in the thesis), since it preserves the topology and relative distance between patterns.

3.7 Summary

The aim of this chapter was to identify some of the underexplored (also some of the more commonly used) computational methods in the seabed image classification domain and hence confine the scope of the computational methods to be considered within the application focus of the thesis. The key points arising from this chapter are:

- Using texture as a discriminative subspace is probably the commonest approach to classifying seabed imagery. The most frequently used texture features are derived from Grey level co-occurrence matrices (GLCM). Many researchers and commercial processing packages use these features.
- Descriptive statistics can be used as features on sidescan imagery since mean backscatter intensity can be correlated to sediment type.
- Signal processing methods, such as Gabor Filter banks, despite their widespread use elsewhere are rarely considered for feature creation in the seabed classification domain and are not currently used in any commercial processing packages.

- Image texture formation models may offer a promising approach, as part of a machine system for the discrimination of specific textural targets such as Sabellaria. However this was not a viable research pathway for the thesis as there were no data available for validation.
- In order to work effectively, parametric feature kernels need some degree of tuning for the specific data dependent classification objectives. There is no universally accepted method for feature (and feature parameter) evaluation on seabed imagery. There have been very few domain specific studies concerning feature evaluation methods and none regarding the robustness of the methods.
- There are no published studies on the machine discrimination of Sabellaria textures. It is not known what features could be useful for the novel task of discriminating these textures in waterfall and mosaic imagery.
- Most studies use their own data, classification objectives, features, classifier, processing strategies and class labeling system. In some published studies, there is no external validation. It is all but impossible to evaluate the work of different researchers and each has to be considered on its own merits for the specific problem being studied. This issue is explored further in the next chapter.
- There have been few published case studies on machine discrimination of specific landform objects in DBM's and what features are suitable for these novel tasks.
- More sophisticated classification approaches, using hybrid methods and unsupervised ensembles have received very little attention in the domain.

The aim of this chapter has been achieved. As will be seen as the thesis progresses, many of the points raised above concerning the computational methods will be revisited and explored further, within the case studies. In the next chapter, the research analysis and design is presented. The issue of evaluation and standardisation of results is also discussed as it is not easy to specify at this stage.

Research analysis and design

Chapter 4

Contents

4.1 Introduction

4.2 Research analysis and design

4.2.1 Background

4.2.2 Aim, objectives and scope of the project and case studies

4.2.3 Outline of research type and general methodology

4.2.4 Generic case study design

4.2.5 Timeline of general project activities

4.2.6 Risk awareness

4.3 Discussion of evaluation issues

4.1 Introduction

This chapter concerns the research analysis and design i.e., the approach to the research, how it was executed and the methods, processes and evaluation issues involved. Section 4.2 covers the main aspects of the research analysis and design, focusing on the framework for the research at global and case study levels. The aims, objectives and scope are stated and a general description of the research type and methodology is given. Generic research design as applied at a case study level is then outlined. As an industrial project, some extraordinary issues arose which led to major changes in the direction, design and duration of the research. Some of these factors are considered further in the research management section of chapter 9 but they are briefly stated in this chapter, in the context of the project timeline (4.2.5). The main (known) risks for the project are also identified.

Section 4.3 discusses some the evaluation issues. The results are uncertain in human interpretative and machine approaches to the subjective seabed target discrimination problems considered. Issues relating to interpretation and mapping, external validation

and standardisation are outlined and discussed. This section includes some original work to elucidate one of the problems and support the argument.

4.2 Research analysis and design

Section 4.2.1 presents a general background to the research project and type of research. In section 4.2.2 the general aims, objectives and scope of the project and the individual case studies are stated. The type of research and general methodology is outlined in section 4.2.3. Research design at a case study level is considered in section 4.2.4. The timeline of general project activities is summarised in section 4.2.5. Two key risks are identified in 4.2.6.

4.2.1 Background

The research in this thesis concerns the creation and evaluation of novel processes and methods (ideas) in a virtual (computational) experimental space. From a practical perspective, these artifacts could potentially be useful if implemented in an industrial context. In particular, for assisting with laborious and expensive office based classification related tasks at the human interpretative end of a data processing pipeline. In order to be described as research, knowledge and understanding must be advanced in some way. The thesis provides evidence to show that a context-limited advancement has been achieved in the course of the project. The purpose of this section though, is to describe the research framework that has facilitated the accomplishment of this goal.

4.2.2 Aim, objectives and scope of the project and case studies

General aim

The general aim of the research is to devise and investigate novel processes and methods that could potentially be used as part of a machine learning system for assisting humans with the subjective (interpretative) discrimination and classification of natural targets in seabed imagery. The general objectives reflect the aims of the case studies.

General objectives

1. Propose and investigate a novel machine learning process for pockmark object discrimination and boundary mapping in a DBM.
2. Investigate if the machine discrimination of Sabellaria textures in sidescan waterfall and mosaic imagery is a tractable task. Determine suitable features, feature configurations and parameters for this task.
3. Design a framework for evaluating the robustness of distance measures for feature evaluation and ranking. Investigate properties of a novel committee of distance measures for feature evaluation and ranking on sonar imagery.
4. Investigate how to apply an unsupervised machine learning processes to larger real-world sonar mosaic imagery. Devise novel methods and processes to meet the challenges of this task. Qualitatively evaluate the merits of a fully automated supervised and novel, hybrid unsupervised approach (including an unsupervised ensemble approach) to a target discrimination task in a real-world mosaic image.

General scope

Technical methods scope is limited to feature based machine learning processes and methods summarised at the end of chapter 3.

The problems are data driven and the data scope is confined to sidescan sonar mosaic, and waterfall imagery and MBES bathymetry in the form of a DBM. All data are rasterised and processed. No further processing is applied. The problems and data are considered at a phenomenological, human interpretive level.

The application scope concerns the two previously identified topics for investigation in chapter 2; pockmark discrimination and Sabellaria discrimination. The sediment discrimination task in case study 4 is a contingent application.

Situated within the general aims and scope are the four case studies. Each case study has its own aim, objectives and scope, as summarised in table 4.1. Thus the general scope is defined implicitly by the scope of the case studies.

[1] Pockmark object discrimination and mapping

<i>Problem statement/question</i>	Manually identifying and mapping pockmarks rendered in a DBM is a time consuming, inconsistent and expensive manual process. Can aspects of this process be automated with a machine learning system?
<i>Aim</i>	Design methods and a process for supervised discrimination and mapping of the pockmarks.
<i>Objectives</i>	<p>1.1. Investigate types and combinations of features that could be useful for representing and discriminating between grid-node neighbourhoods belonging to pockmark objects and those that do not. Identify the most useful features using classification accuracy as the objective function.</p> <p>1.2. Compare the machine identification of the pockmark landform objects to the manual human identification by counting the objects identified.</p>
<i>Scope</i>	DBM data from the Geological Survey of Norway. Supervised feature based machine learning, with manual, human selection of training points. Discrimination of pockmark grid nodes neighbourhoods and mapping the boundaries of pockmark objects.

[2] Sabellaria texture discrimination

<i>Problem statement/question</i>	Manual, interpretative discriminating and mapping of Sabellaria textures rendered in sidescan sonar imagery is a time consuming, inconsistent and expensive manual process. Can aspects of this process be automated with a machine learning system?
<i>Aim</i>	Identify a useful feature creation method and parameter ranges for efficiently discriminating Sabellaria textures from other textures in sonar waterfall and mosaic imagery.
<i>Objectives</i>	<p>2.1. Evaluate different state-of-the-art feature creation methods on waterfall imagery and identify any promising methods. Use Sabellaria classification accuracy and overall classification accuracy as objective functions for the evaluation.</p> <p>2.2. Investigate, in depth, a feature creation method (from objective 1) for Sabellaria discrimination in sonar mosaic imagery and identify a suitable configuration and set of parameters for the chosen feature creation method. Use classification accuracy and cost as objective functions for heuristically evaluating the configurations and parameters.</p>
<i>Scope</i>	Mosaic and waterfall dataset 1. Sabellaria texture discrimination on small sized image patches.

[3] Feature evaluation and ranking using distance measures

<i>Problem statement/question</i>	Which distance measures are useful for evaluating and ranking features and feature parameters on sonar imagery? Can a committee of distance measures provide more reliable results than an individual? How robust are the feature evaluation methods?
<i>Aim</i>	To find out which distance measures can be used for a robust evaluation and ranking of parametric features, applied to sonar imagery and establish if a committee of distance measures can improve the robustness, compared to an individual.
<i>Objectives</i>	<p>3.1. Establish which individual distance measure(s) are most reliable for feature evaluation and ranking on sonar imagery. Evaluate the methods by considering the correlation to classification accuracy, rank correlation and the variability of these properties under different conditions.</p> <p>3.2. Establish if there is any advantage in using a committee of distance measurements on sonar imagery and what the limitations are. Evaluate as described in objective 1.</p>
<i>Scope</i>	Mosaic and waterfall dataset 1, small image patches. Outex benchmark textures. Grey Level Co-occurrence Matrix features.

[4] Unsupervised classification of sonar imagery

<i>Problem statement/question</i>	How can multi-class supervised and unsupervised classification techniques be applied to larger sized, real-world mosaic images? What are the main technical issues, advantages and disadvantages of the supervised and unsupervised approaches? Is there any justification for using an ensemble approach? Which approach is better for discriminating between sediment types?
<i>Aim</i>	Devise an unsupervised process, including a fusion method for multi-class, multi-model unsupervised discrimination and subsequent classification of interesting regions in full-sized sonar mosaic imagery.
<i>Objectives</i>	<p>4.1 Devise a heuristic, hybrid machine learning process for the unsupervised classification of sonar mosaic imagery.</p> <p>4.2 Qualitatively evaluate the plausibility of the process and results by comparison with a supervised classification and a specimen, manually produced classification.</p>
<i>Scope</i>	USGS sonar mosaic imagery, ground truth descriptions and sediment class map. Sediment type (grain size) discrimination.

Table 4.1 Summary of the aims, scope and objectives for each case study.

4.2.3 Outline of research type and general methodology

As an industry-based project, it is predominantly about “problem solving research” as defined by Phillips and Pugh (2005, p 52.) It is also a form of action research (Lewin, 1946, Susman and Evered, 1978) since the consequences of the research could potentially bring about a change in the way certain manual tasks are currently carried out in the workplace. Therefore, by implication, it must be applied research. It is also pragmatic in the sense that an aim is to create knowledge and methods to improve specific situations. Established computing science research methodologies are used in the thesis; case study, proof of concept and empirical investigation (Ramesh *et al.*, 2004.)

Figure 4.1 illustrates a product of the research (the four case studies ordered from top to bottom along a timeline) and the conceptualised framework for the research. The case studies are aim driven, and each has its own individual (but coherent) objectives and scope that are subsets of the overall aim, objectives and scope of the thesis. This serves as one constraint on the research design at the case study level. Another constraint is the desired output or outcome from the case research. These endpoints facilitate the construction of a research design for the studies. This is described further in subsection 4.2.4. The case study research designs are components of the overarching or global methodology concerning the research process, methods and procedures. Each case study therefore is a distinct but coherent component of the thesis.

The research is primarily qualitative, exploratory, descriptive and evaluative. It cannot be described as fundamental or explanatory, since the purpose is not to investigate causality with testable hypotheses. In fact, it is not possible to construct reliable hypotheses when the facts are uncertain. Much of the work carried out concerns observing outcomes or generating numerical results from data in virtual experimental spaces and subsequently interpreting and analysing those results. It is therefore empirical research.

There are three principal benefits of using an experimental computing science approach to the case studies, as in Tichy (1998);

1. The knowledge base can be made more reliable by reducing the uncertainty about, methods and tools which are appropriate for the task.
2. New and unexpected insights can arise in the course of the experimental work.
3. Useless approaches and incorrect assumptions can be identified and eliminated. This helps to align ideas and engineering in more fruitful directions.

The researcher has no control over data collection and processing prior to the point of applying the machine learning processes, so in this respect, the work is phenomenological and specific to the individual data set. Although, a case study structure is suitable for this type of research and the different themes investigated, one limitation is that it is not easy to generalise. Whilst inductive and deductive reasoning processes are used, the research is not fully inductive (or deductive) due to the aforementioned limitations of hypothesis construction and generalisation within the individual case studies.

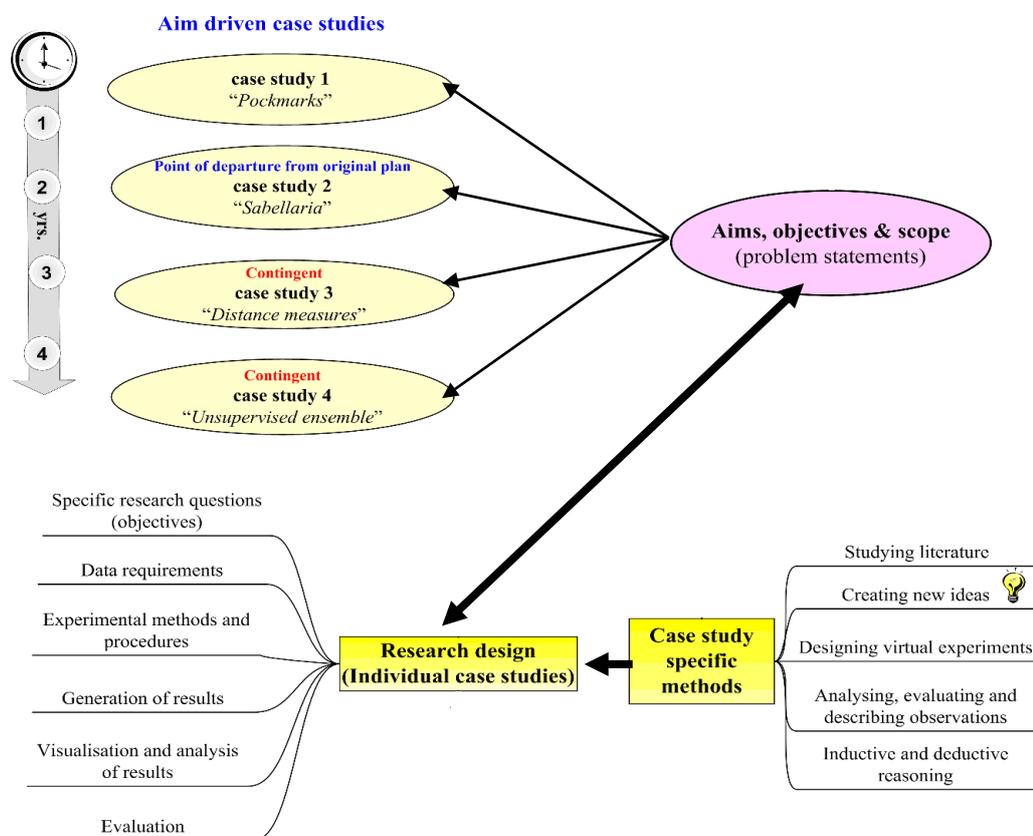


Figure 4.1 Conceptual layout showing the generic research design for the case studies and the methods used.

The research methods are mixed, and a combination of qualitative and quantitative techniques is used. Again, this follows from the fact that the data are qualitative and uncertain and whilst quantitative analyses can be applied, the outcomes are subjective. Mixed methods can lead to a holistic understanding of the processes and outcomes from the process (Bazely, 1999, p 284.)

4.2.4 Generic case study design

Having described some of the global characteristics of the research, attention is now turned to the case study level and the attributes associated with the yellow boxes, “Research design” and “case study specific methods” in figure 4.1.

There are differences in detail regarding the research design and process for each of the case studies. The design and process for the individual studies is implicit within the case study chapters. However, the templates for each study are quite similar. This section describes a generic design and states the methods, broadly applicable to all of the case studies. Evaluation of the results is considered as a separate topic in section 4.3. The key design factors are as follows:

Data

Consideration is given to what data are required (to meet the objectives, if already known) and where the data will come from. It is lower risk and a sensible precaution to devise problem statements and design the research around available data than to propose research based on data that is not yet available (see also, section 4.2.6). Instances where data were available at the inception of the research are indicated in table 4.2.

Problem statement (objectives)

A simple test for the objectives is the SMART rule: Specific, Measurable, Agreed, Realistic and Time constrained. Most of the objectives in each case study satisfy this rule, as summarised in table 4.2.

Objective	Data	S	M	A	R	T
1.1	✓	✓	✓	✓	✓	✓
1.2	✓	✓	✓	✓	✓	✓
2.1	✓	✓	✓	✓	✓	✓
2.2	✓	✓	✓	✓	✓	✓
3.1	✓	✓	✓	✓	✓	✓
3.2	✓	✓	✓	✓	✓	✓
4.1	-	✓	✓	✓	✓	✓
4.2	✓	✓	✓	✓	✓	✓

Table 4.2 Summary of objectives according to the SMART criteria.

Experimental methods and procedures

The experiments are virtual, involving the evaluation of an idea in a virtual experimental space. A computational challenge in this study was the lack of existing software tools and frameworks required to carry out the experimental tasks. Consequently, many of the algorithmic methods (excluding the core learning algorithms such as Naïve Bayes and k -means) implemented within the experimental harnesses and the design and implementation of the harnesses themselves, had to be done from first principles.

In deciding on the dimensions of the experimental space, it is essential to identify a few important metrics that facilitate a sensible measurement or representation of the desired properties of the concept under test. It is not realistic in the cases here, to explore more than a few properties due to the computational time required and the additional difficulty of analysing and making sense of the results. Deciding on suitable metrics (or proposing new ones) capturing the properties of interest therefore requires a thorough understanding of the problem and the domain literature. It is also important to decide which parameters will remain fixed and which can be varied. Detailing the specifications and the parameters for the experimental space constrains the problem to a manageable size and also helps to ensure repeatability (reliability), i.e., given the same data, experimental set-up and parameters, an independent researcher would obtain identical results.

Generation of results, visualisation and analysis of results

In all of the case studies, the raw results output from the experimental harnesses are numerical. Visual representation is preferable for interpretation, understanding and communication of the results. This entails another layer of processing, applied to the results output from the experimental harnesses. Each study required its own results analysis and visualisation layer. Generally, the simplest method of analysis was chosen.

4.2.5 Timeline of general project activities

The timeline of research activities (in their most general terms) is summarised in figure 4.2. Blue boxes correspond to part-time activities and orange boxes to full-time. Eight months were spent on preliminary investigations prior to officially commencing the full-time project. Much of the first-year work was devoted to studying domain literature, investigating possible research pathways and on practical development and experimental work relating to the discrimination of pockmarks in MBES bathymetry data. Some time after the fundamental direction of the research was changed to the discrimination of Sabellaria colonies in sidescan sonar data, it became apparent that data required to continue with and bring to conclusion this line of research would not be available. Contingent investigations were thus carried out in another direction using different data. A more detailed consideration is given in chapter 9.

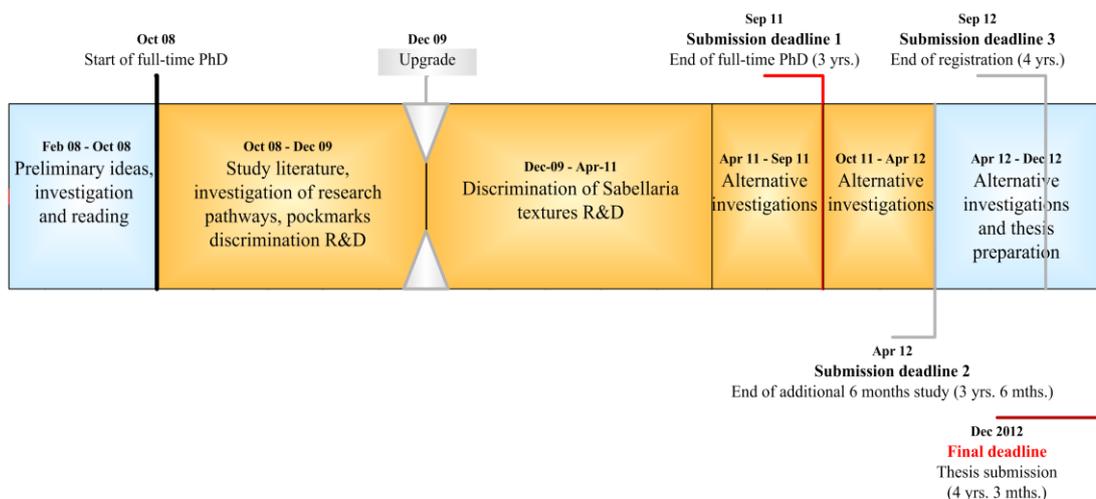


Figure 4.2 Research timeline (general themes and deadlines)

4.2.6 Risk awareness

A research project whose outcomes depend on several, intensive software design and development stages for the experimental harnesses and also on the provision of data has a very high level of intrinsic risk.

An interesting viewpoint is that building a thesis from a diverse collection of data and experimental harnesses in the case study contexts mitigates the risk of problems due to dependency on a single theme centred on one experimental harness and data set. For example, if all of the work were based on a single piece of software that was discovered to be faulty at a late stage, it could potentially have implications for the entire project and all results in the thesis.

The data provision/acquisition had three contingencies (in addition to the original data set which later turned out to be incomplete.) Again, this is considered further in chapter 9.

4.3 Discussion of evaluation issues

The purpose of this section is to describe and discuss some of the evaluation issues associated with subjective, interpretative manual target discrimination and classification tasks and the machine automation of these tasks. The results in general, from interpretative human classification are not as repeatable or consistent as those obtained from statistical, acoustic seabed classification methods. Anderson *et al.* (2008) present an argument in favour of statistical, objective classification procedures over manual interpretative classification. However, Christensen (2007) found that some of the results from automatic classification software were rather poor and claims that the automatic classification usually has greater uncertainty than manual classification. As has been pointed out, the work in this thesis concerns the automation of interpretative classification tasks (which is quite different from automated acoustic seabed classification). The application focus of the thesis is on tasks that cannot currently be achieved by more objective acoustical seabed classification methods. Thus, measuring and assessing the efficacy of the machine processes in the thesis presents a number of challenges.

There are several examples of published work on seabed classification problems that have good external validation, even though the methods, processes and classification objectives may not be standardised. In these cases, the results of the machine classification process applied to the sonar imagery are supported by various forms of ground truth, for instance; (1) PSA from grab samples for confirming sediment types, (2) shallow seismic sections for confirming subsurface geology and regions of rock outcropping and surface sediments, (3) photographic or video evidence of seabed substrate. See, for instance, Dartnell and Gardner (2004), McGonigle *et al.* (2009) and Brown and Collier (2008).

As an example of a validation issue, in the case of pockmarks, it is not easy or inexpensive to go to the bottom of the sea and take the various measurements required to confirm the efficacy of a human or machine discrimination of these objects in a DBM. Further issues arise, for instance, if the pockmark boundary is not defined in some way. There may be considerable disagreement amongst experts on what exactly the boundary is, as visually rendered in the DBM. Invariably, some simplifying assumptions have to be made in order to make progress with machine processes for discriminating such targets. In Gafeira *et al.* (2012), a watershed method is applied and identified objects are treated geometrically as polygons. A combination of two thresholds for minimum area and area/perimeter ratio are used to reject polygons that are less likely to be pockmarks. It is different to the approach of case study 1, in the next chapter of the thesis, where, the assumption is made that all identified objects are pockmarks. They are both plausible assumptions, based on knowledge and context. In case study 1, the object densities are 2-6 times greater than the highest density case considered by Gafeira *et al.* (2012). The feature based approach is different too and the resolution of the features is tuned to identify objects of a particular size range. Thus small, irrelevant objects are likely to be filtered by the features as they are below the resolution threshold. Having made these simplifying assumptions, a quantitative comparison (evaluation) between machine and human object discrimination is facilitated, such as the number of objects identified and the speed with which the task was done. However, because of the differences in the methods, processing flows and assumptions, the lack of common benchmarking data and standardised evaluation metrics, it is not easy to make a comparison of the approach with the work of other researchers. The same is true for much work in the seabed classification domain, as

there is such a great diversity in classification objectives, methods, processes and labelling of entities, generally with no widely accepted common framework or data sets for benchmarking new methods.

Natural entities on the seabed are often heterogeneous and distributed, occurring at a variety of spatial scales, such as the case with Sabellaria colonies. Human interpretation, despite its subjectivity, is perhaps the most widely used method for identifying and delineating Sabellaria texture surrogates in sonar imagery.

To highlight some of the issues involved in trying to establish a meaningful evaluation, of human (manual) vs. machine, a simple controlled experiment was carried out. The example in figure 4.3 illustrates inter-rater variability between five independent human raters (sonar interpreters) when tasked with discriminating regions of potential Sabellaria colonisation from the background in a small region of waterfall imagery.

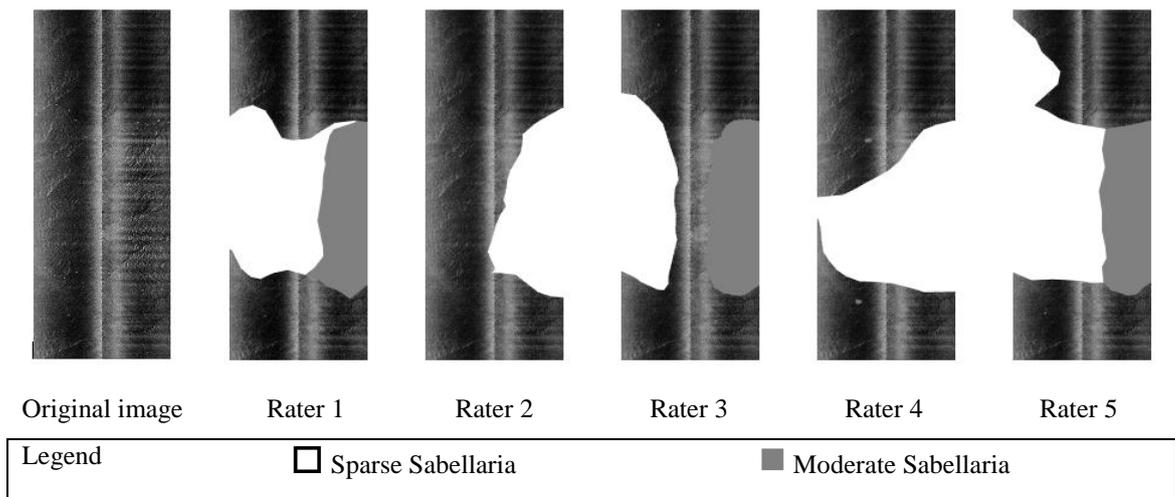


Figure 4.3. Manual segmentation and classification of the same waterfall image section (far-left) by 5 independent human raters. Block white areas indicate Sparse Sabellaria and block grey areas, moderate Sabellaria. All other regions not block shaded represent sand (as in fig. 4.5.)

The raters were presented with identical versions of the original image shown on the far left of figure 4.3. Water column removal and slant range correction had already been applied beforehand in CodaGeosurvey processing software. Usually, individual interpreters will apply their own preferred enhancements to the imagery to assist with the visual examination of the textures, including changing brightness and contrast, filtering and histogram equalisation.

In the experiment, the raters could not reprocess the raw data to alter the image. This circumvents the potential problem pointed out by Johnson and Helferty (1990) that sidescan imagery can be enhanced by a human, to match a preconceived idea of how the processed image of the seafloor should appear. The interpretation was constrained to the identification of up to three prescribed classes in the imagery, as shown in figure 4.4.

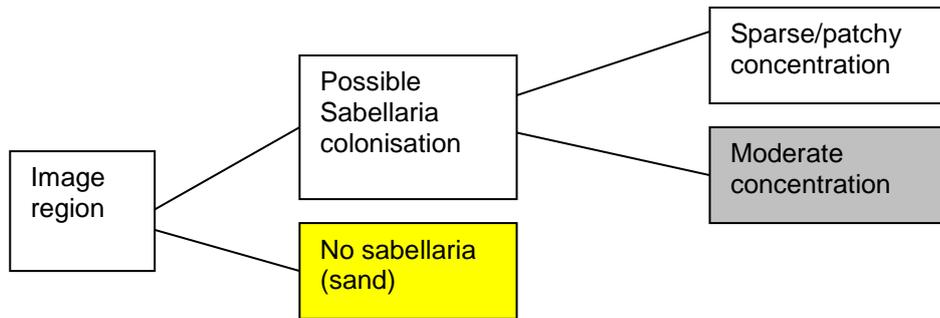


Figure 4.4. Class prescriptions for the image region shown in figure 4.3

The target class, (T) is defined as the textural region(s) in the image which potentially represent Sabellaria colonisation. The target has two possible sub-class definitions, “sparse/patchy concentration” and “moderate concentration”, shaded white and grey, respectively (if distinguished) in the interpreted images of figure 4.3. Regions not belonging to the target class ($\neg T$) are defined as “sand”. The regions are associated with class labels by ground truth but the true boundaries of the regions are undefined (and unknowable). The meanings of the sub-class labels are imprecise and describe the qualitative degree of patchiness or ground coverage of the colonies. There is no defined, quantifiable distinction between the two subjective labels.

Individual raters expressed a high (100 %) degree of confidence in the correctness their interpretations. The result of the experiment was five different interpretations, with unequal numbers of classes and significant inter-rater variability in the delineated class regions, shown in figure 4.3. Although the experiment was not repeated later, it is highly likely that intra-rater variability would also be observed, i.e., the same interpreter presented with the same information would produce a different interpretation at a different time. Despite the uncertainties in the class labelling and the boundary locations, it is clear that each interpreter must have some notion, in their internal visual

database, of the texture type that characterises regions of potential Sabellaria colonisation.

Given the individuals were presented with the same image and they all recognise the Sabellaria textures, why are the interpretations different? There are many factors involved here. Unlike the clean, relatively homogeneous representation in figure 2.3 (a), in this image, the Sabellaria target is far more ambiguous. Textural surrogates are heterogeneous and spatially distributed with varying density (coverage) and lacunarity (variability in the size and shape of the distributed target elements). The image is also quite visibly corrupted with noise. Therefore, it is very difficult for the human interpreter to reliably discriminate regions of Sabellaria textures from the background. The natural boundaries of Sabellaria colonies are not well-defined, so there may be many possible interpretations of the spatial extent (Hendrick and Foster-Smith, 2006). Our eyes can discern differences in tone and texture of the grey levels between the different regions due to differing acoustic reflectivity and roughness of the materials, so naturally, physiological differences such as visual acuity will affect interpretation. There may also be institutional bias, where individuals have been trained to interpret imagery in a certain way by a particular institution or individual. The rater bias and variance is impossible to measure in absolute terms, because the true boundary of the classes is unknown. Georeferenced video footage or photographs would provide a more detailed representation of the Sabellaria colony and its delineation from surrounding sediments but these data are not available for this study. In fact, detailed, ground truth data are seldom available for evaluating seabed classification results (Martin *et al.*, 2006).

A basic assumption in seabed interpretation and classification is that if a ground truth point lies within a corresponding region of imagery with textural (or intensity) properties associated with the ground truth, the homogeneous region can be assigned the class label associated with the ground truth point. Ground truth data, in addition to being highly localised is usually very sparsely distributed due to the acquisition costs.

In the next stage of this experiment, using an un-weighted majority vote scheme, class labels are assigned to the image regions with the combined (consensus) classifications of the independent raters. This suppresses the inter-rater variability leading to a more reliable class map, with unknown accuracy (bias) in the true class boundaries. The consensus class map is shown on the far left of figure 4.5.

The results of two supervised, binary classification experiments (A), (B), are also shown in the centre and at the right of figure 4.5. In both cases the white areas represent the target class, Sabellaria (no sub-class distinction is made) and the black areas correspond to the background (sand class). The model induction and prediction parameters and sample locations for each experiment are identical but different feature creation methods are used (the methods are discussed in much greater detail later in the thesis.)

In machine classification (A), features are generated using a Gabor filter bank. Features derived from a GLCM produced classification (B). The blue vertical line in (A) and (B) is the image centre line and the parallel red lines enclose a region containing poor quality or noisy pixels either side of the nadir region. Superimposed magenta lines show the outline of the consensus boundary of the raters. The regional, machine based segmentation is a by-product of the classification.

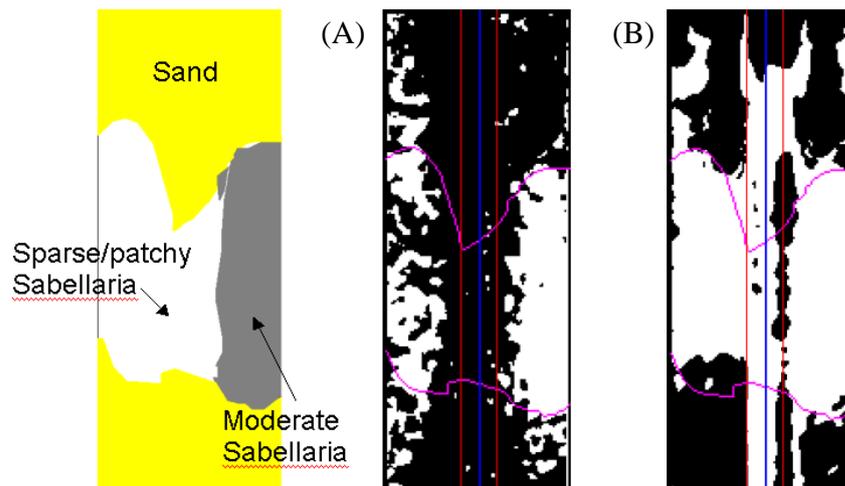


Figure 4.5. (left) Consensus classification of the image region in figure 4.3, by five independent raters. (A) machine classification using Gabor filter bank features. (B) machine classification using GLCM features.

Qualitatively, it is clear that the spatial extent of the human consensus classification and the two example machine classifications have some similarity. A quantitative assessment of similarity can be made by counting the number of sub-regions (small tiles, or image neighbourhoods), excluding the central region, where the human and machine classification are in agreement, i.e. by using a contingency table for Sabellaria and non-Sabellaria classification. If the assumption is made that the human interpretation is the benchmark or gold standard, the similarity, compared to the

benchmark can be called accuracy. In this case, classification (A) has an accuracy of 72.6 % and (B) 80.6 %. The word accuracy is inappropriate though, because these accuracies compare the proportions of a bulk area that have been assigned a class label that is in agreement with an uncertain human classification.

Further, in the absence of visual information, the numerical values of accuracy indicate that method (B) is more accurate at identifying Sabellaria than method (A). Yet, the visual result of (A), despite having an accuracy which is 10% lower than (B), is more representative of the actual physical distribution of the Sabellaria clumps. It not only captures the patchiness but shows a clear distinction between the degree of patchiness on the left and right sides of the centre line, corresponding to the descriptions of sparse and moderate concentrations of Sabellaria, respectively. For offshore engineering purposes, where contractors are concerned with avoiding bulk areas of a particular habitat, a classification similar to (B) may be useful. But for scientific purposes, where researchers are interested in quantifying the spatial distributions of biogenic structures at a particular scale, classification (A) would be more suitable. It is not then, a simple matter of claiming that one method is more accurate than another and is thus a better method. This further justifies a mixed methods approach to evaluating such results. It also highlights a point that repeatability of the classification approach and results may in fact, not be an issue for certain types of study where bespoke methods and analyses are required in an interpretative classification. It is the opinion of the author that instead of trying to replace interpretative approaches with objective statistical approaches, computational methods should be devised to assist the interpretative approach and facilitate an integration of human knowledge with objective methods, to provide a holistic view of the submarine environment.

External validation of tentative classifications in sidescan sonar imagery with ground truth is also uncertain. The very limited numbers of georeferenced locations where physical ground truth samples are taken in a survey do not usually coincide with the corresponding georeferenced points in the imagery or may correspond with image regions which are noisy or distorted. There is always a mismatch due to positioning errors in the image acquisition process and in the estimation of the ground truth location. If ground truth locations lie in the nadir region (almost directly below the instrument), the image data will be saturated, containing no discernible textural information. An example of this problem is shown in figure 4.6. The green markers

indicate ground truth locations in the nadir region of the image (the bright, vertical band of noise, running from top to bottom). Sabellaria were identified at this ground truth location. Visual inspection of the image textures reveals the possibility of Sabellaria colonisation in the region to the left, bounded by the red line, several metres from the ground truth locations. There is no discernible texture characteristic of Sabellaria near to ground truth points.

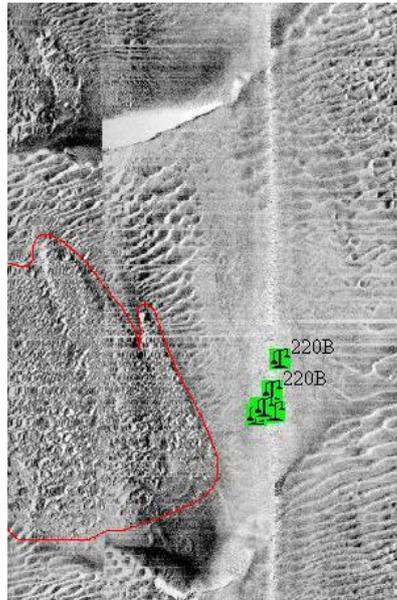


Figure 4.6. Ground truth points (green markers) are located in the nadir region of the image where there is no discernible textural information. A possible location of a Sabellaria colony is the region of texture bounded by red, to the left.

The hard evidence that Sabellaria are in the vicinity is in the ground truth samples. The ground truth sample points may (or may not) be accurately georeferenced and accurately collocated in the image. It might have been the case that small, isolated patches were sampled and they are acoustically transparent or unresolved at the 2 pixel per metre resolution in the mosaic or obscured by the nadir distortions. It is also possible that the sampling locations were actually inside the region bounded by red (or some other region) and there is a mismatch in the image location due to positioning errors of the sidescan instrument and the ground truth sample points. This is speculative though.

Regardless of the reasons why there is no discernible Sabellaria texture at the ground truth location, great care is needed when designing a machine process applied to the

discrimination task, because of such spatial mis-matches between real-world ground truth locations and corresponding target textural regions in the mosaic image. In this instance, if points near the ground truth locations are selected as target class seeds for training a supervised learning algorithm to identify Sabellaria textures, the induced model will likely be a poor representation and the prediction results erroneous. A human interpreter can recognise textures characteristic of different regions in the image, so a visual selection of representative training regions for building a machine learning model of the textures can be a more reliable approach, as in Blondel and Gómez-Sichi (2009.) Naturally, for this to work, the mosaic needs to be created and rendered at a resolution appropriate for visual identification of the textures of interest.

The approach taken in case studies 2 and 3 (chapters 6 and 7 respectively) in the thesis is to use small homogeneous regions of mosaic and waterfall imagery containing textures representative of Sabellaria and other seabed classes. The texture samples were selected by visual inspection, subject to the condition that selected regions tentatively identified as belonging to a particular class must have nearby external validation, in the form of ground truth, confirming the class type. This reduces the uncertainty in the experiments concerning discrimination of Sabellaria from other textural entities, primarily by removing the boundary and heterogeneity issues associated with mixed natural textures on larger sized image regions or full-size mosaics.

According to Anderson *et al.* (2008), many researchers prefer statistical, objective approaches to seabed classification. Current objective methods are based on established physics or empirical models of backscatter-sediment interaction, as in Fonseca and Mayer (2007), Fonseca *et al.* (2008) and Hughes Clarke *et al.* (1997). A reason why these methods are preferred by some researchers (and in some industrial data acquisition and classification pipelines) is that substrate classification results obtained from calibrated backscatter data (as in remote substrate characterisation) are standardised and repeatable (albeit uncertain.) However, seabed substrate alone does not capture the richness and diversity of the seabed environment. Remote substrate characterisation is one of representation of this environment. Sidescan offers another view, usually on a smaller scale and at a higher resolution than the MBES. It further captures details of important seabed targets such as Sabellaria colonies, a task which is unproven with current MBES technology.

The focus of the work in the thesis is specifically on novel computational methods and processes that could be used for automation of decision (human) level qualitative interpretation and classification tasks on qualitative data, such as Sabellaria discrimination. It is important to appreciate that the results and their subsequent evaluation is context specific and uncertain. It is not a limitation of the research work itself but an intrinsic property of the problems under consideration. A pragmatic research approach is appropriate for these problems since they can be specified (implicitly or explicitly) as: If the interpretation (classification) task is done by machine subject to parameters A, B, C, D, \dots, N the outcomes will be P, Q, \dots, Z . Thus reliability and repeatability (reproducibility) are always guaranteed.

Ideally, experimentation with different features, classification algorithms and processing pipelines should be performed on a common seabed texture benchmarking dataset. This point has recently been expressed by other researchers, e.g., Schumann *et al.* (2010). A common benchmarking dataset would enable the various feature creation methods and classification paradigms to be compared against each other directly and the merit of new algorithms and processes could be evaluated within a coherent measurement framework. A prescribed set of tasks for evaluating the methods and processes together with clearly defined metrics for the tasks such as classification accuracy, reliability and processing time, would facilitate a meaningful, quantitative comparison between the work of different researchers. Presently, common seabed texture benchmarking data sets and prescribed tasks and metrics have not been established. It would require collaboration of the broader research community to set up such a database and define the benchmarking scenarios, tasks and metrics.

In the next chapter, the first of the four experimental case studies is presented. Case study one concerns a feature based machine learning approach to the discrimination of pockmark objects and the mapping of their boundaries in a DBM.

A new approach to the automated mapping of pockmarks in multi-beam bathymetry

Chapter 5

Some of the work presented in this chapter resulted in the publication of a paper, “A new approach to the automated mapping of pockmarks in multi-beam bathymetry” by Harrison *et al.* (2011). The companion poster presentation is in Appendix 1.

Contents

5.1 Introduction

5.1.1 Pockmarks

5.2 Problem outline and previous work

5.3 Methods and process design

5.3.1 Overview of process and methods

5.3.2 Test data

5.3.3 Boundary definition

5.4 Feature creation

5.4.1 Terrain features

5.4.2 Texture features

5.5 Feature evaluation

5.5.1 Qualitative feature evaluation

5.5.2 Quantitative feature evaluation

5.6 Real -World object discrimination

5.7 Evaluation and discussion of results

5.8 Evaluation, conclusions, recommendations and scope for further work

5.8.1 Evaluation

5.8.2 Conclusions

5.8.3 Recommendations

5.8.4 Scope for further work

5.1 Introduction

This chapter is the first of four case studies covering the core experimental work carried out for the thesis. As has been pointed out previously, discriminating and mapping pockmarks in multibeam bathymetry is an important interpretative task. The aim of the case study reported in this chapter is to establish if it is feasible to automate the process using feature based machine learning methods. The two main objectives concern (1) the investigation of different types and combinations of features for discriminating between pockmark and non-pockmark neighbourhoods in a DBM and (2) comparing the machine discrimination of pockmark objects with a human discrimination. The remainder of the chapter is organised as follows. The short following subsection (5.1.1) briefly describes what a pockmark is. The problem and some related previous work is outlined in section 5.2. Methods and process design are covered in section 5.3. Feature creation methods are described in section 5.4. The next three sections present the main investigative components of the case study; section 5.5 elucidates the feature evaluation process, section 5.6 explains the real-world object discrimination process in a DBM. The results are evaluated and discussed in section 5.7. Finally, an overall evaluation, conclusions, recommendations and scope for further work are presented in section 5.8

5.1.1 Pockmarks

Pockmarks are naturally occurring depressions found in soft sediments at the bottoms of lakes, estuaries and oceans around the world. It is widely believed that their principal formation mechanism is fluid flow from the seabed into the water column. Their presence can be indicative of underlying hydrocarbon reserves, as in Pinet *et al.* (2008), Chand *et al.* (2009) and Judd (2001). Pockmarks support unique ecological niches as described in Webb (2009). They are designated as Special Areas of Conservation (SAC) if they contain Methane Derived Authigenic Carbonate (MDAC) structures¹, such as the Scanner Pockmark² and Braemer Pockmarks³ in the North Sea. Additionally, they are associated with geohazards, including seabed instability, trapped gases and submarine

¹ Directive 2006/105/EC, Annex 1 (1180 Submarine structures made by leaking gases) [accessed 10-06-2012]

² http://jncc.defra.gov.uk/pdf/ScannerPockmark_SelectionAssessment_4.0.pdf [accessed 10-06-2012]

³ http://jncc.defra.gov.uk/pdf/BraemarPockmarks_ConservationObjectives_AdviceonOperations_4.0.pdf [accessed 10-06-2012]

slides. Furthermore, increased fluid escape from active pockmarks may be precursory to earthquake activity (Hovland *et al.*, 2002, Lastras *et al.*, 2004).

King and MacLean (1970) first discovered these features off the coast of Nova Scotia and described them as “cone-shaped depressions” in the seabed. More recent discoveries and investigations include; the Victoria Land Basin, Antarctica (Lawver *et al.*, 2007), the St. Lawrence Estuary, Canada (Pinet *et al.*, 2008) and the Black Sea (Naudts *et al.*, 2008).

Their coverage can be extensive; Chand *et al.* (2009) report they have been observed in most of the Norwegian offshore region. On the Bering shelf, Nelson *et al.* (1979) noted areal pockmark densities of up to 1340 per km² over an area of 20 000 km². These ubiquitous objects occur in various shapes and sizes, and possess diverse morphological properties, hence, they can be classified into different generic types. Hovland *et al.* (2002) define “regular-” and “asymmetric-normal pockmark” types as circular depressions from 10 - 700 m diameter and 1 – 45 m deep (Figure 5.1).

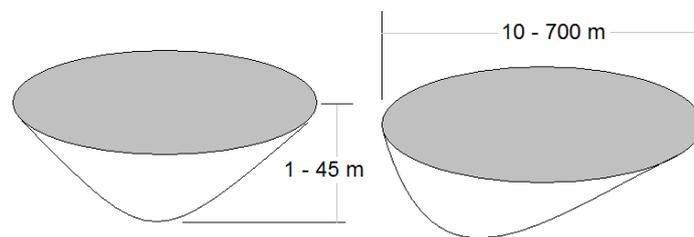


Figure 5.1. Shapes of “regular-normal” (left) and “asymmetric-normal” (right) pockmarks, based on a figure in Hovland *et al.* (2002).

If one axis is much longer than the other, the pockmark is described as “elongated.” Pockmarks, submarine sand dunes and canyons are just a few examples of a diverse range of morphological (landform) objects that can be rendered in a digital Bathymetry Model (DBM). A DBM is the end product of a processing chain, applied to acoustic sounding data, often collected by a Multi-Beam Echo Sounder (MBES) during hydrographic survey operations. It comprises accurately georeferenced depths in a raster, at fixed interval node spacing, representing the sampled terrain of the seafloor.

5.2 Problem outline and previous work

The process of identifying pockmarks in a DBM and of placing and mapping their individual boundaries is still predominantly a manual task. Manual identification and mapping of even a relatively small number of pockmarks can consume many hours of a skilled analysts' time. Human decisions on boundary size, shape and placement are likely to be subjective too and it is difficult to achieve consistency from one pockmark to another or between different interpreters working on the same data set.

Simplifying assumptions can be made about the boundaries, for instance, using a circular boundary, as in Rogers *et al.* (2006), although most pockmarks are not circular in shape. Manually placing a polygon around the pockmark in a Geographical Information System (GIS) software package is another, arbitrary approach to establishing boundary size, shape and location. However, if a detailed study is required of a dense pockmark swarm containing thousands of individuals, the task of objectively placing boundaries and mapping each individual is virtually impossible for a human to undertake. It would therefore be very useful to have an automated machine learning procedure for performing this task relatively expeditiously, objectively and consistently.

As pointed out earlier in chapters two and three, there have been relatively few investigations into the quantitative geomorphometry of DBM's and machine approaches to discriminating specific landform objects. A feature based approach to semi-automatically extracting pockmarks from MBES data was proposed by Fowler *et al.* (2008). Their feature vector comprised components such as terrain slope, curvature and fractal dimension, generated in a Geographical Information System (GIS) package.

More recently, Gafeira *et al.* (2012) proposed a process for discrimination and morphological characterisation of pockmarks in the northern North Sea region, with tools implemented using a script in ArcGIS. A watershed method is applied in a seven stage "*pockmarks identification script*" which is followed by a "*pockmark characterisation script*."

Interestingly, the planar shapes of impact craters in images of the lunar and planetary surfaces (although more circular) can appear similar to the shapes of seabed pockmarks.

Automated detection and segmentation of impact craters in radar and optical images is the subject of ongoing research. Some of the more recent approaches include, using a boosting algorithm for binary discrimination (Martins *et al.*, 2009), a three-stage workflow, comprising focusing, edge organisation and refinement processes (Kim *et al.*, 2005), and a multi-stage supervised process using mathematical morphology and shape descriptor features with a Decision Tree (DT) classifier (Urbach and Stepinski, 2009). Craters usually have a quite well-defined rim which can be used as a means of delineating the crater from the background. Pockmarks on the other hand, due to the formation process and the erosional/depositional environment may not have such clearly defined morphological boundaries.

There are several possible approaches to the pockmark discrimination problem. An image processing method could work well for boundary-based segmentation of pockmark objects. For instance, using the Circular Hough Transform (CHT) (Duda and Hart, 1971) or a zero-crossing edge detection method, such as the Laplacian of a Gaussian (Marr and Hildreth, 1980). However, the proposed approach used in this work is a region-based segmentation, achieved as a by-product of a feature based, supervised binary classification. The aim is to try to learn which neighbourhoods (centred on grid nodes) of the DBM belong to a pockmark (P) object (or not, $\neg P$) and use this as a means of object discrimination.

5.3 Methods and process design

5.3.1 Overview of process and methods

The process comprises seven stages, as illustrated in figure 5.2.

- 1. Feature creation:** A feature vector is generated, as described in the next section. The full feature vector (FFV) contains all features at multiple resolutions if feature evaluation and pre-selection has not been performed.
- 2. Feature selection:** (Optional) feature saliency is evaluated in a filter and wrapper, applied to control data. Bhattacharyya's distance measure is used to rank features in the filter approach, whilst exhaustive feature subset evaluation

and feature clamping can be carried out within the wrapper. Results are used to select an appropriate feature sub-set for the task.

- 3. Classifier model induction and class prediction:** A Ball Vector Machine (BVM) classifier is trained with user-selected samples. The classifier then predicts the labels of all grid nodes in the raster as either pockmark (P) or non-pockmark ($\neg P$) classes.
- 4. Segmentation of the landform objects:** Classifier output (label predictions) are transformed into a raw (unprocessed) binary image, referenced to the local coordinates of the DBM grid.
- 5. Filtering:** A morphological closure operation is applied, to remove false positives and false negatives from the binary image.
- 6. Boundary extraction:** A kernel detects the boundaries between the two classes in the filtered image.
- 7. Mapping:** Boundary locations are super-positioned on the original DBM raster and the end result is displayed to the user.

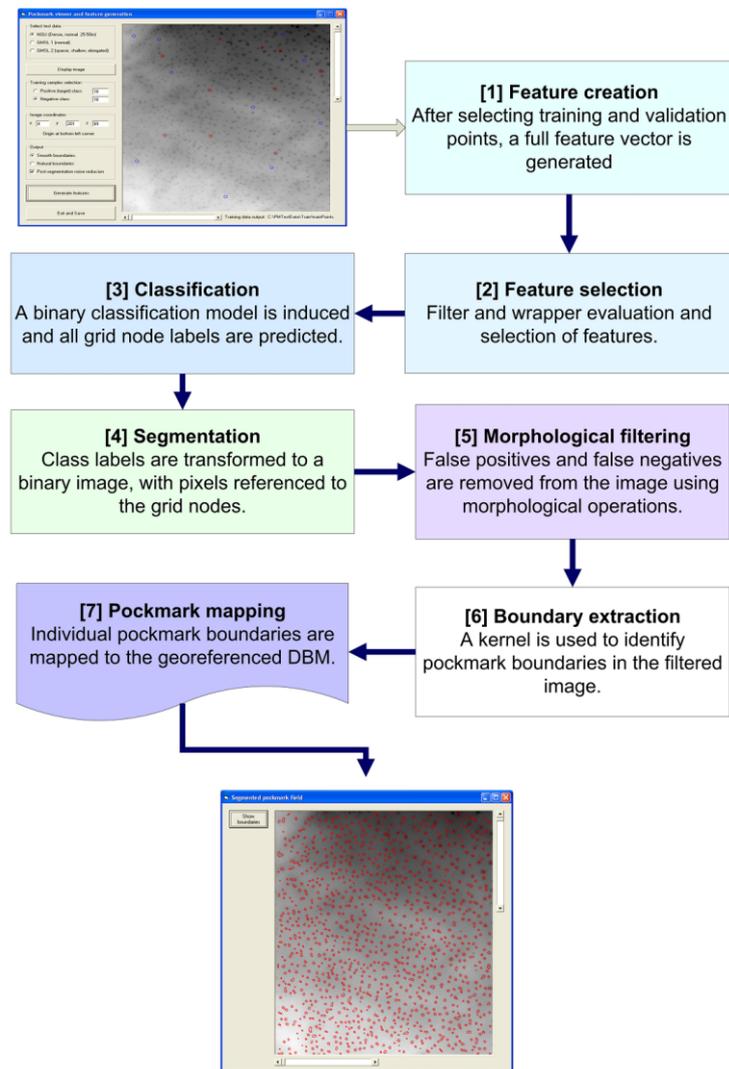


Figure 5.2. Sequential stages in the pockmark discrimination and mapping process

5.3.2 Test data

A rectangular sector of rasterised (5 metre node spacing) bathymetric data 7.5 km x 2.8 km was provided by the Geological Survey of Norway (NGU.) These data are a small subset of a 3500 km² survey conducted during 2005-2006, in the Barents Sea. A Kongsberg EM 1002 MBES was used to collect the bathymetric data. See Chand *et al.* (2009) for further details of the surveyed area and data acquisition.

5.3.3 Boundary definition

There is no consensus of opinion on what, precisely, constitutes a morphological boundary or boundary region between a pockmark and the surrounding seabed. In fact, boundary delineation (specifically its ambiguity) is inherent in many landform object discrimination problems, e.g., Fisher *et al.* (2004).

In this work, a pockmark boundary is defined to exist where a P class grid node is next to a $-P$ node. Thus, object segmentation and boundary delineation can be derived from binary classification of the nodes, as a pockmark object is a localised spatial cluster of P nodes. The aim is to learn which neighbourhoods could be part of a pockmark (or not), rather than define the whole pockmark object in terms of its intrinsic morphological properties. The DBM surface (unlike the real, continuous object surface) is a sample of points from the topography soundings. As the underlying data representation of the surface is discrete and a hard classification technique is applied, this hard boundary definition will be the most practical. The basic concept is illustrated in figure 5.3. Figure 5.3 (a) shows the rendered DBM surface and 5.3 (b) a specimen classification of P (blue) and $-P$ (red) nodal points. In this figure, targets are defined by the collection of blue nodes and their boundary elements defined to occur where red and blue nodes are next to each other.

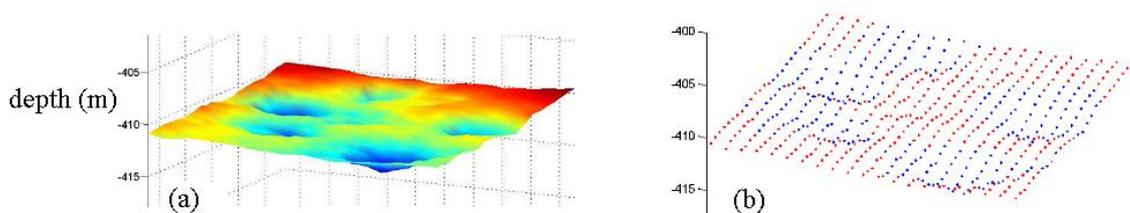


Figure 5.3 (a) example of a rendered section of DBM (b) corresponding specimen binary classification of node points in the DBM section, showing how targets (collections of blue-coloured nodes) and their boundaries are represented, in accordance with the definition in the text.

5.4 Feature creation

It is important to represent neighbourhoods forming part of a pockmark, (P), and non-pockmark, ($\neg P$) with discriminatory patterns so that an efficient (low dimensional) and accurate machine model can be induced with a small number of pattern instances. A fundamental problem to solve then, is to establish if the pattern created by the features, on a particular nodal neighbourhood in the DBM identifies that node as belonging to a pockmark object or not. It is a binary or two-class problem, with the P nodes as the target class. The features need to be able to capture the context of the node, relative to its surroundings and hence, collectively, capture the geometric signature of the landform objects so they can be partitioned from the surroundings, using the boundary definition in 5.3.3.

Resolution of the features is also important. If the computational kernel is set at an inappropriate resolution, the patterns will not be as discriminatory. A much larger scale (lower resolution) feature kernel would be useful for discriminating regions of the DBM covered by pockmarks, facilitating the delineation of a pockmark field. Such a large scale would be unsuitable though, for distinguishing individual landform objects, which is the goal of this work. In a visual inspection of the rendered DBM, the pockmark diameters ranged from approximately 20 m - 50 m, this is corroborated by Chand *et al.* (2009) who observed diameters generally less than 50 m. The node spacing in the raster is 5 m, indicating that computational kernel scales of 5, 7, 9 and 11 nodes (25 m - 55 m) should be appropriate, considering the size ranges of these pockmarks. Feature generators chosen for the task are subdivided into two generic groups comprising "terrain" and "textural" features as described in the following subsections.

5.4.1 Terrain features

[1] Prototype crater matching kernel (PCM)

A novel, prototype crater matching kernel is defined, based on the description of pockmarks as "concave, crater-like depressions" King and MacLean (1970).

Moving outwards from the lowest (assumed central) point inside a normal pockmark, whilst remaining inside the pockmark boundary, it is reasonable to suppose, for the purposes of modeling, that the altitude of the location increases monotonically with distance from the centre. This follows directly from the notion of the pockmark as “concave” and “crater-like.” A prototype kernel for evaluating if a local spatial arrangement of DBM nodes conforms to this notion is proposed. By comparing the relative vertical positions of nodes in directional sequences (defined by the unit vectors below) on the DBM, with a formal mathematical description of the “crater like” prototype, the degree of similarity between the local DBM structure and the prototype can be assessed. Figure 5.4 illustrates a kernel spanning seven grid nodes of a hypothetical DBM neighbourhood, with the central node under test. The kernel is non-parametric (aside from its size) and contains no pre-assigned operator values. The objective is to assign a numerical weight to the node under test to assess its similarity to the notion of a crater-like object within the context of the spatial neighbourhood of nodes.

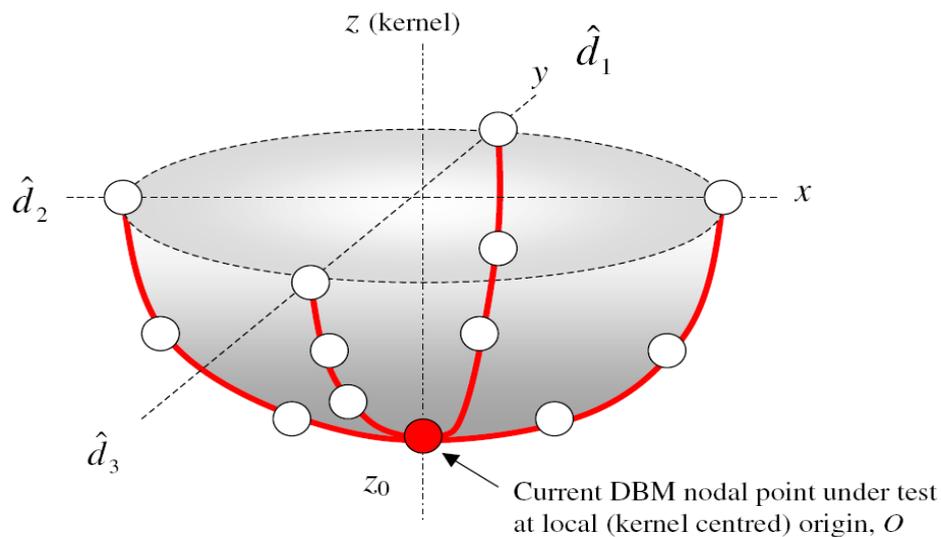


Figure 5.4. PCM kernel, spanning seven grid nodes of a hypothetical DBM neighbourhood, with the central node under test at the local (kernel centred) origin.

Four unit vectors, \hat{d}_n , correspond to four orthogonal directions, within the computational kernel, relative to the centre of the kernel at the current location in the DBM, $O(x, y)$.

Moving in an anticlockwise direction starting in the direction of the positive x -axis ($+x$):

$$\begin{aligned}
 \hat{d}_0 &\rightarrow (1, 0) \\
 \hat{d}_1 &\uparrow (0, 1) \\
 \hat{d}_2 &\leftarrow (-1, 0) \\
 \hat{d}_3 &\downarrow (0, -1)
 \end{aligned}
 \tag{5.1}$$

The kernel size, $k \in \{5, 7, 9, 11\}$ (grid nodes) used to generate a sub image at a particular resolution, places a bound on the discrete number of nodes in the sequence, from the local origin O , in the directions \hat{d}_n . The current position in the node chain, relative to the origin is p , where,

$$\left\{ 1 \leq p \leq \frac{k-1}{2} \right\}$$

The altitude, z of a node at p , in a direction \hat{d}_n is therefore, $z(p\hat{d}_n)$. At the local origin of the kernel, $z_0 = z(0\hat{d}_n)$. A discrete, unit Heaviside function, $H[f(p, \hat{d}_n)]$ can now be defined to compare the relative, sequential vertical displacements of the nodes in the directional sequences,

$$H[f(p, \hat{d}_n)] = \begin{cases} 0, & z(p\hat{d}_n) - z((p-1)\hat{d}_n) < 0 \\ 1, & z(p\hat{d}_n) - z((p-1)\hat{d}_n) \geq 0 \end{cases}
 \tag{5.2}$$

Finally, a normalised weighting function, $W_k: \{0 \leq W_k \leq 1\}$ is defined, which is applied at each grid node tested, to generate the feature value at a specific resolution, k . In compliance with the definition of the relative spatial arrangement of nodes in a ‘‘crater-like’’ context, $\max\{W_k\} = 1$ (a localised node structure that is identical to the definition of being ‘‘crater-like’’ as in fig. 5.4) and $\min\{W_k\} = 0$ (a localised node structure that is least like the definition of being ‘‘crater-like’’),

$$W_k = \frac{1}{2(k-1)} \sum_{n=0}^3 \sum_{p=1}^{\eta} H[f(p, \hat{d}_n)], \quad \text{where } \eta = (k-1)/2
 \tag{5.3}$$

There are some important advantages of this novel prototype crater-matching feature;

- There is no need to fit a local surface to the nodes or compute curvatures or gradients.
- Nodal values in the DBM can be used directly, no smoothing is required.
- It is not necessary to compare numerical kernel prototype values with the DBM nodal values to establish the degree of correlation, as is the case with many correlation based pattern-matching templates.
- The approach can easily be modified to fit other structural variants such as “elongated pockmark,” or “ridge.”
- The spatial coherency of the real, individual morphological features is captured in the set of multi-resolution sub-images. This could potentially be useful for high-frequency noise rejection and for automated determination of pockmark size.

In summary, the PCM feature should provide a strong indication of the context of the central pockmark nodes, as will be demonstrated later. It is computationally less expensive than existing template approaches, since relative spatial relationships are quantified, as opposed to computing absolute numerical measures such as gradient, curvature and shape template correlation.

[2] Deviation from a local mean plane (DMP)

This measure is similar to the Bathymetric Position Index (BPI) (Weiss, 2001). The BPI measures the deviation of a central node, from the mean altitude of nodes within a circular patch of surrounding nodes. Using the (GIS) notation of Wilson *et al.* (2007),

$$BPI = Z_{\text{grid}} - \text{focalmean}(Z_{\text{grid}}; \text{circle}, r) \quad (5.4)$$

where, Z_{grid} is the current node at which BPI is being calculated. The focal mean provides the mean value of the DBM nodes in a circle, radius r , centred on Z_{grid} . In the application used here, the deviation of the central node from the mean of a square kernel of dimension, 5, 7, 9 and 11 nodes is computed after the local nodal values under the kernel are first transformed to a zero-mean normal distribution, i.e $N \sim (0, 1)$. After computing the residual at each of the nodal positions, the four feature spaces are globally rescaled to $[-1,1]$. Any node that is low, relative to its surroundings (such as a node near the centre of a pockmark) would take a negative value. Relatively high nodes have positive values and any node at the same level as the local mean (zero) plane has a deviation of zero.

[3] Laplacian of a Gaussian (LGA)

The Laplacian of a Gaussian was first defined by Marr and Hildreth (1980). It can be used as an edge detector (Gonzalez and Woods, 2008). In the application here, there is no need to calculate the zero crossings to find edges explicitly (as would be the case in a boundary based segmentation), since it is the response of the function to a pockmark as a distinguishing feature that is of interest. The function is obtained by applying the Laplacian operator ∇^2 to the isotropic Gaussian convolution mask $G(x, y)$ giving the expression,

$$\nabla^2 G(x, y) = \frac{(x^2 + y^2 - 2\sigma^2)}{\sigma^4} \exp\left\{-\frac{(x^2 + y^2)}{2\sigma^2}\right\} \quad (5.5)$$

where, x and y are displacements from the kernel centre node and σ is the standard deviation of the Gaussian function, which is varied in proportion to the resolution of the computational kernel. The feature sub-images, $z(x, y)$ are generated at the four specified kernel resolutions by convolution of a normalised Laplacian of a Gaussian with the DBM, $Z(x, y)$,

$$z(x, y) = \frac{1}{N} \nabla^2 G(x, y) * Z(x, y) \quad (5.6)$$

where the resolution dependent normalisation coefficient, $N(k)$, is defined as,

$$N(k) = \sum_{i=-\kappa}^{+\kappa} \sum_{j=-\kappa}^{+\kappa} \nabla^2 G(x, y), \quad \kappa = \frac{k-1}{2} \quad (5.7)$$

5.4.2 Texture features

Textural features derived from the co-occurrence matrix, a.k.a the Grey Level Co-occurrence Matrix (GLCM) of Haralick *et al.* (1973) have been applied to numerous problems in, for example, remote sensing (Clausi, 2002), terrestrial terrain analysis (Wood, 1996) and classification of multi-beam backscatter imagery (Blondel and Gómez Sichi, 2009). However, it is not known if an individual pockmark can be discriminated from its immediate surroundings by considering local co-occurrence based measures of textural patterns in the terrain.

Further technical details of the GLCM and a method for evaluation of the parameters and features are given in chapter 7. A brief outline is given here. In order to capture the texture of individual pockmarks it is necessary to use a kernel size comparable to (or smaller than) the size of the feature of interest. Kernel sizes of 5 and 7 nodes, corresponding to a morphological feature size of 25m-35m are used. The inter-pixel distance $\|\mathbf{d}\|$, is fixed at $\|\mathbf{d}\| = 1$, and nodal altitudes are quantised locally, under the kernel, to 3 bits ($Q = 3$), i.e. 8 levels in all experiments. This is a necessary step, as DBM rasters typically have a vertical resolution of 1 cm. In a data set like this one with a vertical range of a few tens of metres, there are potentially thousands of discrete levels. It would be extremely inefficient computationally, to use a GLCM at the full, vertical resolution due to the size and sparsity of the matrix.

Directions of the displacement (sampling) vector used in this study are $\left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{2}\right\}$.

P_θ is a normalised co-occurrence matrix at orientation θ and $f_k\{P_\theta\}$ a feature derived from the co-occurrence matrix at orientation θ for the kernel size k . A rotationally invariant feature, F_k^{RI} , can be defined by taking the mean over the four orientations,

$$F_k^{RI} = \frac{1}{4} \sum_{\theta=0}^{\frac{3\pi}{4}} f_k\{P_\theta\} \quad (5.8)$$

An advantage of using a rotationally invariant GLCM is that textural anisotropies and the effects any other orientation dependent factors are averaged to produce a mean representation of the texture. Using the mean of the feature values in each direction also reduces the dimensionality of the feature space by a factor of 4, compared to generating the features at each orientation. Three features are derived from the co-occurrence matrix; contrast (CON), correlation (COR) and entropy (ENT) i.e., $F_k^{RI} = \{\text{CON}, \text{COR}, \text{ENT}\}$. Clausi (2002) also used CON, COR and ENT as a preferable choice of features for classifying Synthetic Aperture Radar (SAR) sea-ice imagery. Further, through experimentation, Clausi (2002) found that the peak classification accuracy using these features occurred at coarse quantisation levels (hence lower computational complexity and the choice of $Q = 8$.) The three features are defined as follows;

Contrast (CON) - a measure of differences in the quantised values between neighbouring nodes.

$$\text{CON} = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} (i - j)^2 P_\theta \quad (5.9)$$

Correlation (COR) - a measure of how correlated the quantised value of a node is to that of its neighbours.

$$\text{COR} = \frac{\sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} ijP_\theta - \mu_i\mu_j}{\sigma_i\sigma_j} \quad (5.10)$$

Entropy (ENT) - a measure of the randomness of the quantised nodal values.

$$\text{ENT} = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} P_d \ln(P_\theta) \quad (5.11)$$

Where μ_i , μ_j and σ_i , σ_j are the mean and standard deviations of the (normalised) co-occurrence distribution, given by,

$$\begin{aligned} \mu_i &= \sum_{i=0}^{Q-1} i \sum_{j=0}^{Q-1} P_\theta & \sigma_i &= \sum_{i=0}^{Q-1} (i - \mu_i)^2 \sum_{j=0}^{Q-1} P_\theta \\ \mu_j &= \sum_{j=0}^{Q-1} j \sum_{i=0}^{Q-1} P_\theta & \sigma_j &= \sum_{j=0}^{Q-1} (j - \mu_j)^2 \sum_{i=0}^{Q-1} P_\theta \end{aligned} \quad (5.12)$$

All features are summarised in table 5.1. The legend applies to all subsequent figures.

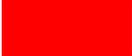
		Kernel resolution, k (grid nodes)				
Feature type	Feature	5	7	9	11	Legend
Terrain	Prototype crater matching	PCM(5)	PCM(7)	PCM(9)	PCM(11)	
	Deviation from mean plane	DMP(5)	DMP(7)	DMP(9)	DMP(11)	
	Laplacian of a Gaussian	LGA(5)	LGA(7)	LGA(9)	LGA(11)	
Texture	Contrast	CON(5)	CON(7)	-	-	
	Entropy	ENT(5)	ENT(7)	-	-	
	Correlation	COR(5)	COR(7)	-	-	

Table 5.1. Summary of the features and resolutions used in this study. Texture features are not applied in this test case at resolutions of 9 and 11 nodes.

5.5 Feature evaluation

5.5.1 Qualitative feature evaluation

This qualitative evaluation provides a preliminary insight into the discriminatory capacity of the individual features. Responses of an individual feature to a grid node in the context of a pockmark and non-pockmark terrain patch are investigated. Differences in the magnitude of the feature responses in these two contexts may be indicative of the capacity of the feature to discriminate between P and $-P$ grid node classes in general. Qualitative evaluation is not an integral part of the processing.

[1] Qualitative evaluation of features at a typical P -node

A normal, asymmetric pockmark (approx. width = 35m, length = 45m, depth = 2m) is used to investigate the behaviour of the features, applied to a line of nodes through a single interior control node, as shown in figure 5.5.

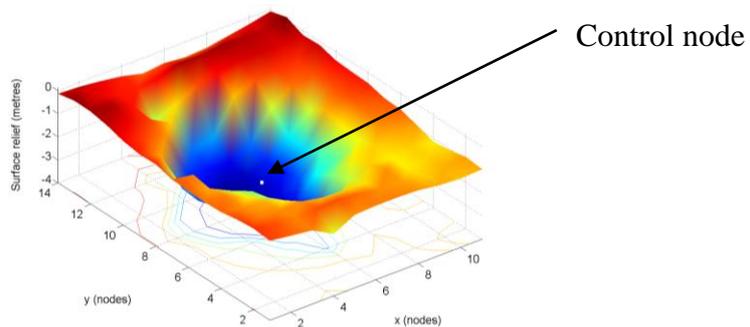


Figure 5.5 A typical asymmetric normal pockmark selected arbitrarily from the DBM: width = 35m, length = 45m, depth = 2m.

The normalised numerical responses of the features generated on neighbourhoods centred along a line of 11 sequential grid nodes, through the control node, parallel to the x -axis are shown by the bar charts in figure 5.6. They are grouped row-wise according to the feature type and column-wise according to the feature kernel resolution, k . The first row (kernel sizes 5,7,9,11, from left-right) shows the response of the PCM kernel. At all four resolutions, the feature value peaks, as the kernel moves through the pockmark. It is clear that the response of the feature is correlated at the different

resolutions. Row 2 shows the response of DMP, with the zero crossing points approximating the boundary between the pockmark and the surroundings. A similar shaped response, but with a greater amplitude, is produced by LGA in row 3. The similarity between the responses of LGA and DMP is a possible indicator of redundancy, and their magnitudes may also be correlated to the response of PCM. The two charts in row 4 show the variation of ENT at length scales of 5 and 7 nodes, respectively. CON values are shown by the charts, in row 5, again at 5 and 7-node scales. The first two charts of the final row show COR at resolutions of 5 and 7 nodes. A small contour plot of the bathymetry and the terrain profile through the control node, are also shown on the right of the final row. Whilst the 3 features from the terrain group in rows 1-3 show distinctive responses to the pockmark, it is not clear from this initial inspection if the textural features will prove to be useful discriminators.

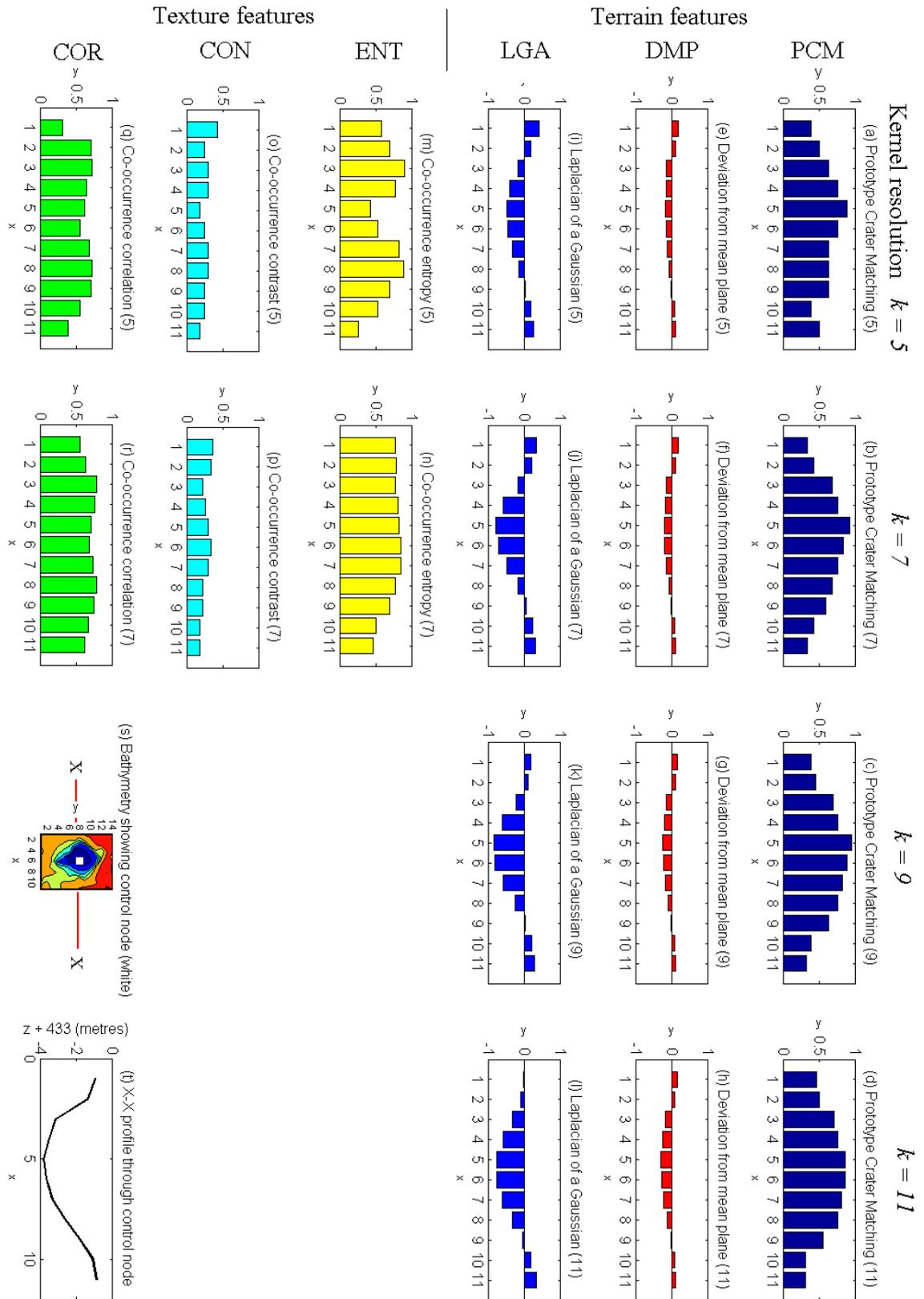


Figure 5.6 Bar charts showing the normalised values (y-axes) of the terrain and texture features at node points (1-11) (x-axes) along a line, x-x, through the control node, inside the pockmark, parallel to the x-axis. The bar charts are colour coded according to the legend in table 5.1. A contour plot (bottom row, third right) represents the bathymetric surface and the location of the control node (white square marker.) The vertical profile of the DBM nodes along x-x is shown in the plot at the bottom right corner.

The numerical responses of all 18 features at the control node inside the pockmark, i.e., the full-feature vector (FFV) at the control node are summarised by the bar chart in figure 5.7. The first 4 bars show the response of PCM. Bars 1-3 increase in height as the kernel size increases, indicating PCM is responding more strongly. There is a slight drop in the amplitude of the feature as shown by bar 4, at a scale of 55 m, which is at least 10m greater than the longest dimension of the pockmark. The peak response at bar 3 (9 nodes or 45 m) corresponds to the maximum dimension of the pockmark, of 45 m. Response patterns of DMP and LGA in bars 5-8 and 9-12 respectively produce values which increase monotonically with kernel size. ENT values are indicated by bars 13-14, CON, bars 15-16 and COR, bars 17-18.

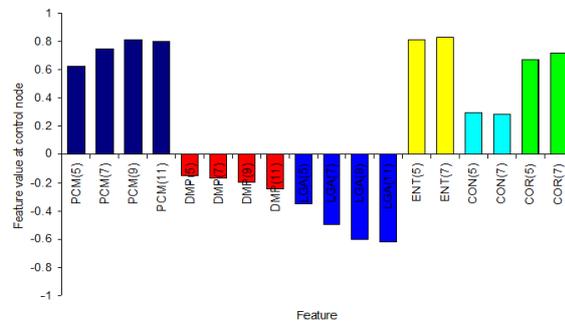


Figure 5.7 Bar chart showing the numerical responses of all the individual features in figure 5.6 at the control node inside of the pockmark. The bars are colour coded (legend in table 5.1.)

[2] Qualitative evaluation of features at a typical $\rightarrow P$ node

A control node is selected, clear from any pockmark boundaries, in a square patch of terrain. To avoid repeating previous details, the numerical performance of the features at the control node is summarised in figure 5.8.

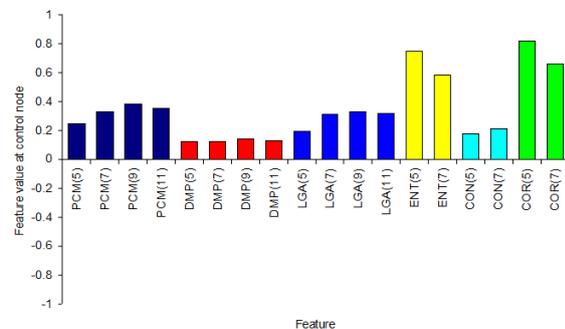


Figure 5.8. Bar chart showing the numerical values of the FFV at a $\rightarrow P$ control node.

There are clear differences between the feature profile at the $\neg P$ node in figure 5.8 and the P node in figure 5.7. The amplitude of the PCM feature is much lower at the $\neg P$ node. The sign of DMP and LGA is reversed at the $\neg P$ node. The remaining 6 bars indicate the responses of the 3 textural features at the two resolutions. ENT and COR (bars 13-14 and 17-18, respectively) show a slight increase in amplitude between the 5 and 7-node scale at the P node. The amplitude decreases slightly, in moving from the 5 to 7-node resolution at the $\neg P$ node. CON values (bars 15-16) have greater amplitudes inside of the pockmark.

By applying a feature selection process to a larger sample of P and $\neg P$ control nodes, in the next section it is shown that these differences are not coincidental and they can be used to discriminate between P and $\neg P$ classes.

5.5.2 Quantitative feature evaluation

The goal of this process is to identify the least number of features capable of discriminating between the nodes, with the highest accuracy. Preselecting features reduces the dimensionality of the feature vector. This expedites the processing and generally improves accuracy. Two independent methodological approaches to feature evaluation are applied;

- Filter approach: Features are evaluated individually and independently of the classification algorithm, using a distance measure to assess their discriminatory potential. The features are ranked in decreasing order of their intrinsic capacity to separate the classes.
- Wrapper approach: A classification model is induced and the features are evaluated on the basis of their performance (sensitivity, specificity and accuracy) with the specific classifier. Within the wrapper, features are investigated using a feature clamping approach and by performing exhaustive evaluation of feature sub-set combinations. Further details of these methods can be found in, for instance, Guyon and Elisseeff (2003), Tan *et al.* (2006) and Webb (1999).

[1] Filter approach

Bhattacharyya's distance measure (Bhattacharyya, 1943) is used to compute the distance between the estimated probability density function (pdf) of the feature values for P and $\neg P$ class instances, sampled from the control data. Denoting the probability distribution estimates for P and $\neg P$ class instances of an individual feature x_f , by $P(x_f)$ and $\neg P(x_f)$, respectively, the Bhattacharyya distance measure is,

$$B(P(x_f), \neg P(x_f)) = \arccos \left\{ \sum_{i=1}^n \sqrt{P(x_f) \cdot \neg P(x_f)} \right\} \quad (5.13)$$

Where, $B \in [0, \frac{\pi}{2}]$, the maximum and minimum values corresponding to disjoint and congruent, coincident distributions, respectively (i.e. maximum and minimum class separation.) The control data used in the feature evaluation and selection stages comprise approximately 1000 manually (visually) selected and labeled DBM nodes in each class, i.e. the selected control nodes were labeled as P or $\neg P$. Care was taken to ensure the P points were selected from inside independent pockmarks away from the boundaries and $\neg P$ control points from the surrounding terrain, away from the boundaries of nearby pockmarks. Computing the distance measure for each individual feature facilitates ranking of the features in terms of their class discriminatory capacity, as shown in figure 5.9.

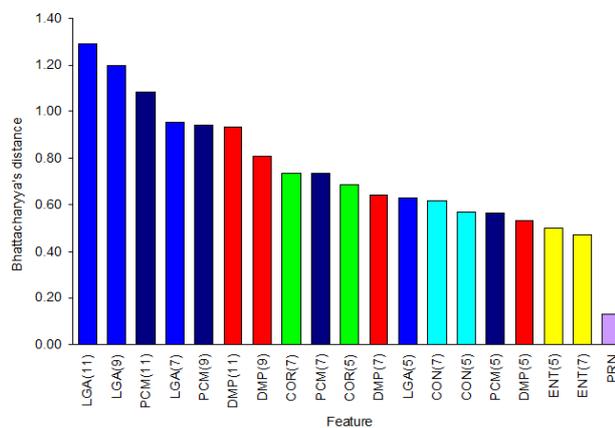


Figure 5.9 Bar chart showing a ranking of the individual features, computed using Bhattacharyya's distance measure.

The most salient individual feature is LGA(11). It has the greatest intrinsic capacity to discriminate between the P and $\neg P$ classes. The pseudo random noise feature (PRN) is the least effective (and naturally, we should expect all discriminatory features to perform better than random noise.) The left-half of the ranking chart is populated mostly by the terrain features at larger kernel sizes, confirming an appropriate choice of resolutions for the scale of the objects. Terrain features with the smaller kernel sizes and the textural features are grouped more to the right of the chart indicating a relatively low discriminatory capacity as individual features. Any one of the top 3 features, LGA(11), LGA(9), PCM(11) may be suitable for inclusion in a feature vector.

[2] Wrapper approach

A Ball Vector Machine (BVM)⁴ classifier is used as the core learning algorithm for the dual tasks of feature evaluation in the wrapper and classifying grid nodes in the real world application. It is a fast Support Vector Machine (SVM) that uses a core set approximation on very large data sets (Tsang *et al.*, 2007). The binary classification function, $\Omega(x)$, for determining class membership of an unseen pattern, x is, $\Omega(x) = \text{sgn}(w \cdot \Phi(x) - \beta)$ where w is the normal vector to a discriminating hyperplane and β is the distance of the plane from the origin. A default radial basis function (RBF) kernel, $\Phi(x)$, is used. Default settings are also used for model cross-validation, training cost, C , and the hyper-parameter for the RBF kernel function, γ . A sound practical reason for choosing the BVM as the binary, hard discriminative classifier Practical reasons for choosing the BVM are that it is as accurate as an SVM but faster, and can deal with larger data sets (Tsang *et al.*, 2005.)

(a) Feature clamping technique

In a feature clamping technique, one feature at a time can be removed from the n -dimensional FFV and then replaced. The classifier is trained and tested on each combination (containing $n-1$ features) replacing the feature before taking out the next one, until all 18 combinations have been tested in this way. Examining the sensitivity, specificity and overall accuracy of the classifier output highlights any features whose

⁴ LibCVM Toolkit Version: 2.2 (beta), How to use. <http://www.cs.ust.hk/ivor/cvm.html> [Accessed 23-07-2010]

removal from the FFV produces an increase (or reduction) in class discrimination performance. The individual features are then ranked according to the change in overall accuracy brought about by their removal from the FFV. Features whose removal causes the greatest reduction in accuracy are ranked the highest. The clamping experiment was repeated 10 times, using different numbers of samples and ratios of P and $\neg P$ control nodes. The ranking of the features is shown in figure 5.10.

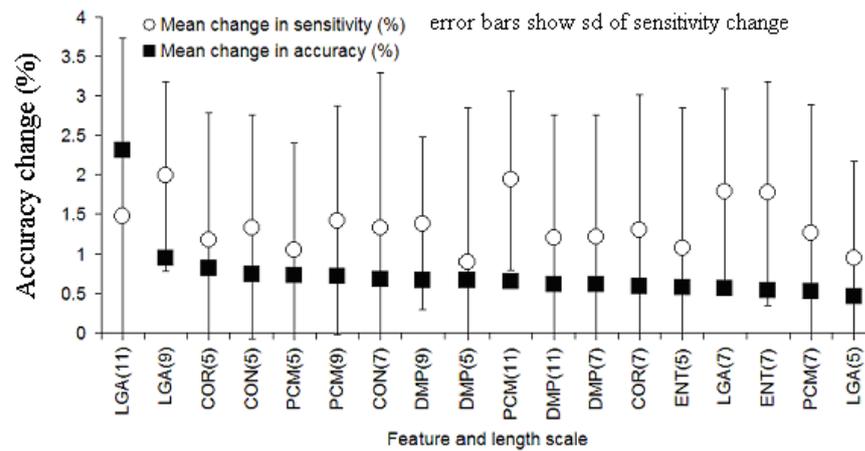


Figure 5. 10 Features ranked, in order of decreasing influence on the FFV, from left-right. Mean change in accuracy is used as an objective function for the rankings. Mean change in sensitivity is also shown and the error bars indicate the standard deviation of sensitivity change over multiple runs..

The most salient feature is LGA(11), followed by LGA(9) and COR(5). The overall mean accuracy of the FFV for discriminating between P and $\neg P$ control nodes in all test cases is 93.3 %. Mean overall sensitivity is 93.4%. Sensitivity changes are shown on figure 5.10, the error bars represent 1 standard deviation from the mean sensitivity change. LGA(11), brought about a 2.31 % mean reduction in accuracy, when it was removed from the FFV indicating that it must have a relatively important contribution to make in relation to the other features.

The filter investigation and the feature clamping experiments provide evidence that LGA(11) is probably the most useful individual feature of the group. The feature clamping results indicate is it important, synergistically to the FFV which could in turn, be connected to its overall importance in relation to other feature combinations. Removal of any of the other features individually from the FFV reduced the accuracy by less than 1% and the differences in these accuracy changes, between features other than LGA(11) is very small, less than 0.5% between LGA(9) and LGA(5), ranked second

and bottom, respectively. Considering the changes in sensitivity in figure 5.10, removal of the individual features LGA(9), PCM(11), LGA(11) or ENT(7) reduces the sensitivity by between 1.5 - 2.0 %. It is possible that a combination of 2 features, such as LGA(11) or PCM(11) with ENT(7) (or any other textural feature) would be suitable for the classification task. The reason for this is that the terrain features are not strongly correlated to the textural features. For instance, it may at first seem logical to construct the feature space from say the top-3 ranked features in figure 5.9. However, they are highly correlated (which was easily verified using a correlation matrix), since the top two features are the same feature applied at different scales of analysis. In other words, they may contain redundant information that will not be useful to the classifier.

(b) Exhaustive evaluation of feature subsets

Several further experiments were carried out to exhaustively evaluate all the feature combinations in pre-selected feature subsets. The reason for doing this is to identify specific feature combinations that have a good classification performance. Using information from the filter, feature clamping and exhaustive evaluations, a reliable decision can be reached, concerning which feature combinations would be suitable for maximising the accuracy and reliability of the classification task. Additionally, another objective is to minimise the computational complexity by reducing the dimensionality of the feature space. The overall goal is to find the combination(s) of features with the smallest dimension producing the highest classification accuracy, subject to a threshold of minimum sensitivity, specificity and accuracy of 80 %. As an illustrative example, all possible combinations of the six different feature groups at a fixed resolution of seven nodes (35m) were evaluated on the control data. The feature combinations are encoded as binary strings, with "1" indicating the presence of the feature in a combination and "0" its absence. Ranking of the features was based on using overall accuracy as the objective function, subject to the additional constraint that accuracy, sensitivity and specificity must all exceed the 80 % threshold. Only seven feature combinations out of the 64 evaluated in this experiment satisfied the criteria. They are listed, with numerical results, in table 5.2. A smaller difference between sensitivity and specificity is also preferable over a relatively large difference, so this condition can be applied as a further constraint if desired, in selecting a good feature subset for the classification task.

Combination code	sensitivity %	specificity %	accuracy %	rank	dimension
101011	91.8	85.1	89.0	1	4
001111	93.7	82.4	89.0	1	4
111100	90.7	85.8	88.7	2	4
111111	87.3	89.7	88.3	3	6
101110	93.7	80.3	88.1	4	4
111010	88.4	86.9	87.8	5	4
111110	85.6	89.4	87.2	6	5

FFV code	1	1	1	1	1	1
Attribute	PCM	DMP	LGA	CON	COR	ENT
Example	1	0	1	0	1	1
Combination	PCM	-	LGA	-	COR	ENT

Table 5.2 Top seven feature subsets at a resolution of 7 nodes (35 m). The feature codes key and an example code is also provided.

All these combinations include at least one feature from both the terrain and textural feature groups. Of the two joint top-ranked feature combinations {LGA(7), CON(7), COR(7), ENT(7)} has slightly higher sensitivity and is therefore marginally better at identifying the P nodes, compared to {PCM(7), LGA(7), COR(7), ENT(7)}. The former also has a smaller difference (6.7 %) between sensitivity and specificity.

Considering the correlation between the generated feature distributions on the two classes (at a kernel size of 7 nodes), the linear similarity in the distributions of values in the control data, for each feature can be quantified with Pearson's correlation coefficient. The results are shown as a matrix, in figure 5.11. The upper-right triangular part of the matrix shows the degree of correlation for the P class and the lower-left triangular part shows the degree of correlation for the $\neg P$ class. The features are grouped as terrain and texture types, using the same codes as before.

$k = 7$	Morphological features			Textural features			PRN
	PCM	DMP	LGA	CON	COR	ENT	
PCM		-0.75	-0.69	0.31	0.06	0.47	0.02
DMP	-0.79		0.86	-0.41	0.19	-0.48	0.01
LGA	-0.60	0.71		-0.48	0.21	-0.55	0.00
CON	0.03	0.01	-0.02		-0.66	0.82	-0.03
COR	0.00	0.01	-0.02	-0.65		-0.17	0.03
ENT	0.04	0.01	-0.07	0.77	-0.06		-0.02
PRN	-0.02	0.03	0.01	0.05	-0.05	0.04	

Figure 5.11. Correlation between different feature distributions at a fixed kernel size of $k = 7$. The upper-right triangle shows the correlation coefficients for the pockmark class (P) distributions and the lower-left shows the correlation coefficients for the non-pockmark ($\neg P$) class.

Some very interesting information is shown here. The region shaded in green shows strong correlations between the different features in the terrain group generated on the control data for pockmark and non-pockmark classes at the same scale (a strong correlation was also observed between the same feature generated at different scales.) The pink shaded region also shows strong correlations, i.e., COR(7):CON(7) and ENT(7):CON(7), between the textural features for the pockmark and non-pockmark control data. Correlation between the terrain features and the textural features (yellow shaded region) is negligible for the non-pockmark class, with correlation coefficients of comparable magnitude to the correlation with the PRN. However, the blue shaded region shows a relatively greater correlation between the textural features and the terrain features for the P class. This may indicate a natural correlation between the real texture and morphology of interior of the pockmarks (P) relative to the surrounding terrain of the seabed ($\neg P$).

Similar experiments were carried out using pre-selected subsets containing up to 12 features, at multiple resolutions. Combinations including LGA(11) or PCM(11) paired with one of the textural features were found to produce very good classification results on the control data. For instance, with the combination {LGA(11), CON(7)} sensitivity, specificity and accuracy of 96.9%, 92.0% and 94.8% respectively, were attained. It is possible to improve on these values slightly but a case of diminishing returns. Gaining ~1% in accuracy was typically achievable with a 4-5 fold increase in the dimension of the feature space. The overall best accuracy result was 95.9 % with a combination of 10 features. In terms of achieving the goal of accuracy and low dimensionality though, a two-dimensional combination of LGA(11) or PCM(11) with CON(7), COR(7) or

ENT(7) would be appropriate. The combinations {LGA(11), ENT(7)} or {PCM(11), ENT(7)} are preferable since ENT(7) has a lower computational complexity compared to CON(7) or COR(7). Sensitivity, specificity and accuracy of up to 95.4%, 90.1% and 93.1% respectively were attained with {LGA(11), ENT(7)}. This specific combination is used for object discrimination in a real-world test case on a DBM containing a dense pockmark swarm.

Whilst overall binary classification accuracy on the labelling of individual grid node points is a key metric, it is important to consider sensitivity and specificity too in the preceding analysis. Since, the evaluation and validation data used in the feature evaluation process comprised sets of balanced samples of carefully selected control points. However, in the bathymetry data set, the surface area covered by the pockmarks is much less than that of the surrounding terrain. Thus, there are far fewer P than $\neg P$ node points. At the validation stage, it is possible to have a relatively high accuracy with large differences between sensitivity and specificity. However, this can lead to unsatisfactory results when such a model is applied to the unbalanced test data. So in general then, machine models should be sought based on feature sets that yield relatively high validation accuracy, low dimension of feature space *and* a relatively small ($< 10\%$) difference between sensitivity and specificity. In terms of comparing actual object discrimination between the machine and the human (considered in section 5.7), these metrics are less suitable and an alternative scheme is proposed. Some care is needed in specifying a “best” choice of features for the task and furthermore, the choice of features is highly dependent on the methods of measurement, i.e., how the features are evaluated (a topic considered in further detail in case study 3, chapter 7.) These factors have to be considered then, within the context of the abstract measurement framework and the evaluation metrics used. The selection of features is a type of optimisation problem within this methodological framework. Some of the more important objectives that can be considered in an optimisation are (in no particular order):

1. Maximising feature saliency (i.e., maximising accuracy, and or sensitivity and specificity.)

2. Maximising the feature stability and robustness to data and parameter perturbations (this concept is revisited in case study 3, chapter 7).
3. Minimising the difference between sensitivity and specificity.
4. Minimising the dimension of the feature space (choosing fewer features but conceding relatively little accuracy.)
5. Minimising the input size in the feature creation process.
6. Minimising computational complexity of the feature creation methods.
7. Correlation constraints, such as removing features subject to a correlation threshold.

For the analysis here, objectives 1, 3, 4 and to a lesser extent, 6 and 7 were considered. Further, the selected features should have strong semantics such as a physical (or intuitive) relationship between the collective feature set (or individuals within it) and the real properties the features capture information about. Although this is not always easy or possible with high dimensional feature sets, as there may be no obvious connection between the combination of features and the outcomes of the classification task. In the case here though, for one of the best choices of features sets, {LGA(11), ENT(7)}, LGA(11) responds strongly to edges and blobs at this scale and ENT(7) responds to the degree of randomness of the texture. Pockmarks have intrinsic morphological differences compared to the surrounding terrain of the seabed. These properties can therefore be represented meaningfully and used to discriminate the objects from the surroundings by using a machine to learn differences in the randomness of the texture and the occurrence of edge-like structures in the terrain with information from this feature combination. There are also other combinations that may produce similarly good results.

5.6 Real-world object discrimination

To demonstrate and evaluate the prototype process in a practical context, it is applied to pockmark object discrimination in a dense swarm. The DBM, supplied by the NGU, covers a 21 km² section of the Barents Sea. Horizontal resolution is 5 metres between the grid nodes and the DBM contains 844 124 nodes. Processing proceeds in the following stages;

1. The process is initialised manually by visually selecting a small sample of training points comprising approximately 10 P and 10 $\neg P$ class instances from the Graphical User Interface (GUI).
2. A pattern vector (comprising the pre-selected features {LGA(11), ENT(7)} in this specific test case) is generated for the training points and on all of the DBM grid nodes. In the previous analysis on the evaluation and validation control data, accuracy exceeded 90%, the difference between sensitivity and specificity is less than 10% and the feature space is 2-dimensional, i.e., computationally efficient.
3. A BVM classification model is induced with the user selected training points and pre-selected features.
4. Using the model from the previous stage, the classifier hard-labels all DBM nodes as belonging to P or $\neg P$ classes.
5. Sequential labels output from the classifier are transformed into a binary image by encoding the labels as [0 1] and mapping them to the local raster coordinates of the DBM.
6. The binary image is filtered in a morphological closure operation.
7. Boundaries are extracted from the closed image and super-positioned on the DBM raster.

After selecting the training instances all stages are completed sequentially and the boundaries automatically placed, within a few seconds.

DBM Data is loaded into the GUI and displayed as a grey scale. Numerous experiments were conducted to determine an appropriate number of samples to choose. It was found the best results were obtained by selecting a very sparse, balanced sample of approximately 10 instances of each grid node type (<0.003 % of the data.) The instances can be selected individually from widely distributed locations. Red and blue markers in figure 5.12 indicate locations where respectively, P and $\neg P$ training instances, have been selected.

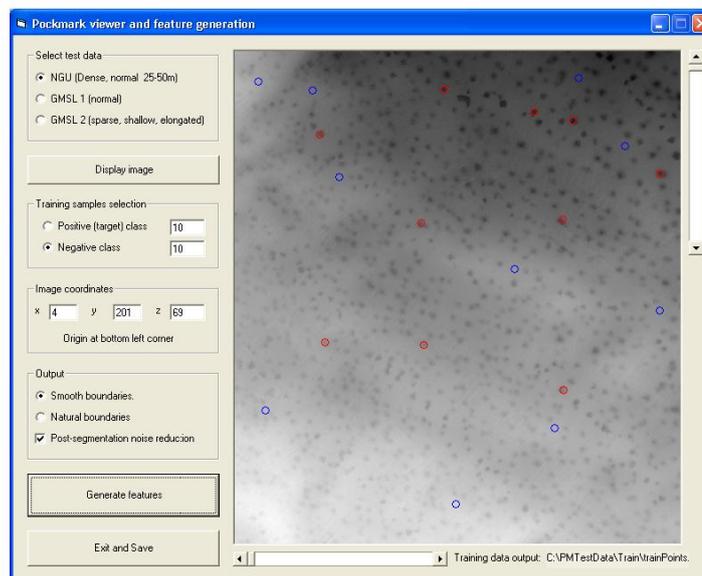


Figure 5.12 The user selects a balanced sample of about 10 instances from the P (red circle) and $\neg P$ (blue circle) node classes by mouse clicks on the rendered DBM. The visible area in this particular case is approximately 7 km², i.e. about one third of the DBM area.

A single BVM model is induced with the pattern instances created at each of the previously selected training nodes. Output from the classifier is a sequential stream of hard [0 1] class labels. This output stream is encoded, as a binary image, with pixels referenced to the grid coordinates of the DBM raster. Effectively, by classifying individual nodes as P and $\neg P$ the DBM is implicitly segmented by mapping the class labels to the binary image space, B .

The segmentation may contain noise in the form of false positives and false negatives and these are removed by filtering, prior to boundary extraction. Morphological closure is used as the filtering process, the closure operation is defined as,

$$B \bullet S = (B \oplus S) \ominus S \quad (5.14)$$

where S is the structuring element. Dilation and erosion are applied with a square structuring element spanning 3 nodes. Further details of morphological operations can be found in, for instance, Gonzalez and Woods (2008).

Finally, after morphological closure, all object boundary locations can be identified in accordance with the earlier definition of the boundary in subsection 5.3.3. The boundaries are superpositioned on top of the DBM raster and a new viewing window opens (figure 5.13) displaying the boundaries of the identified individual pockmark objects in red.

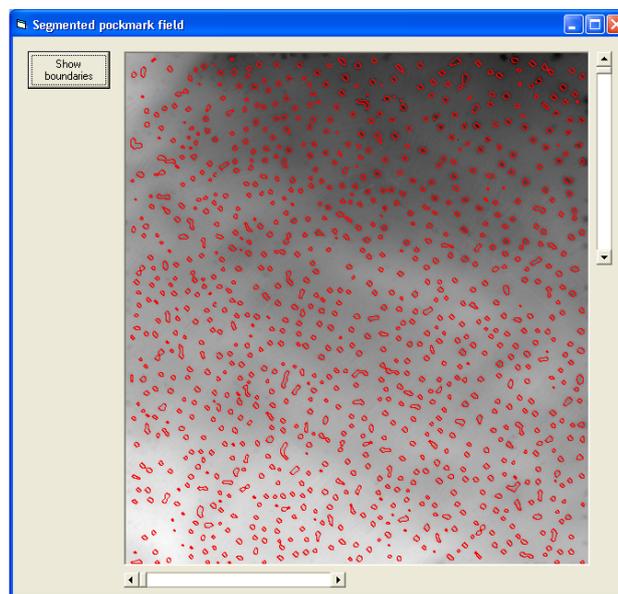


Figure 5.13 The end result of the process is a display showing the boundaries of the identified pockmarks, overlaid on the DBM.

In terms of efficiency, completion of the processing stages including boundary extraction and mapping of approximately 2000 pockmark objects can be accomplished

in seconds on a reasonably specified PC. A human analyst manually mapping say, 100 pockmarks per hour, working non-stop for 10 hours per day, would take two days to complete this task. There would also be intra-rater variability in the boundary placement and shape (and inter rater variability if two or more analysts worked on the same data set.)

Therefore, the two key advantages of the machine process are;

- (1) it can perform the task 3 orders of magnitude faster than a human and
- (2) it produces more precise boundary placement and shapes, according to the object boundary definition.

In the next section, the quality of the object discrimination is considered.

5.7 Evaluation and discussion of results

The machine discrimination of individual pockmarks is compared against human discrimination in 7 sub-regions of the data, shown in figure 5.14, using results obtained from the application set-up, described previously. Each of the regions has an area of 1 km² (40 000 grid nodes) and areal pockmark densities ranging from 76 - 190 km⁻².

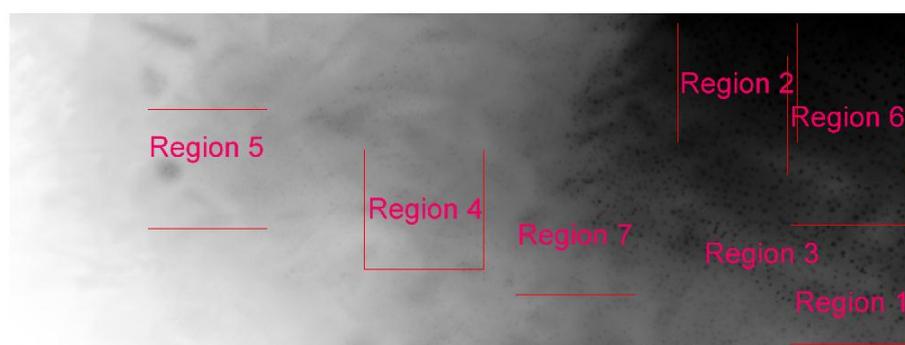


Figure 5. 14 Locations of the 7 regions used to compare machine against human identification of the pockmarks. Each region is 1km² (40 000 grid nodes).

It is perhaps inappropriate to consider “errors” or “accuracy” in this evaluation since in common with much of the other work in the thesis (and many other related studies in the domain) the outcome of the human or machine interpretation/classification is uncertain. There is little or no ground truth available to establish conclusively whether the machine or the human has correctly identified and delineated a meaningful target or not. Seabed depressions are created by a variety of processes, e.g., the base of an iceberg impinging on the seabed may leave an imprint that looks like a pockmark in size and shape. However, as geologists have identified this area as a pockmark swarm, the assumption has to be made that the majority of the landform objects within the area are pockmarks. In other words, if it is about the same size and shape as a pockmark and is surrounded by similar landform objects in a known pockmark field, then it is more plausible that the object is a pockmark than some other type of object formed by a different mechanism.

The comparison between human and machine approaches is based on the number of occurrences of agreements and disagreements in the machine and human identification of an (assumed pockmark) object at the same location. An agreement occurs when the machine and human have identified a discrete object at the same location. Agreement of the boundary shape and position is not considered. There are four types of disagreement, defined as follows;

- Type 1. The human identifies an object not identified by the machine.
- Type 2. The machine identifies an object not identified by the human.
- Type 3. The human identifies multiple objects in one location but the machine identifies one object.
- Type 4. The machine identifies multiple objects in one location but the human identifies one object.

Set M contains machine identified objects, set H , human identified objects. T_1 , T_2 , T_3 and T_4 are the sums of the instances of disagreement types 1-4, respectively, as defined above. O_{T_3} and O_{T_4} are the sums of the individual object counts identified by the human

(for instances of Type 3 disagreements) and for the machine, respectively (for instances of Type 4 disagreements.) The number of objects identified by the machine and human for which there is an agreement, is $|H \cap M|$. The total number of identifiable objects is,

$$O_{total} = |H \cap M| + T_1 + T_2 + O_{T3} + O_{T4}. \quad (5.15)$$

The total number of disagreements on individual objects is,

$$D_{total} = T_1 + T_2 + (O_{T3} - T_3) + (O_{T4} - T_4). \quad (5.16)$$

Overall agreement, A , between the machine and the human can now be quantified as,

$$A = (O_{total} - D_{total}) / (O_{total}) \times 100\%. \quad (5.17)$$

Total disagreements are computed in this way so as not to excessively penalise the human or the machine for agreeing on localisation of an object but disagreeing on the quantity of objects within the localised area. For instance, the machine may identify one object in a specific localised area and the human may identify two smaller, discrete objects. Both approaches have identified the location but are in disagreement over the quantity of objects. The cost in this instance is 1 disagreement. If the machine identified 1 object and the human identified 3 objects, the cost would be 2 disagreements. This evaluation method is then, more complex than a simple contingency test for object identification, since the type 3 and 4 disagreements concern object resolution. It should be expected that very closely spaced objects with a low contrast to the background are harder to resolve individually than higher contrast, well separated objects.

Figures 5.15 shows the comparisons between the machine (left image) identification of pockmarks and the (contrast enhanced) image used by the human analyst for identification of the pockmarks (right) in sub regions 1 - 7 respectively. Type 1 and type 2 disagreements are highlighted by the light blue boxes, type 3 and type 4 disagreements by the magenta boxes. Pockmark boundaries placed by the machine approach are shown in red. The grid (green lines) divides each region into 16 smaller

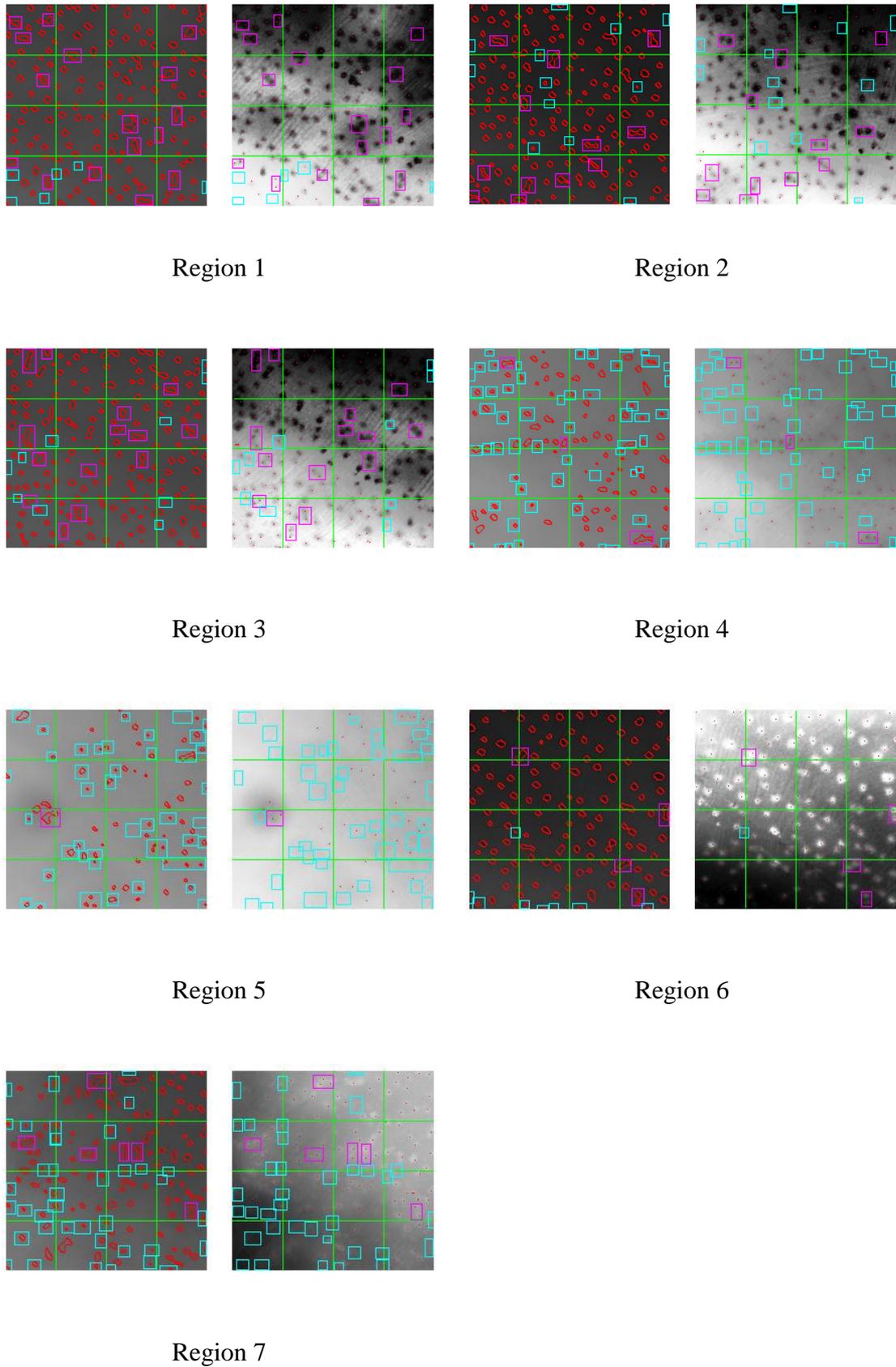


Fig. 5.15. Seven image regions used in the evaluation. For each region, the left image shows the pockmarks and boundaries identified by machine. The right image is used for human identification and counting of individual pockmarks.

(250 m × 250 m) boxes, to help with the comparison and counting of objects identified in the corresponding areas of the left and right images. The number of agreements and disagreements, between the machine and human approaches are summarised in table 5.3. Total number of identifiable objects, total number of object disagreements and overall agreement are summarised in table 5.4.

Region	$ M \cap H $	T_1	T_2	T_3	T_4	O_{T3}	O_{T4}
1	133	0	4		16	0	35
2	139	1	6		12	0	24
3	151	1	6		14	0	32
4	61	1	43		3	0	8
5	288	3	43		1	0	2
6	114	0	2		4	0	6
7	94	0	32		6	0	14

Table 5.3 Summary of agreements and disagreements between the machine and human approaches in each of the 7 subregions.

Sub-region	O_{total}	D_{total}	$A(\%)$
1	172	23	86.6
2	170	19	88.8
3	190	25	86.8
4	113	49	56.6
5	76	47	38.2
6	122	4	96.7
7	140	40	71.4
Σ	983	207	

Table 5.4 Summary of total agreements, disagreements and the overall agreement, A , in each sub-region.

The mean overall agreement, \bar{A} (%), for a number of regions, R , is given by,

$$\bar{A} = \frac{\sum_{i=1}^R O_{total}(i) - \sum_{i=1}^R D_{total}(i)}{\sum_{i=1}^R O_{total}(i)} \times 100 \% \quad (5.18)$$

For a total of 983 objects, in the 7 regions, mean overall agreement, $\bar{A} = 78.9\%$. Highest localised agreements, up to 96.7%, occur in regions 1, 2, 3 and 6, where the contrast between pockmarks and the background terrain is strong and the individual objects are easily identifiable by the human. There are relatively few instances of type 1 and type 2 disagreements in these regions - mostly type 2, situated in localised areas where the visual appearance of the pockmark object is not clearly defined. Some type 3 disagreements are also evident in these regions. This may in part be due to the choice of the LGA(11) feature. In raw segmentation images (fig.5.16) boundaries obtained from LGA(11) tend to be smoother, compared to using PCM(11). Further, LGA(11) can occasionally coalesce two or more neighbouring individuals into a single segmented object, whereas PCM(11) can segment very closely spaced individual pockmarks more effectively.



Figure 5.16 Raw segmentations: PCM(11) (left), LGA(11) (right).

However, the morphological closure operation also has a tendency to coalesce individual objects that are very close to each other. Thus, whichever feature combination is used there will probably be a few neighbouring objects merged during closure.

In all of the regions, the number of type 1 disagreements is very low, indicating it is quite rare for the human to identify an object that cannot be identified by the machine. Small sized and low contrast (presumably very shallow) pockmarks in regions 4, 5, and 7 are quite difficult for the human to identify even when the DBM image contrast is enhanced. There are numerous type 2 disagreements in these regions, indicating that the machine may be more effective than the human at identifying these particular types of pockmarks in this terrain. Of course, the machine also has speed advantages, having placed boundaries around all of the objects identified in the DBM, in less than 20 seconds.

5.8 Evaluation, conclusions, recommendations and scope for further work

5.8.1 Evaluation

The problem of pockmark discrimination in a DBM was identified and solved using a novel feature based machine learning process, capable of accomplishing in a few seconds, a task that would take days for a human analyst to complete. A peer reviewed conference publication arose as a consequence of the research.

Before undertaking the case study it was not known what features would be useful for the discrimination task or how effective the devised process would be. The investigation carried out in the case study provided evidence to demonstrate that it was possible to discriminate a large quantity of pockmarks, and map their boundaries expediently using the proposed process. Different types of features were considered at a variety of resolutions and it was found, through multiple methods of evaluation, that a combination of two features provided very good classification results on validation data. This was subsequently verified on a real-world test case and the quality of the discrimination evaluated by comparing machine and human object identification. The aim and objectives of the study have therefore been successfully achieved.

Compared to planetary and terrestrial land surfaces there has not been much research on the machine discrimination of specific landform objects in DBM's or the types of

features and scales of analysis to use for landform object discrimination. The study contributes generally to this wider domain. Specifically though, this is the first bespoke machine learning process designed for the purpose of pockmark discrimination. Fowler *et al.* (2008) first proposed a process based on conventional land surface attributes such as slope and curvature, derived in a GIS package. More recently, a script based watershed method was proposed by Gafeira *et al.* (2012).

Further research could be carried out to extend the evaluation of the methods and process on areas covered by different types of pockmark. It may also be revealing, in future work, to apply the unsupervised process of case study 8 (with modifications) to establish if the discrimination can be fully automated. The work that has been completed so far can also be built upon further by combining bathymetry and backscatter data to classify different types of pockmark.

Aside from the BVM implementation, all of the software used for the experimental harness was designed and developed *ab initio*, using C++. The main harness comprised several C++ dynamic link libraries (dll) and executables, with a Visual Basic user interface. The reading of intermediate results files and the control of data communication processes between modules in the harness were managed by MS-DOS batch files. Some of the results output from the harness were subsequently post processed and rendered visually using Matlab.

5.8.2 Conclusions

A new feature based machine learning process has been proposed and evaluated, for segmenting and mapping pockmarks in a real-world DBM. Good classification results can be expected by combining either LGA(11) or PCM(11) with one co-occurrence textural feature. For instance, the combination {LGA(11), CON(7)} attained sensitivity, specificity and overall accuracy of 96.9%, 92.0% and 94.8% respectively on the control data. Sensitivity, specificity and accuracy of up to 95.4%, 90.1% and 93.1% respectively were attained with {LGA(11), ENT(7)}. This specific combination was used for object discrimination in a real-world test case on a DBM containing a dense pockmark swarm. The agreement between machine and human on the identification of pockmarks is on average, in this test case, 78.9%. The machine approach may be better

at identifying low contrast objects that are difficult for a human to identify. It is also significantly faster than a human analyst, taking seconds rather than days, to identify and objectively map the boundaries of approximately 2000 individual pockmarks. Furthermore, the placement and shape of the individual boundaries is consistent.

The aim of the case study was to establish the feasibility of automating pockmark discrimination using feature based machine learning methods. A novel process and methods were proposed and evaluated. The two main objectives concerned an investigation of the different types and combinations of discriminatory features for the task and comparing the machine discrimination of pockmark objects with a human discrimination. As has been demonstrated in the chapter and the conclusions, the aim and objectives have been successfully achieved.

5.8.3 Recommendations

In relation to the data under consideration, the BVM classifier gave good results in terms of classification accuracy and the efficiency of model induction and class prediction. Due to its speed and accuracy it is recommended as a hard discriminative classifier for this task, using a RBF kernel. It is likely that other base learners will perform well on this task, such as the Naïve Baye's classifier or SVM. The Naïve Baye's classifier also generates numerical support (probabilities) for the class labels.

A two-dimensional feature space is recommended for this task, comprising LGA(11) and ENT(7). LGA(11) responds to edges and blobs and ENT(7) responds to the degree of randomness of the texture. Pockmarks have morphological differences compared to the surroundings. These properties are captured by the features and used to discriminate the objects by using a machine to learn represented differences in the randomness of the texture and the occurrence of edge-like structures in the terrain.

In order to apply the process to similar types of pockmarks in different data, it is recommended that the approximate range of sizes of the pockmarks (in grid nodes) is first estimated by visual inspection. The range of sizes in the data considered was approximately 20-50 meters. Setting the LGA kernel size corresponding to the upper range of the physical sizes of the pockmark objects and the texture feature at the mid-

range value may be a suitable choice for other data. Resolution will also need to be taken into consideration in this case, to estimate the size of the feature creation kernels to be applied. A feature evaluation process is still therefore recommended when using different data sets, to identify the most useful features and their resolutions for the specific object properties in the data. It is also recommended that the evaluation process is holistic, using a variety of methods to assess the usefulness of individual features and combination of features.

Morphological closure should be applied to remove false positives (FP) and false negatives (FN). In general, choosing a small kernel size for the features leads to a noisier classification result with a greater proportion of FN and FP instances. The kernel sizes specified previously will not result in a large number of FP or FN and will tend to filter out smaller irrelevant objects that are probably not pockmarks.

5.8.4 Scope for further work

There are several ways in which this work could be extended. A few possibilities are outlined here;

- Fusing features derived from DBM and backscatter imagery for discrimination/classification of different pockmark types and extracting other information that could be used in commercial applications (see the mini-proposal in chapter 10, section 10.2.1.)
- Automated counting of the identified objects and determination of object properties such as planar area, eccentricity, volume, depth, and spatial distribution statistics.
- Experimentation on other data sets (subject to availability) containing different types, sizes and areal densities of pockmarks.
- A comparison of supervised and unsupervised approaches, e.g comparing the proposed approach with a full automation using a pre-clustering stage (as applied to sonar mosaic imagery in chapter 8.)

- A comparison with non-feature based approaches such as watershed methods and crater discrimination algorithms.
- Representing the uncertainty in the discrimination of the individual objects by making use of the decision values (distance of pattern instances from the separating hyperplane in the BVM or an SVM.) Alternatively, a Naïve Bayes classifier could be used to generate probabilistic support.
- Evaluating the agreement between human analysts and machine, in the placement and shape of pockmark boundaries. Multiple interpreters used in object identification.

In the next chapter, feature creation and evaluation concepts are advanced and considered in a more challenging case study on another important natural seabed target discrimination problem – the discrimination of Sabellaria textures in sidescan sonar imagery.

Feature based discrimination Chapter 6 of Sabellaria spinulosa textures in sidescan sonar imagery

The work presented in section 6.2 of this chapter resulted in the publication of a paper, “A texture analysis approach to identifying Sabellaria spinulosa colonies in sidescan sonar imagery” by Harrison *et al.* (2011).

Contents

6.1 Introduction

6.2 Sabellaria texture discrimination in waterfall imagery

6.2.1 Methods

6.2.2 Experimental work

6.2.3 Results and discussion

6.3 Sabellaria texture discrimination in mosaic imagery

6.3.1 The Gabor filter bank

6.3.2 Experimental work

6.3.3 Results and discussion

6.3.4 Further results and considerations

6.3.5 Summary

6.4 Other useful features

6.5 Evaluation, conclusions, recommendations and scope for further work

6.5.1 Evaluation

6.5.2 Conclusions

6.5.3 Recommendations

6.5.4 Scope for further work

6.1 Introduction

In the previous chapter, the identification of targets (pockmark objects) in a DBM was studied. The case study reported in this chapter approaches the novel task of discriminating a target textural class, representing potential *Sabellaria spinulosa* (*Sabellaria*) colonisation in sidescan sonar imagery. Manual (visual) discrimination and mapping of the *Sabellaria* is an important but time consuming and expensive task. The aim of this case study is to establish if machine discrimination of *Sabellaria* textures from other seabed texture classes in sonar imagery is a tractable problem. The objectives concern the evaluation of a selection of state-of-the-art feature creation methods on waterfall imagery to identify any promising methods for *Sabellaria* discrimination (section 6.2) and a more in-depth analysis of a single method, applied to a similar task in mosaic imagery (section 6.3). The main findings indicate that a tuned Gabor filter bank is a useful feature creation method for *Sabellaria* discrimination in waterfall and mosaic imagery. Other potentially useful features for *Sabellaria* discrimination are briefly considered in section 6.4.

Due to (documented) external issues affecting the provision of data, it was not possible to investigate any of these methods for *Sabellaria* discrimination on multiple or larger sized real-world data sets. However, it is not uncommon for published studies in the domain to focus on smaller sections of imagery from a single data set, as opposed to much larger mosaics. Using smaller patches of verified imagery greatly reduces the uncertainty of the study. Despite the noise, distortions and intra-class variability, there is more certainty that a smaller image patch contains a relatively homogeneous region of texture, as in the examples of figures 2.3 (a) and (b) in chapter 2. When there is ground truth and expert visual inspection, this helps to confirm that the texture is in some way representative of the seabed class. It is pointed out in section 6.5 that the establishment of a seabed texture benchmarking database would be very useful for further increasing the understanding of and advancing texture analysis methods in this domain.

6.2 Sabellaria textures in waterfall imagery

This section summarises an experimental investigation into the performance of some statistical, signal processing-based and morphology-based texture features for classifying seabed classes in qualitative sidescan sonar imagery. The primary concern is the detection of the tube-building worm *Sabellaria spinulosa*. (*Sabellaria*). *Sabellaria* inhabits sub-tidal and intertidal zones in UK and European waters (Limpenny *et al.*, 2010). Dense accumulations of such worms can form biogenic reefs which are protected under the EU Habitats Directive. Naturally, this has planning and commercial ramifications for the offshore renewable energy industry as well as other construction activities that cause a seabed disturbance.

Sidescan sonar is often the instrument of choice for high-resolution acoustic imaging of the seabed (see for instance, Blondel 2007, Fish and Carr, 2001.) The imagery contains useful textural information that can indicate the presence and extent of *Sabellaria* colonies (Limpenny *et al.*, 2010, Hendrick and Foster-Smith, 2006, Birchall, 2007.) The sonar is usually towed on a cable behind a survey ship. It generates pulsed, fan-shaped lobes of acoustic radiation from a transducer array that doubles as a receiver for returning signals. The acoustic energy undergoes a complex environmental interaction as it propagates through the water column, into the seabed and back to the instrument, as illustrated in figure 6.1.

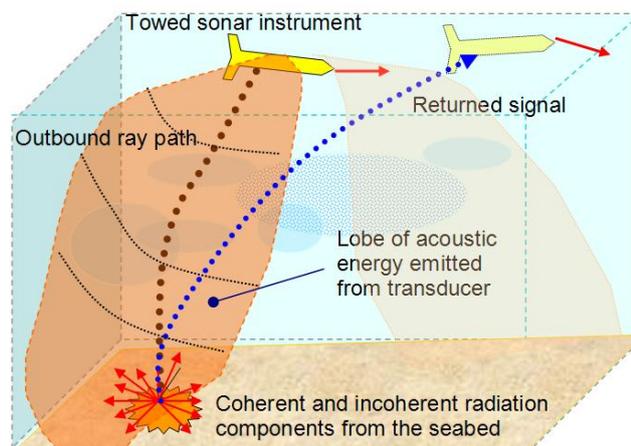


Figure 6.1 Illustration showing the basic operation of a sidescan sonar

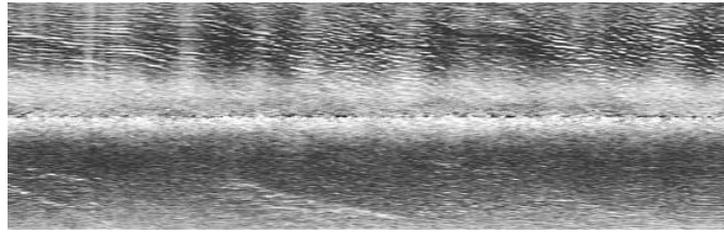


Figure 6.2 A waterfall image segment is built up progressively, from a sequence of transverse scan lines as the instrument is towed above the seabed.

Returning signals are captured as a time series of amplitudes, comprising the sum of spatiotemporally varying specular components and boundary and volumetric backscatter from the seabed and water body. As the instrument advances, a waterfall image of the seabed (figure 6.2) is built up from consecutive scan lines. The imagery contains noise and various radiometric and geometric distortions from multiple sources (Cervenka and de Moustier, 1993, Cobra et al., 1992.) Image degradation can be exacerbated in the shallow water regimes where Sabellaria are found, due to wave motion and surface noise, such as from rainfall.

Optical imaging of the Sabellaria offers much higher resolution and is sometimes used for localised, non-intrusive ground truthing over small areas. Optical and sonar imaging are both stochastic processes but there are fundamental differences in the imaging physics and the transducers. Underwater optical cameras (still and video) use charged coupled device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) arrays of pixel elements for visible-spectrum imaging (Jonsson *et al.*, 2009.) Absorption of electromagnetic radiation by seawater, backscattering due to turbidity and the limited ground coverage make this a less practical option for surveying large areas though. Recent advances in underwater imaging technology are described in Kocak *et al.*(2008).

The heterogeneous structure and spatial distribution of the Sabellaria colonies produces diverse textural proxies in sonar imagery. Identification, segmentation and mapping of these textures is predominantly a manual task performed by a human expert. However, this procedure is tedious, subjective and therefore expensive. An automated, objective approach to reliably discriminating textural regions characteristic of Sabellaria would assist in the process of mapping the colonies. Further, automation would expedite

processing of the large volumes of data currently being acquired in site surveys for offshore wind farm developments.

Different methods for the generic textural analysis of sonar images have been studied widely and include; Gabor filters (Samiee and Rad, 2008), Wavelets (Wang *et al.*, 2007, Celik and Tjahjadi, 2011) and Co-occurrence matrices (Reed and Hussong 1989, Blondel *et al.*, 1998, Karoui *et al.*, 2005.) Yet there is no published research on the features that could be used for our specific task - the automated discrimination of Sabellaria in sidescan imagery. Due to data dependent performance, we need to evaluate a set of methods on this novel texture classification problem, as a preliminary stage in the design of an expert system for automatic Sabellaria discrimination.

The remainder of section 6.2 is organised as follows. In 6.2.1 the methods included in the comparison are briefly reviewed. Experimental set-up including image acquisition, preparation and accuracy estimation is outlined in 6.2.2. Results are presented and discussed in 6.2.3.

6.2.1 Methods

During the last 20 years many approaches to texture analysis have been described in the literature. A comprehensive and up-to-date review can be found in Xie and Mirmehdi (2008.) Here, we consider methods from three groups: Statistical, Signal processing-based and Morphology-based. All the approaches included in the comparison are rotation invariant as we do not assume any preferred orientations of natural textures in the imagery. Given the large palette of textural descriptors available, we restrict our subset to include some well-known techniques in the sonar imaging domain and others whose performance on this type of imagery is untested. The saliency of all these features in relation to the specific task is unknown though. Thus, the aim of our investigation is to find out which of these feature groups is the most promising for the task.

[1] Statistical methods

Statistical methods are based on the statistical distribution of greyscale values at predefined relative positions. The methods considered here are: Local Binary Patterns

(LBP), Improved Local Binary Patterns (ILBP), Coordinated Clusters Representation (CCR), co-occurrence matrices (COOC) and the recently introduced Binary Gradient Contours (BGC) (Férandez *et al.* 2011). LBP, ILBP and CCR have received significant attention in recent years due to their ease of implementation, low computational cost and high discrimination accuracy in many applications, see Férandez *et al.* (2010). Co-occurrence matrices are included as a useful benchmark in many comparisons and are appropriate here due to their widespread use in the sonar domain.

The basic version of the LBP ($LBP_{3 \times 3}$) uses the 256 possible binary patterns that can be defined in a 3×3 window, taking the central pixel value as a threshold. The 3×3 subscript is used to indicate this. A rotation invariant (*ri*, as indicated in the superscript) version ($LBP_{8,1}^{ri}$) is obtained by replacing the original 3×3 window with a circular one and considering all rotated versions of the same pattern to be equivalent (Ojala *et al.*, 2002.) This reduces the number of histogram bins (features) by a factor of 8 (since there are 8 directions separated by $\pi/4$ radians) to 36. The subscript 8,1 is used in this case and unless stated otherwise, the same definitions for the superscripts and subscripts are applied in the features subsequently described.

ILBP (Jin *et al.*, 2004) is an extension of the LBP. The main differences are: (1) The threshold is the mean value of the nine pixels in the window. (2) The central pixel is included in the definition of the binary patterns (in the LBP it is excluded). Hence, there are 511 possible binary patterns for the basic version ($ILBP_{3 \times 3}$), since one of the patterns (all black pixels) is impossible by definition. The number of features is reduced to 71 in the rotation invariant case ($LBP_{8,1}^{ri}$), which is obtained as described previously.

The BGC considers the binary gradients defined between pairs of adjacent pixels lying on the peripheral closed path of a 3×3 grayscale window. This approach, despite discarding the central pixel, has been shown to be potentially superior to the LBP on a theoretical basis (Fernandez *et al.*, 2011.) In this work the version referred to as BGC1 is used. This descriptor, which generates 255 features in the basic version ($BGC1_{3 \times 3}$), can be easily made rotationally-invariant ($LBP1_{8,1}^{ri}$), in the same way as the LBP. In this case the number of features reduces to 35.

The CCR (Kurmyshev and Sanchez-Yanez, 2003) differs from the previous methods in the thresholding approach. Whereas this is *local* in both LBP and ILBP, it is *global* in the case of the CCR. This gives 512 possible binary patterns for the basic version ($CCR_{3 \times 3}$), which reduces to 72 in the rotation invariant version ($CCR_{8,1}^{ri}$).

Co-occurrence matrices (Haralick *et al.*, 1973) estimate the joint probability distributions of gray level combinations for pixel pairs at fixed displacements and orientations. In the implementation used here eight displacement vectors are considered: $\{(1,0), (1,1), (0,1), (-1,1), (-1,0), (-1,-1), (0,-1), (1,-1)\}$. The following five features are derived from the distributions: contrast, correlation, energy, entropy and homogeneity. Rotation invariance is achieved by averaging the feature values over the orientations.

[2] Signal processing-based methods

Signal processing methods usually involve feature generation by filtering the image through suitable filter banks and computing global statistics from the filter responses. Gabor filtering, the Dual-Tree Complex Wavelet Transform (DT-CWT) and Ring filters are considered in this study.

Gabor filters are one of the most effective approaches to extracting textural features. It is commonly believed that their effectiveness is related to their ability to model the frequency-orientation decomposition performed by simple cells in the mammalian visual cortex. In these experiments two Gabor filter banks are used: one with four frequencies and six orientations (Gabor 4-6) and the other with six frequencies and eight orientations (Gabor 6-8). In both cases the other parameters are: maximum frequency = 0.327, $\eta = 0.5$, $\gamma = 0.5$, frequency ratio = half-octave. These values have been chosen in compliance with the guidelines suggested in Bianconi and Fernández (2007). Texture features comprise the mean and standard deviation of the absolute value of the transformed images. Rotation invariance is obtained through DFT normalisation, generating 32 and 60 features for the two banks, respectively.

The DT-CWT (Kingsbury, 1998) has interesting properties, such as moderate redundancy and directional selectivity. It operates on the directions $\pm 15^\circ$, $\pm 45^\circ$, and $\pm 75^\circ$. To achieve consistency with the number of scales and orientations in the Gabor

filtering, four scales and the six orientations are used (DT-CWT 4-6). For each sub-band the mean and the standard deviation of the absolute value of the CWT coefficients are used as texture features. The features generated at each scale are averaged over the six orientations in order to obtain rotational invariance (DFT normalisation is not recommended in this instance, due to differing sensitivities of the complex wavelets). This configuration results in $4 \times 2 = 8$ features.

Ring filters (Coggins and Jain 1985) are also well known in texture analysis. Being based on circular Gaussian transfer functions, they are intrinsically invariant against rotation. In the experiments presented in this paper we employed a bank of five filters with centre frequencies 1, 2, 4, 8 and 16 cycles/image. The other filtering parameters are set as suggested in Coggins and Jain (1985.) Texture features are the mean and standard deviation of the absolute value of the transformed images, resulting in a feature vector of dimension 10.

[3] Morphology-based methods

Morphology-based methods extend the classical digital morphological operations to greyscale images. Two well-known methods are considered: granulometry and variogram.

Granulometry (Hanbury *et al.*, 2005) is the normalised sum of the pixel values of an image, when transformed with a family of openings and closings, as a function of the size of the structuring elements. In the experiments, we use four linear structuring elements with orientations $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ for opening and closing. The dimension of the elements ranges incrementally, in steps of 4, from 2 to 50 pixels. Rotationally-invariant features are produced by averaging the four granulometry vectors corresponding to each direction, giving $13 \times 2 = 26$ features.

The variogram (Wackernagel, 2003) estimates greyscale difference as a function of the distance between pixels. Given two generic pixels x_1 and x_2 the average greyscale difference is defined as $1/2E\{[g(x_1) - g(x_2)]^2\}^{1/2}$, where g represents the greyscale value and E the expected value. In the experiments, eight variograms were considered, corresponding to the displacements $(d_1, d_2) = \{(n, 0), (n, n), (0, n), (-n, n), (-n, 0), (-n, -$

n), $(0, -n)$, $(n, -n)$ for $n = 1, \dots, 20$. For rotation invariance the average variogram is computed over the eight displacements, which yields a 20-dimensional feature vector.

All of the individual feature generation methods described above were implemented in a Matlab harness.

6.2.2 Experimental work

[1] Data acquisition and preparation

An Edgetech¹ 500 kHz towed sidescan sonar was used to capture the acoustic imagery. Range on port and starboard channels was set at 100m and along-track samples acquired at approximately five pings (scan lines) per metre. A subset of survey line waterfall segments were down-sampled in the across-track direction at five pixels per metre and slant range corrected in Coda GeoSurvey² processing software. No further geometric or radiometric corrections were applied and the waterfall image segments were output as greyscale [0, 255] Tiffs, with a fixed (software specific) five-bit radiometric resolution. The image sample regions were chosen from transects oriented in approximately the same directions, and as far as possible, from the central region of the imaged seabed on the same side of the transducer (i.e., sonified from approximately the same directions and incidence angles) where the towfish was moving relatively smoothly, in a straight line and the textures of interests were present. Generally the representative samples were chosen within about 100 metres of each other.

For the classification experiments, 40 image regions (20 Sabellaria and 20 non-Sabellaria) each 256×256 pixels (approximately $50 \text{ m} \times 50 \text{ m}$ ground coverage) were extracted from the processed waterfall segments. The non-Sabellaria regions are further sub-classified as Sand, Mussels, Bedforms and Boulders. These classes are defined in table 6.1 and the number of class instances listed. Typical appearances of the five textural classes are shown in figure 6.3.

¹ <http://www.edgetech.com/edgetech/gallery/category/side-scan-sonar-systems> [accessed 13-07-2011]

² <http://www.codaoctopus.com/coda-geosurvey/> [accessed 13-07-2011]

Class	Image region description	No of images
Sabellaria	Moderate Sabellaria egggregations	20
Sand	Sandy and mixed sandy sediments	6
Bedforms	Linear dunes, wavelength, $\{2 \leq \lambda \leq 5\}$ m	6
Mussels	Mussel beds on a sandy substrate	6
Boulders	Boulders on a sandy substrate	2

Table 6.1 Summary of textural classes

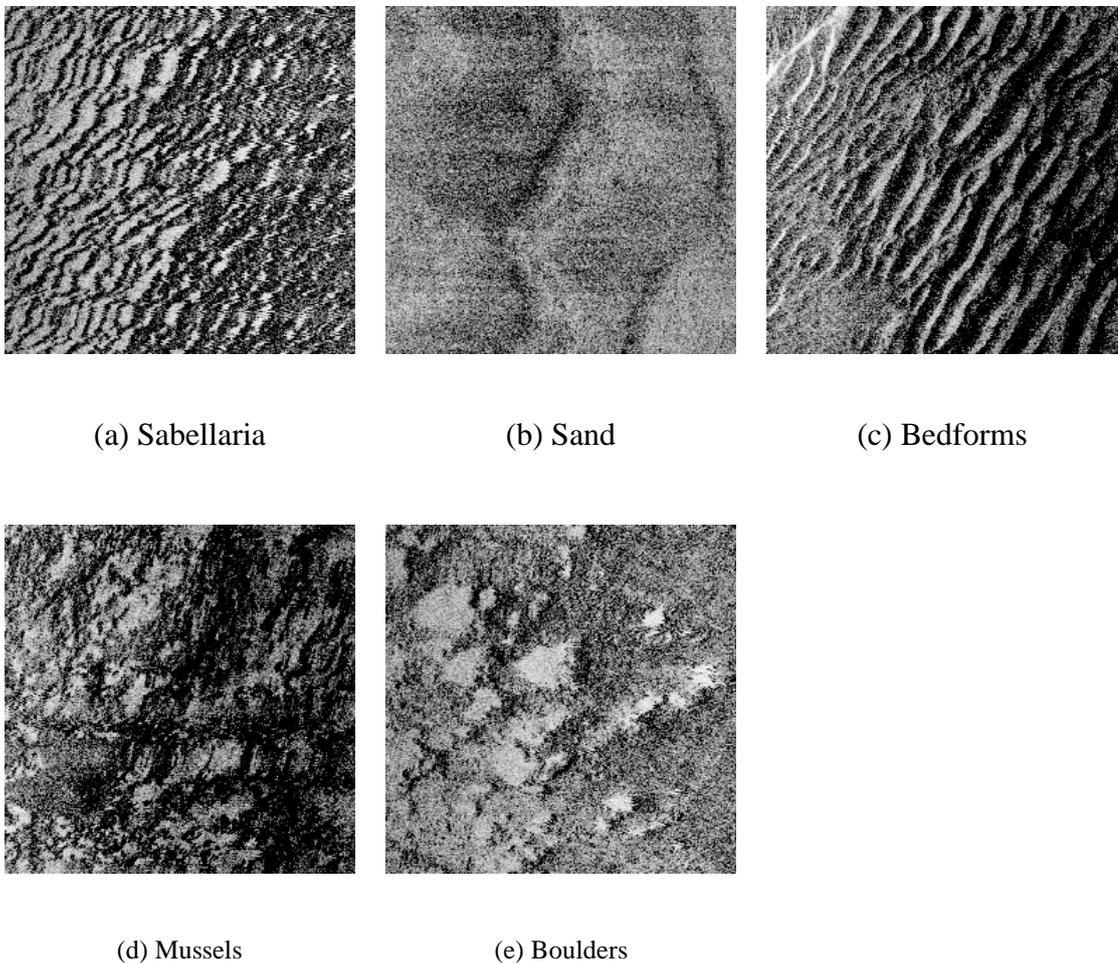


Figure 6.3 Examples of the five textural classes

Natural textural variability is clearly present. For instance, bedforms exhibit varying wavelengths and coherencies in individual image regions as in (figure 6.4) and within the class. These clearly reflect true natural variability of the target together with any distortions generated through unknown variations in the image acquisition parameters.

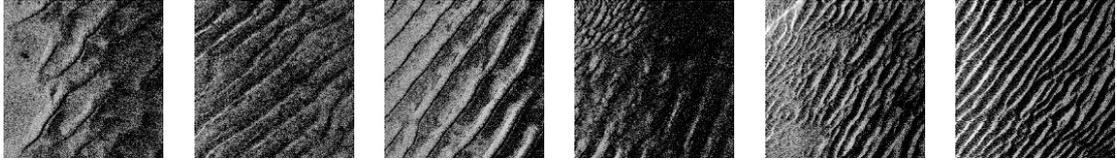


Figure 6. 4 intra-class variability of bedform wavelength, amplitude and coherency.

[2] Classification and accuracy estimation

To assess the effectiveness of the methods described earlier, a multi-class supervised classification experiment was carried out using a nearest-neighbour (1-NN) classifier with L_1 (Manhattan) distance. The choice of 1-NN is motivated by its absence of tuning parameters, its ease of implementation and by the asymptotic property that its error is bounded by twice the Bayes error as the number of samples tends to infinity. The choice of distance (similarity) measure can affect the k-NN results, since different distance measures can produce different sets of k-neighbours. When there is no prior knowledge of what distance measure is most suitable for a particular classification task, the Manhattan or Euclidean distances are conventionally used (Wang *et al.*, 2006.) Results for the lowest error rates on five benchmark classification tasks (unrelated to the problem being considered here), in Wang *et al.* (2006) indicate that the mean error rate is slightly lower for the Manhattan measure, compared to the Euclidean. It does not necessarily mean that the error rates will be lower by using the Manhattan, as opposed to Euclidean measure for the test case here due to the different data and objectives. In the absence of any information to suggest otherwise though, it is not an unreasonable decision to choose this measure. Accuracy estimation is based on leave-one-out cross validation. The classifier and validation process was implemented using Matlab library routines.

6.2.3 Results and discussion

A potentially useful feature generation method for this task should yield good classification accuracies on relatively small sub-image sizes. Table 6.2 shows mean classification accuracies (%) for each method on the different sized sub-images. Results are partitioned into overall multi-class accuracies (left) and Sabellaria-specific accuracies (right) - i.e. sensitivity with Sabellaria as the target.

Method	Sub-image size (pixels)							
	Overall multi-class accuracy				Sabellaria-specific accuracy			
	32	64	128	256	32	64	128	256
BGC1 _{8,1} ^{ri}	62.0	78.3	89.4	82.5	70.1	87.5	95.8	83.3
LBP _{8,1} ^{ri}	62.0	72.0	75.0	62.5	78.3	83.4	86.3	75.0
ILBP _{8,1} ^{ri}	66.7	80.9	85.6	75.0	82.3	91.6	97.5	90.0
CCR _{8,1} ^{ri}	63.7	74.1	87.5	70.0	80.9	91.3	98.8	85.0
COOC	58.2	56.9	51.9	45.0	70.9	67.8	63.8	65.0
Gabor 4-6	66.8	82.7	89.4	72.5	77.5	90.3	93.8	90.0
Gabor 6-8	68.9	83.4	88.8	82.5	79.4	89.1	92.5	95.0
Ring filters	75.5	84.1	86.3	60.0	88.0	95.8	91.7	66.7
DT-CWT 4-6	67.5	74.2	76.3	77.5	80.7	86.6	83.8	85.0
Granulometry	62.9	68.8	67.5	60.0	75.6	80.6	81.3	85.0
Variogram	59.4	69.5	66.6	65.0	76.7	87.5	86.3	85.0

Table 6.2 Overall multi-class classification accuracy (sensistivity) for Sabellaria.

A ranking of the methods is defined through the following pairwise comparison rule: A method outperforms another method if its accuracy is greater on each sub-image size. Assigning +1 for a win, -1 for a loss and 0 for a tie we obtain the results shown in table 6.3. The overall accuracy as a function of sub-image size for the top five and bottom ranked texture descriptors is shown graphically in figure 6.4.

Features	Score	No. features	Rank
Gabor 6-8	7	60	1
Gabor 4-6	5	32	2
DT-CWT 4-6	3	8	3
ILBP _{8,1} ^{ri}	3	71	4
BGC1 _{8,1} ^{ri}	2	35	5
Ring filters	1	10	6
CCR _{8,1} ^{ri}	1	72	7
LBP _{8,1} ^{ri}	-3	36	8
Granulometry	-4	26	9
Variogram	-5	20	10
COOC	-10	5	11

Table 6.3 Total scores and rank for each method

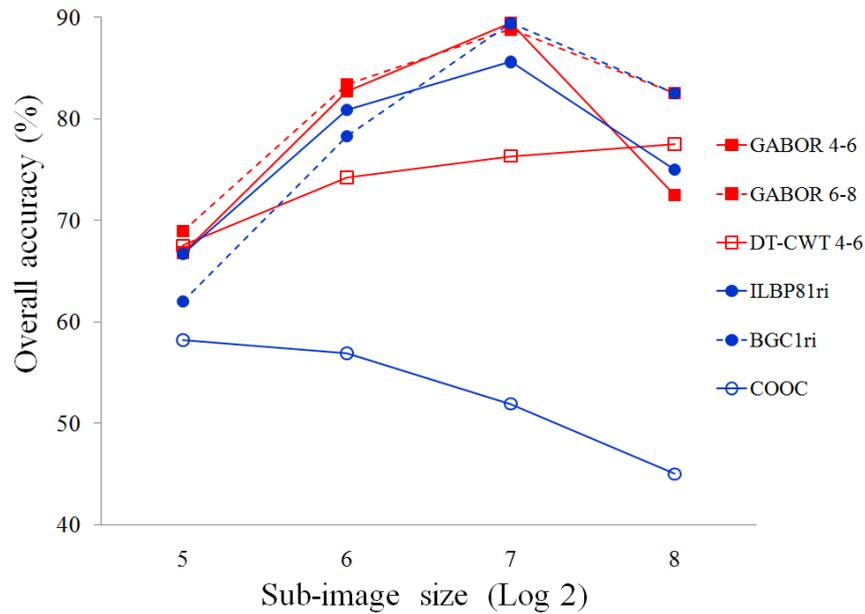


Figure 6.5 Variation in overall accuracy with sub-image size

These results clearly indicate the superiority of signal processing methods on this texture classification task. Both Gabor and ring filters produce good results, with an overall classification accuracy as high as 89.4% and an accuracy for Sabellaria-specific images exceeding 90%. These two methods also yielded some of the highest accuracies on specific classes of Sand and Bedforms. This is very useful, as in their natural habitat, Sabellaria require sand to construct their tubes and they are often found in regions where relatively strong currents (hence bedforms) exist. Features producing good classification accuracies on the Sabellaria, Sand and Bedform classes should also be effective for discriminating between these classes and performing the desired image segmentation.

The classification could be extended to include distinguishing between different periodicities in bedform textures using a Fourier Transform approach. Variations in the spatial-frequency can occur over short length scales as shown in figure 6.3 (c). In this instance, wavelengths range from approximately 2 - 5 metres over the region, so a short-time, windowed Fourier Transform with good spatial and spatial-frequency localisation would be required. The choice of windowing function for the transform would therefore be Gaussian as this optimises the joint spatial and spatial-frequency localisation. Further, we would want to analyse at multiple frequencies. Clearly though, a Gaussian windowed short time Fourier Transform applied at different centre frequencies is a

Gabor expansion. The filter banks are computational implementations of this expansion and therefore have the added versatility of being able to perform this type of periodicity discrimination, should it be needed.

Among the statistical methods investigated, the ILBP and BGC1 gave good results too. As far as is known, these two methods have not been applied to sonar imagery before. Their performance here suggests that they could be useful additions to the pool of methods currently being used in the domain. Surprisingly the LBP does not appear to perform particularly well in this context. It is believed this could in part be a consequence of the image formation process. Downsampling, radiometric compression and other processing factors will result in a loss of some high-frequency components present in the original signal, as noted in Stewart *et al.*, (1994.) These high frequency components are vital for methods based on small scanning windows such as the LBP.

Most methods exhibited a peak overall accuracy at a sub-image size of 128 pixels, corresponding to a square ground patch of approximately 25×25 m (625m².) This indicates that there may well be an optimal size of imaged area for the classification task, under the specific parametric conditions of the feature creation methods used in this test case. The COOC features were ranked bottom out of all the methods considered and as can be seen in figure 6.4, accuracies from the COOC features decreased monotonically by 13% over the range of sub-image sizes tested. This is most likely due to the small magnitude of the displacement vectors in relationship to the scales of the textural variations and also noise in the imagery. It is possible to achieve good results with co-occurrence features on sonar data but this requires careful parameter and feature tuning for the specific data set and texture discrimination tasks. Further details of an approach to co-occurrence parameter optimisation can be found in Blondel and Gómez-Sichi (2009) and in chapter 7 of this thesis.

6.3 Sabellaria texture discrimination in mosaic imagery

The previous investigation of different features on sonar waterfall imagery revealed the Gabor Filter Bank (GFB), under the prescribed parameterisation, as one of the top methods for Sabellaria discrimination. In this section, the parameterisation and design of a GFB, for the discrimination of Sabellaria textures in of sonar mosaic imagery is

considered in greater detail. Mosaic imagery is obtained by applying a mosaicing process to waterfall imagery. The mosaicing process georeferences the pixel values and in so doing, the geometry and radiosity of the waterfall pixels is transformed. The processing steps applied in the production of the supplied mosaic imagery in this study are unknown (and unknowable). The aim of this section is to identify a useful set of filter parameters and filter bank configurations that would be useful for discriminating Sabellaria textures from other textures in sonar imagery.

6.3.1 The Gabor filter bank

[a] *Background*

Research concerning the design and parameterisation of Gabor filter banks and their numerous applications to image texture representation, image retrieval, segmentation and classification is extensive, e.g Manjunath and Ma (1996), Bovik *et al.*(1990), Jain and Farrokhnia (1991), Dunn *et al.* (1994). However, for reasons that are unclear, their use in sonar image processing has not been widely adopted. There are presently no commercial software packages for sonar processing that include Gabor filter banks as a textural analysis and segmentation method. Published research on the application of Gabor filter banks to sonar imagery is scarce compared to methods such as GLCM. Further, aside from the study on waterfall data, in the previous section, there is no published work at all, relating to the specific case of discriminating Sabellaria from other textures in sonar images.

One of the earliest applications of Gabor filters to sonar imagery was carried out by Cexus and Boudraa (2003). They used texture energy features derived from the filter output channels, in an unsupervised approach to segmentation of seabed textures. Atallah (2004) experimented with different parameterisations of Log-Gabor filter banks for the classification and segmentation of sidescan and bathymetric sidescan imagery. He found that adding more orientations to the filter bank improved classification results by up to 20% on the sidescan data set. However, with a different data set containing imagery of artificial coral reefs surrounded by sediment, he found that increasing the number of orientations in the filter bank did not greatly improve the classification results. Atallah (2004) claims this is probably due to random orientations of the textures. Further, the comparative effects of dyadic (octave) and non-dyadic

implementations of frequency scales on the results of the reef – sediment discrimination case were not significant. Sun and Shim (2008) used a fusion of Gabor filter bank features and Fuzzy Fractal Dimension (FFD) applied to the classification of sand, rock and mud textures. They found that fusing the two feature sets led to an improvement in classification performance, compared to using the features independently. Their Gabor filter bank was implemented using only the even symmetric (real) kernel (a Gabor filter kernel is complex valued and comprises real and imaginary components). Similar implementations, in different applications, were adopted by Jain and Farrokhnia (1991) and Camapum Wanderley and Fisher (2001). Malik and Perona (1990) provide some justification for an even symmetric kernel choice, based on psychophysical grounds.

High-level texture perception in the Human Visual System (HVS) depends on many factors but is strongly dependent on repetition (frequency), orientation and complexity properties of the texture (Rao and Lohse, 1993). Prior to this work, Campbell and Kulikowski (1966) had found that an aspect of HVS functionality could be described in terms of orientation tuned channels. One approach to modelling the perception of texture by the HVS is therefore to use a multi-channel filter bank, which captures frequency-orientation characteristics of a texture. However the parameters used to model this aspect of the HVS in the context of a multi-channel filter bank (such as Gabor) are not necessarily appropriate for machine based textural discrimination problems. For instance, use of the even symmetric Gabor function in isolation may lead to inconsistent classification results compared to using the complex filter pairs, particularly when no post filtering is applied (Clausi and Jernigan, 2000). Further, cells in the visual cortex respond to approximately dyadic frequency bandwidths (Pollen and Ronner, 1983). Clausi and Jernigan (2000) claim that a filter bank model of the HVS, with 30° angular separation and one octave spacing/bandwidth is preferred. However, in the experimental work presented in this section for the specific task of Sabellaria discrimination, half-octave separation is shown to be preferable.

Numerous different post processes can be applied to the output channels of the Gabor filter bank, for instance; Gabor-energy (magnitude), complex moment, grating cell operators and full-rectification. These responses are often used directly as textural features, several are compared in a study by Grigorescu *et al* (2002) and also in Clausi and Jernigan (2000). Features derived from the complex phase output can also be used to detect discontinuities within relatively homogeneous textural classes (Du Buf and

Heitk, 1991). Since the focus here though is on binary discrimination of textural classes rather than intra-class textural discontinuities, the phase information is not considered further.

There are a multitude of ways of applying a Gabor filter bank, configuring the output channels and deriving features. The precise approach used will depend on the application objectives and the desired output for the subsequent processing pipeline. In this study, the filter bank is used as a combined feature generator and down-sampler. The objective is to represent small, contiguous, non-overlapping image neighbourhoods with a feature vector capturing the frequency-orientation properties of textures within the pixel neighbourhood.

[b] *Filter bank definition and implementation*

The theory of Gabor filter banks and their mathematical description is widely documented. An individual Gabor filter is built from a 2-D Gaussian function modulated by a complex, planar sine wave. It is sometimes called a Gabor Elementary Function (GEF) or a Gabor Wavelet. The GEF was originally defined by Gabor (1946) and subsequently extended to a 2-Dimensional space by Daugman (1985). The Gabor filter bank, Ψ comprises a total of $|\Theta||\Omega|$ individual filters, at orientations, $\Theta = \{\theta_0, \dots, \theta_{|\Theta|-1}\}$ and centre spatial frequencies (wave numbers), $\Omega = \{\omega_0, \dots, \omega_{|\Omega|-1}\}$. So, for example, if there are eight orientations ($|\Theta| = 8$) and five frequencies ($|\Omega| = 5$) there are 40 individual filters in the filter bank. A spatial (x, y) domain template function for a two-dimensional, individual Gabor filter kernel can be defined as,

$$\psi(x, y) = \frac{\omega_j^2}{\pi\sigma_p\sigma_n} \exp\left\{-\pi\left(\frac{x_r^2}{\sigma_p^2} + \frac{y_r^2}{\sigma_n^2}\right)\right\} s(\mathbf{k}, \mathbf{x}) \quad (6.1)$$

where,

$$s(\mathbf{k}, \mathbf{x}) = \exp\{2\pi i(\mathbf{k}\cdot\mathbf{x})\} - \exp\{-\pi(\sigma_p^2 u_{0(r)}^2 + \sigma_n^2 v_{0(r)}^2)\}. \quad (6.2)$$

In equations 6.1 and 6.2, the σ_p , σ_n terms represent the standard deviation of the Gaussian envelope in directions parallel and normal to the x -direction, respectively. The modulating spatial frequencies of the sinusoid in the corresponding directions are u_0 and v_0 . The spatial frequency (wave number) and coordinate vectors, \mathbf{k} and \mathbf{x} respectively (defined below) are used for compactness of the notation, which varies widely (*c.f.* equation (91) in Movellan (2002), equations (2), (3) and (4) in Dunn *et al.* (1994) and equation (1) in Bianconi and Fernández (2007)). The ‘ r ’ subscript indicates transformed (rotated) components and as usual, $i = \sqrt{-1}$.

Individual filter kernels have real and imaginary components, as shown in figure 6.6. Output from the kernels may be used independently or combined, commonly as the magnitude response. Although, as noted previously, many implementations use only the real (cosine) component of the filter kernel. A cosine or sine kernel only implementation is phase sensitive, since the complex sine kernel is a $\pi/2$ phase shifted version of the real cosine kernel. Further, unless the DC component of the real kernel is compensated or filtered, it will not be intensity invariant as the output will be biased by intensity. The output from the complex sinusoidal kernel has a zero DC component and is therefore unbiased. Using a normalised (zero DC) magnitude response from the real and complex kernels ensures that the filter output is insensitive to intensity and to phase, as in Movellan (2002).

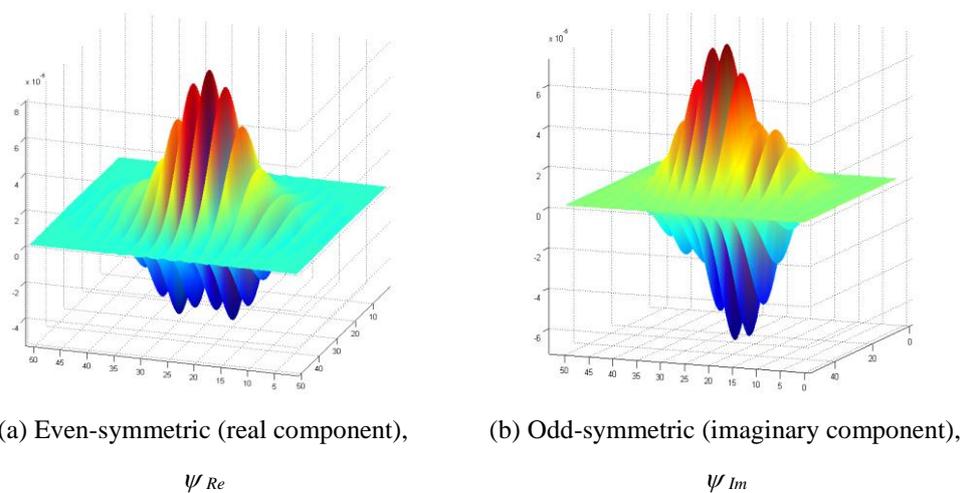


Figure 6. 6 Real and imaginary components of a Gabor filter kernel.

The spatial frequency domain and Cartesian spatial domain vectors are $\mathbf{k} = (u_0, v_0)^T$ and $\mathbf{x} = (x, y)^T$, respectively. Rotated vectors \mathbf{k}_r and \mathbf{x}_r are obtained by applying the rotation matrix, $\mathbf{R}(\ast)$, where,

$$\mathbf{R}(\ast) = \begin{bmatrix} \cos(\ast) & -\sin(\ast) \\ \sin(\ast) & \cos(\ast) \end{bmatrix} \quad (6.3)$$

hence, $\mathbf{k}_r = \mathbf{R}(\phi) \cdot \mathbf{k}$ and $\mathbf{x}_r = \mathbf{R}(\theta) \cdot \mathbf{x}$. By setting $\phi \equiv \theta$, the vector parallel to the wavefront direction is aligned with the corresponding σ_p component of standard deviation in the 2-D Gaussian. The normal component of the Gaussian envelope, σ_n is then aligned normally to the wavefront direction. Multiple filters, $\psi_{\{\theta(q), \omega(j)\}}$ can be derived from the template function to construct a filter bank. Individual filter outputs can be computed by convolving the filter kernel with correspondingly sized neighbourhoods of the input image, in the spatial domain. The convolution operation can use the real, ψ_{Re} or imaginary, ψ_{Im} components of the filter kernel or both. The output channels in the experimental implementation are summarised in table 6.4, where \ast is the convolution operator.

Output channel	Abbr.	Definition
Imaginary	I_{Im}	$I_{Im\{\theta(q), \omega(j)\}} = \text{Im}\{\psi_{\{\theta(q), \omega(j)\}} \ast I(x, y)\}$
Real	I_{Re}	$I_{Re\{\theta(q), \omega(j)\}} = \text{Re}\{\psi_{\{\theta(q), \omega(j)\}} \ast I(x, y)\}$
Norm (magnitude or energy)	I_{Mag}	$ I_{Mag\{\theta(q), \omega(j)\}} = I_{Re} + i I_{Im} $
Full rectification	I_{Rec}	$ I_{Rec\{\theta(q), \omega(j)\}} = I_{Re} + i I_{Im} $

Table 6.4 Output channel configuration

Channel responses as defined in table 6.4 can be used directly as features. Frequency progression (or frequency ratio) defines the location of and separation between individual filter centres on a normalised spatial frequency scale. Using J different

frequency centres, and the lowest frequency centre positioned at ω_0 , with $1/(J - 1)$ octave spacing, the frequency progression is,

$$\omega_j = \omega_0 2^{\frac{j}{J-1}} \quad (6.4)$$

The frequency bandwidths and angular bandwidths of filters in the spatial frequency (wave-number) domain (u, v) are functions of the envelope standard deviations, σ_p and σ_n in the spatial domain. Large spatial coverage (i.e., large σ_p, σ_n) in the spatial domain gives rise to narrower frequency bandwidths in the (u, v) domain. This is illustrated in figure 6.7. Response patterns of individual filters are circular when the Gaussian kernel is isotropic, i.e., $\sigma_p = \sigma_n$, as is used in the experiments in this study.

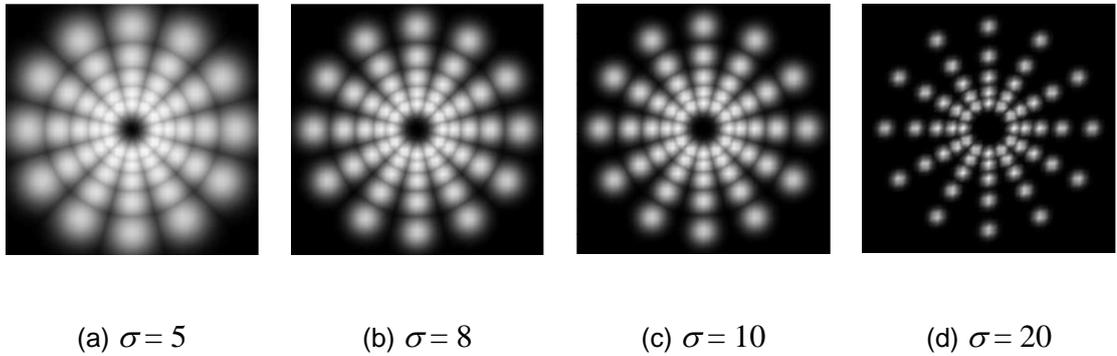


Fig. 6.7 Spatial-frequency (u, v) representation of prototype Gabor filter banks showing how the bandwidth of the filters changes with increasing σ .

The angular spacing in this illustration is $\Delta\theta = \pi/6$. Due to the rotational symmetry of the filter bank, channel responses generated by conjugates, $\psi^*\{\theta_q + \pi, \omega_j\}$ are duplicated. It is therefore unnecessary to implement the conjugates (shown to the left of the red line in figure 6.8) as they increase processing costs and produce redundant information.

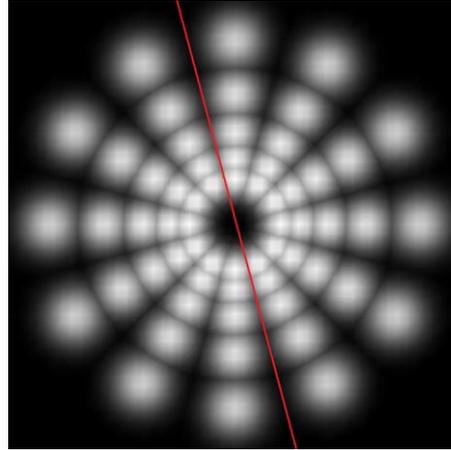


Fig. 6.8 Spatial-frequency (u, v) representation of a prototype Gabor filter bank. Only the filters $\psi\{\theta_q, \omega_j\}$ to the right of the red line are used. The conjugate filters $\psi^*\{\theta_q + \pi, \omega_j\}$ on the left have redundant orientations.

It is not necessary to apply the filter bank kernels at every pixel location in the input image, I (although this can be done if desired.) The filter bank can be used as a down-sampler in the feature creation process, by convolving the kernels with $N \times M$ contiguous, non-overlapping pixel neighbourhoods, centred on $I(n_\alpha, n_\beta)$, as shown in figure 6.9. The dimension of the feature space at this stage is $D = |\Theta||\Omega|$.

Using the feature creation kernel as a down-sampler has the two-fold advantage of greatly reducing the time required for pattern generation and reducing the size of the input space to the classifier. For example, using 25×25 pixel contiguous neighbourhoods reduces the pattern generation time and the size of the input space by a factor of 625, compared to generating pattern instances at each pixel location. Efficiency can be improved further by dimensionality reduction. One way of achieving this is by deriving rotationally invariant features from the orientation dependent output channels. Three well-known post-processing methods applied to the feature vectors after they have been generated are; (1) taking the mean response over all directions, (2) selecting the directional channel with the maximum response at each value of ω_j (Varma and Zisserman, 2005) and (3) applying a Discrete Fourier Transform (DFT) to the output channels (Bianconi *et al.*, 2008). The rotational configurations and the methods used to apply them are summarised in table 6.5.

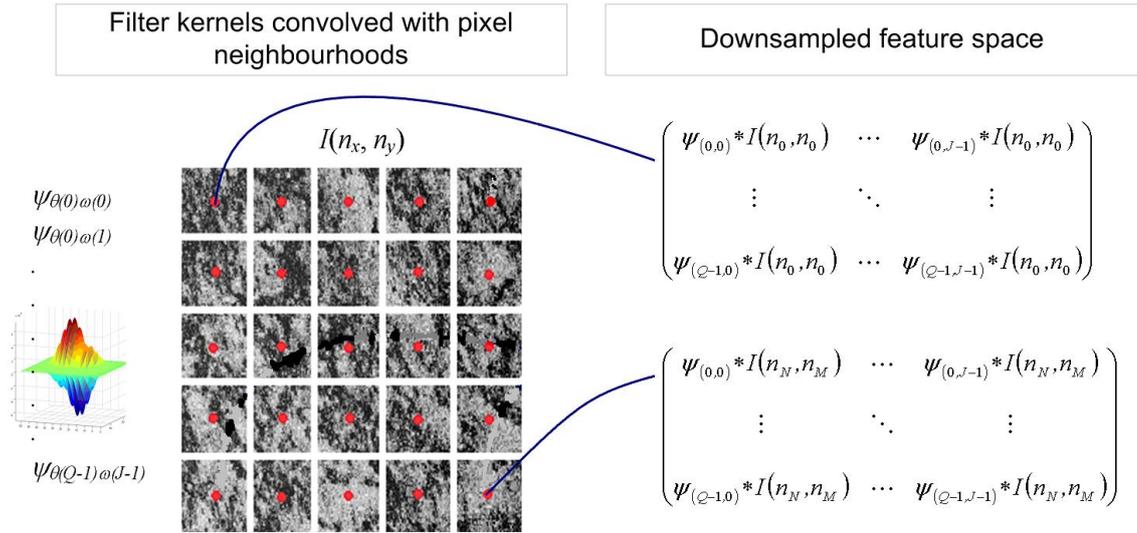


Figure 6.9 Illustration showing how the filter kernels can be used to create a down-sampled feature space from contiguous pixel neighbourhoods. Each matrix contains the patterns (filter channel responses) for the image neighbourhood.

Rotational configuration of output, Ψ^*	Definition	Dimension, D
Rotationally variant, Ψ^{RV}	Directional channels are unmodified	$ \Theta \Omega $
Rotationally invariant, maximum response, Ψ^{max}	$\Psi^{max} = \max\{ \Omega(\theta_0, \dots, \theta_Q) \}$	$ \Omega $
Rotationally invariant, mean response, Ψ^μ	$\Psi^\mu = \frac{1}{ \Theta } \sum_{q=1}^{ \Theta } I\{\theta_q, \Omega\}$	$ \Omega $
Rotationally invariant, DFT response, $\Psi^{(DFT)}$	$\Psi_\gamma^{(DFT)} = \sum_{q=0}^{Q-1} \psi_q e^{\{-2\pi i \gamma q/T\}}$ $\gamma = \{0, \dots, Q-1\}$	$(\Theta - 2) \Omega $

Table 6.5 summary of rotational configurations and methods

As well as reducing dimensionality, an additional advantage of using rotationally invariant features is that they tend to suppress some of the (undesirable) variability captured by the features, due to the effects imparted by sensor platform motion on the imagery (Kalcic and Bibee, 2004). The Dimension, D of the rotationally invariant feature space is given in the right column of table 6.5 as a product of the number of orientations $|\Theta|$ and frequency centres $|\Omega|$ in the filter bank. As an example, if there are 6 orientations and 5 frequencies, there will be 30 rotationally dependent output channels. Computing a mean or maximum directional response will reduce the dimension of the feature space to 5. Applying the DFT will produce a rotationally invariant feature space of dimension 20 as it represents the rotationally invariant features for each frequency band in terms of the Fourier spectral components. The investigation of Bianconi *et al.* (2008) found that the DFT approach performed better than some other methods (total energy, circular shift, brute force and no rotational invariance) in their texture classification experiments. Total energy is the sum of the energies of a specific frequency band over all orientations (dividing this by the number of orientations gives the mean response method.) They point out a limitation of the total energy approach is that it removes orientation dependent information. Further, they also point out that searching for the highest energy orientation (as in the circular shift or similarly, maximum energy approaches) may be prone to errors and sensitive to noise.

6.3.2 Experimental work

The purpose of these experiments is to investigate how different filter bank parameterisations influence the accuracy of binary discrimination of Sabellaria textures and to identify a useful set of filter parameters for this purpose, on the supplied imagery.

The supplied mosaic imagery has a resolution of 6 pixels/m and a radiometric resolution of 5 bits, i.e., 32 discrete grey values [0, 7,..247, 255], specific to the Coda GeoSurvey³ processing software used in the mosaicing process. Details of corrections applied and parameters used in the mosaic processing pipeline are unknown, as is often the case with legacy mosaic imagery.

³ <http://www.codaoctopus.com/coda-geosurvey/> [accessed 13-07-2011]

In the experiments, computational kernel sizes of 11, 17, and 23 pixels are used to generate the filter responses. These correspond to ground coverage of approximately 2, 3 and 4 m respectively. A Naïve Baye's classifier (see, for example, Duda *et al.*, 2001, p 13.) is used in a wrapper, with mean classification accuracy as the objective function for evaluating the parameters. As this series experiments differs from those in section 6.2 and 6.3.4 different processing flows and experimental harness architectures were implemented. The results from each harness were not compared in any detail, since the experimental campaigns, although closely related, were independent, with different objectives. So the fact that different learning algorithms were used in each harness is not relevant.

Multiple, balanced random samples are selected for training the classifier and the multiple models are then applied independently to the remaining samples to predict the class labels. The mean and standard deviation of the accuracies of the model predictions is computed for each run, over the value ranges in an eight-dimensional parameter space, comprising;

1. computational kernel size, k ,
2. initial frequency, ω ,
3. frequency progression (octave ratio),
4. angular separation, $\Delta\theta$,
5. Gaussian S.D, σ_n , (isotropic envelope),
6. channel configuration,
7. rotational configuration,
8. class discrimination task.

Output features are normalised to $[0 \ 1]$. Table 6.6 summarises the output channel and rotational configurations of the filter bank and the corresponding feature space dimensions, D .

Channel configuration	Rotational configuration			
	Ψ^{RV}	Ψ^μ	$\Psi^{(DFT)}$	Ψ^{max}
I_{Mag}	$ \Theta \Omega $	$ \Omega $	$(\Theta -2) \Omega $	$ \Omega $
I_{Re}	$ \Theta \Omega $	$ \Omega $	$(\Theta -2) \Omega $	$ \Omega $
I_{Im}	$ \Theta \Omega $	$ \Omega $	$(\Theta -2) \Omega $	$ \Omega $
I_{Rec}	$ \Theta \Omega $	$ \Omega $	$(\Theta -2) \Omega $	$ \Omega $

Table 6. 6 Summary of the feature space dimensions for output channel and rotational configurations.

The intrinsic filter bank parameters and their value ranges are summarised in table 6.7

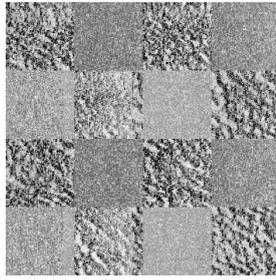
Parameter	Values
ω_0	0.05, 0.07, 0.09, 0.11, 0.13, 0.15
Octave ratio	1, 2, 4: $ \Omega = \{3, 5, 9\}$
Kernel size, k	11, 17, 23
$\sigma_n, \sigma_p \{ \sigma_n \equiv \sigma_p \}$	$\{1, 3, 5, 7, 9\}, \{1, 3, 5, 7, 9, 11, 13, 15\}, \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21\}$
$\Delta\theta$	$\pi/3, \pi/4, \pi/6, \pi/8$: $ \Theta = \{3, 4, 6, 8\}$

Table 6. 7 Summary of parameters and ranges of values considered.

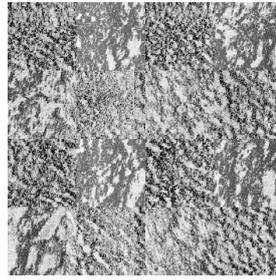
The binary discrimination tasks are;

1. Sabellaria-mixed sediments,
2. Sabellaria-Mussels,
3. Sabellaria-rock,
4. Sabellaria-sand,
5. Sabellaria-bedformed sand (i.e., sand dunes or sand waves.)

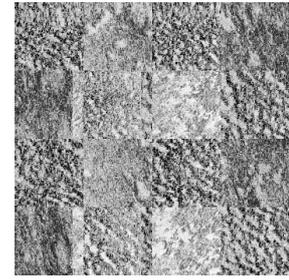
Examples of some of the synthesised test images used and the reference class map are shown in figure 6.10.



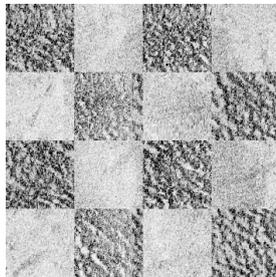
(a) Sabellaria-
mixed sediments



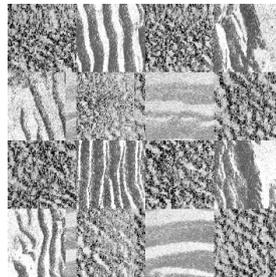
(b) Sabellaria-
Mussels



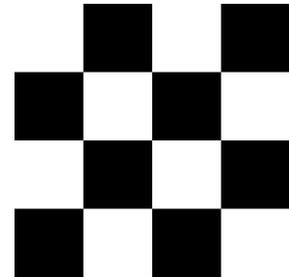
(c) Sabellaria-
Rock



(d) Sabellaria-
Sand



(e) Sabellaria-
Sand dunes



(f) class map: Sabellaria
(white), background
(black)

Figure 6.10 Image patches used for evaluation. The individual squares are sized at integer multiples of the feature kernel, to ensure class homogeneity under the feature kernel. The examples shown are for a kernel size, $k = 11$.

Under the controlled experimental conditions, using evaluation images synthesised from parts of a larger mosaic, it is possible to ensure that the feature kernels are applied only to image regions of the specified classes (regardless of whether those classes are homogeneous, correctly labelled or not). The image patches used in the test cases are sized as integer multiples of the kernel size, so there is no overlap onto tiles having a different class label. This does not remove the problem of within-class heterogeneity or incorrectly assigned class labels but these factors are beyond the control of the experiment. The five classification tasks considered in the evaluation reflect possible

real-world discrimination tasks. A common case would be the discrimination of Sabellaria from smooth sediments or bedformed sediments such as sand dunes.

6.3.3 Results and discussion

Several tens of thousands of experimental runs were carried out, so in order to focus on a meaningful analysis, only a small subset of the results are presented and discussed here. Results are considered from one channel configuration (magnitude, I_{Mag}), one rotational configuration (maximum response, Ψ^{max}) and at a single kernel size of $k = 17$.

Figure 6.11 shows how the mean classification accuracy (%) (y-axis) varies with the filter envelope standard deviation, σ (x-axis) on the Sabellaria-sand discrimination task for the 12 combinations of angular separation (number of orientations) and frequency progression (number of frequencies) of the filter bank. The line colour corresponds to the initial centre frequency (ω_0) in the frequency progression, as indicated in the legend. As, strictly these are wavenumbers rather than angular frequencies in time, the units are cycles per unit distance. Visual inspection of the figures shows that in many cases, maximum accuracy occurs near $\sigma = 9$. It can also be seen that the variability in the accuracies produced at different values of ω_0 decreases as the number of frequencies in the progression increases. For instance, the choice of ω_0 at $|\Theta| = 3$, $|\Omega| = 9$ is less critical than at $|\Theta| = 3$, $|\Omega| = 3$, where average accuracies can change by more than 20% depending on ω_0 . For a fixed dimension of feature space, $D = |\Omega|$ The results are clearly influenced by changing the number of orientations in the filter bank but there is no obvious systematic effect in this particular case.

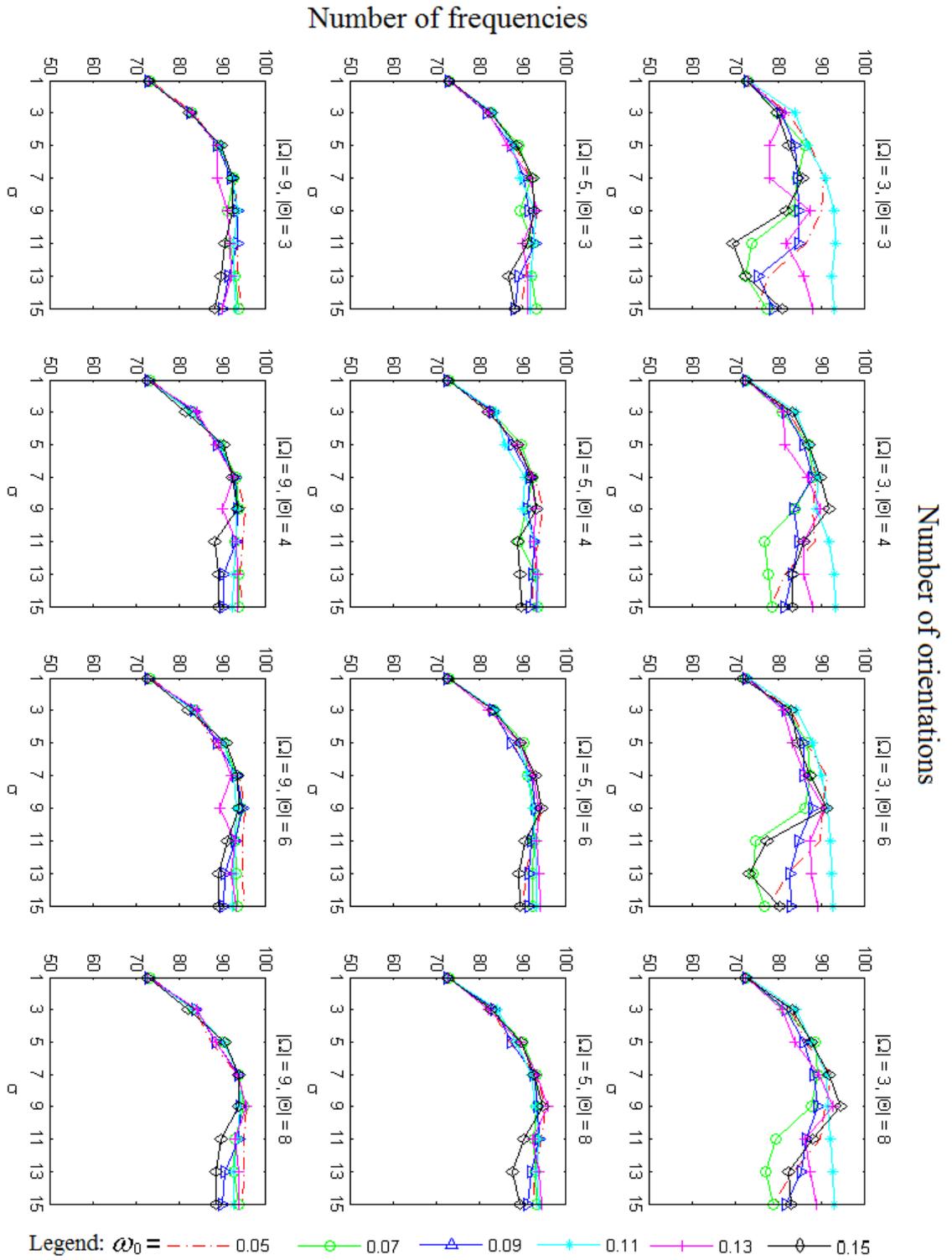


Figure 6.11 Sabellaria-sand discrimination accuracy (mean) over the range of initial frequencies, frequency progressions and angular separations of the filter bank.

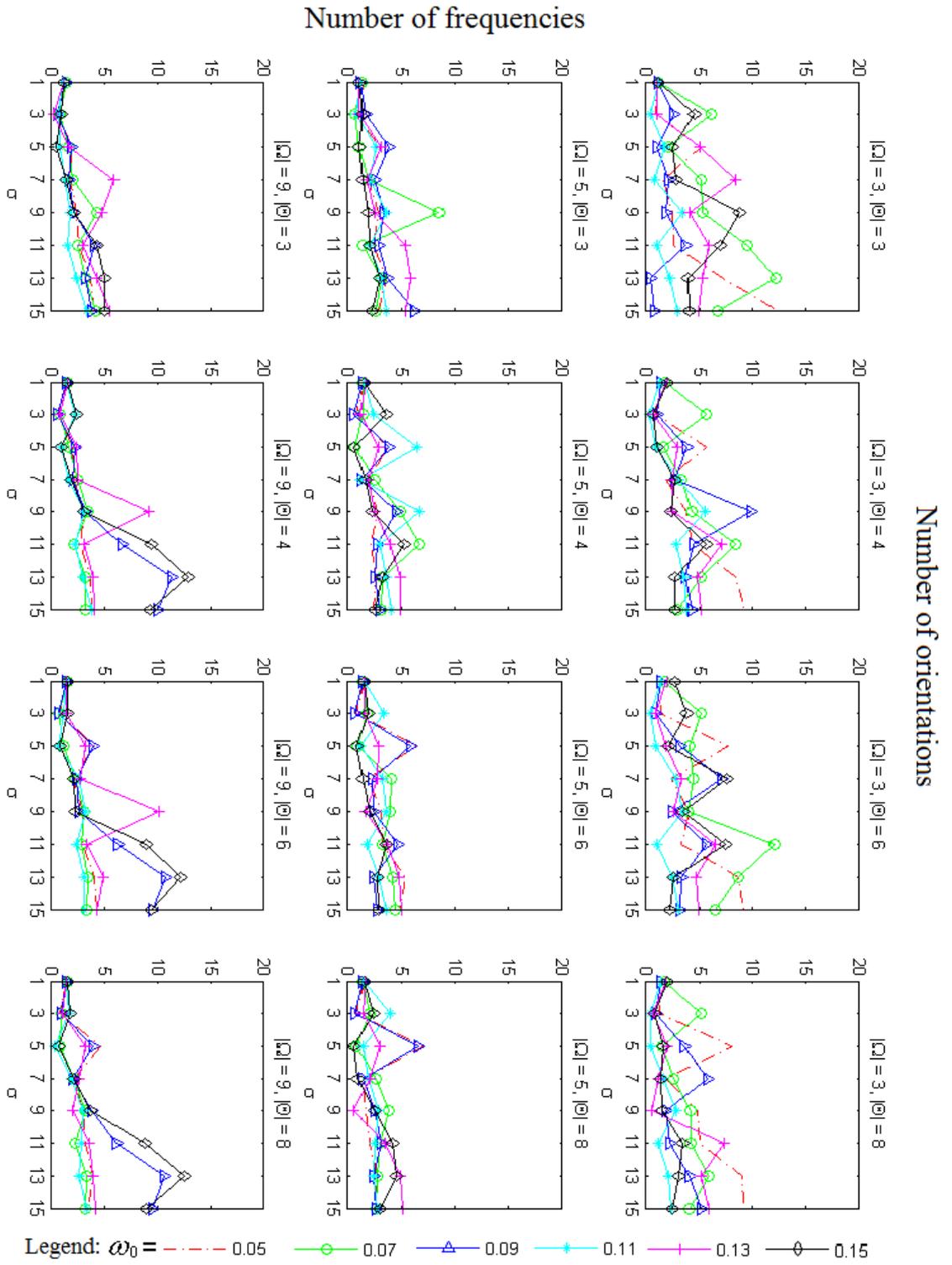


Figure 6.12 Sabellaria sand discrimination sd of accuracy for the corresponding graphs in figure 6.11.

Figure 6.12 shows the standard deviation of the accuracies in figure 6.11. With few exceptions, there is a general trend of an increase in sd with increasing σ . The sd at different values of ω_0 is greatest at $|\Omega| = 3$.

The value of σ , in all cases strongly influences the accuracy, with up to 25% difference in accuracy over the range of σ values, in each configuration. A similar affect was apparent on the other discrimination tasks. As this factor influences classification accuracy more than the other parameters, it is reasonable to first identify the (global) value of σ that produces the highest mean accuracies overall. The relative frequency of *maximum* mean accuracies attained over all class discrimination tasks, frequency progressions, angular separations and ω_0 values, is shown in figure 6.13. This result indicates that if $\sigma = 9$ is selected, the probability of attaining the maximal accuracy (irrespective of the classification task or other parameter settings) is 0.57, more than double that for $\sigma = 11$ and three times that for $\sigma = 7$. Interestingly, considering the results from other experiments (not reproduced here), the value of σ corresponding to the greatest probability of occurrence of maximum accuracies was found in general, at approximately half of the kernel size, i.e., $k/2$.

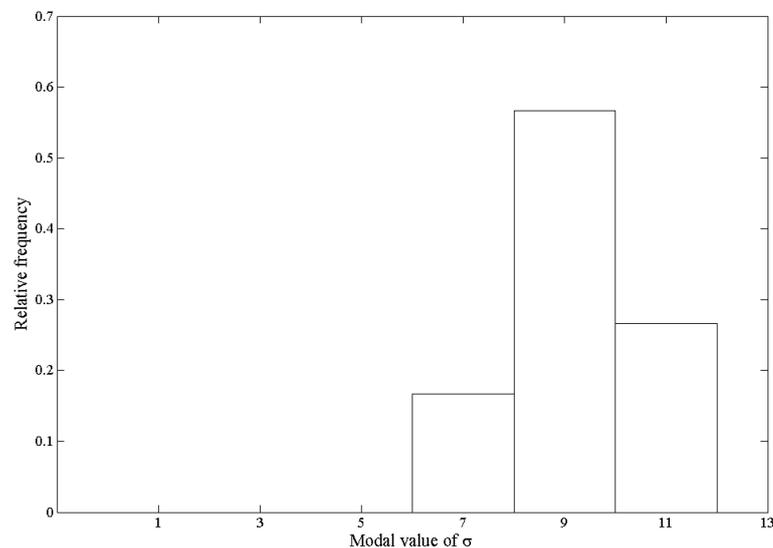


Figure 6.13 Relative frequency of filter envelope standard deviation (σ) producing the maximum accuracy, for $k = 17$, over all parameters and classification tasks.

The choice of ω_0 for half- and quarter-octave frequency progressions, with $\sigma < k/2$ is less critical, since the accuracy produced by using different ω_0 values varies by less than 5 % over the range of ω_0 .

Figure 6.14 shows the mean and SD (error bars) of classification accuracies using $\omega_0 = 0.15$, over all classification tasks. It is clear there is a consistent hierarchy in the level of difficulty of the different discrimination tasks across the various frequency progressions and angular spacings of the filter bank. The discrimination of Sabellaria from Mussels is the hardest task, followed by rock, sand dunes, mixed sediments and finally, the easiest discrimination task, with the highest accuracies, Sabellaria-sand.

The next decision, concerns choice of a suitable combination of frequency progression and angular separation for the scope of classification tasks.

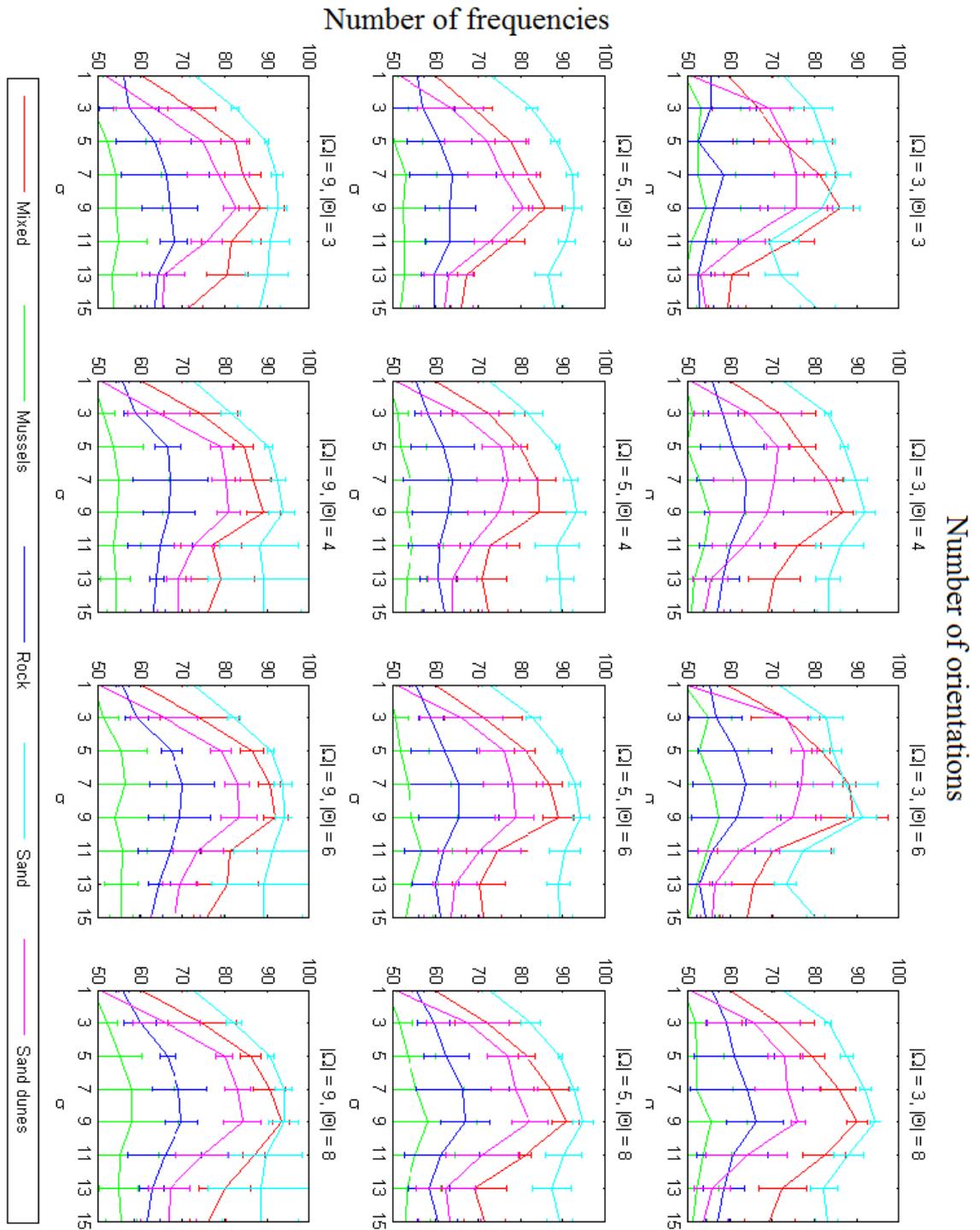


Figure 6.14 Mean and standard deviation of classification accuracy as a function of σ , at a kernel size $k = 17$, initial frequency, $\omega_0 = 0.15$, for all class discrimination tasks. Error bars indicate the accuracy sd.

Table 6.8 shows the mean (bold) and S.D. of the accuracies over all classification tasks, with the 12 different combinations of angular spacing and number of frequency channels. The results in table 6.8 are shown graphically in figure 6.15 (b). The mean

accuracy is greater (by at least 3.5 %) and the sd lower for the filter banks with half and quarter-octave frequency progressions, at all angular spacings. Highest mean accuracy is achieved using an angular spacing of $\pi/8$ and 9 frequency channels.

Frequencies	Number of filter orientations $ \Theta $			
$ \Omega $	3	4	6	8
3	72.2	71.5	74.7	76.0
	2.8	3.6	3.2	2.2
5	76.6	75.8	78.6	79.2
	0.9	2.1	1.4	0.5
9	77.0	76.5	78.3	79.7
	1.1	0.7	1.5	0.6

Table 6.8 Mean and S.D. of classification accuracies over all class discrimination tasks, with $\sigma = 9$, at a kernel size $k = 17$ and initial frequency, $\omega_0 = 0.15$.

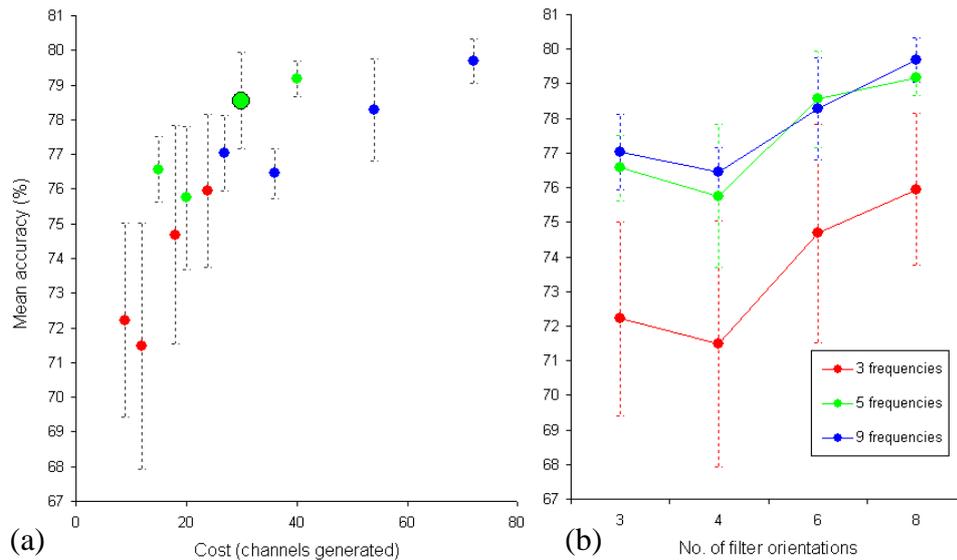


Figure 6.15 (a) Mean classification accuracy over all tasks, at $\sigma = 9$, $k = 17$, $\omega_0 = 0.15$, against cost (number of output channels generated) at different numbers of orientations and frequencies. (b) Mean overall accuracy as a function of octave ratio and number of orientations.

A very important further consideration is one of cost. When generating features from the input image, each orientation and centre frequency channel requires two separate convolutions with the real and imaginary kernels. If (all other things being equal) the cost of this operation is one unit, then the relationship between cost and accuracy can be

represented, as in figure 6.15 (a). This initial unit cost is fixed, regardless of how rotational invariance (and dimensionality reduction) is later achieved and the extent to which dimensionality is reduced, by post processing the output channels. The ratios of gains in accuracy are not of the same order as the gains in cost required to achieve the accuracy improvement. Halving the angular separation increases the cost by 100% but only increases the accuracy by about 5%. For this reason, the filter bank with an angular separation of $\pi/6$ and half-octave frequency progression is chosen (indicated by the larger green marker on fig. 6.15 (a).)

Rotational invariance by the maximum response is an effective method (Varma and Zisserman, 2005) and has a lower dimensional output compared to using the DFT. Good results are possible from the DFT (Bianconi *et al.*, 2008) but where cost is an important consideration, DFT is expensive to implement compared to max and the dimension of the reduced feature space, is always a factor of $(|\Theta|-2)$ greater, compared to using the mean or max invariance methods. Table 6.9 summarises the suggested parameters that are generally useful for discriminating Sabellaria from other textures in this specific test case on the mosaic data.

Parameter (configuration)	Value (description)
ω_0	0.10-0.15
Octave ratio	2 (half-octave progression)
Kernel size, k	≥ 17
$\sigma_n, \sigma_p \{ \sigma_n \equiv \sigma_p \}$	~ 9 (or approx. $k/2$)
$\Delta\theta$	$\pi/6$ (six orientations)
Channel configuration	I_{Mag} (magnitude response)
Rotational configuration	Ψ^{max} (maximum response)

Table 6.9 Summary of parameters and configurations useful for discriminating Sabellaria from other textures in the mosaic data.

The mosaic imagery regions under test had a ground resolution, $R_G = 6$ pixels per metre. It can be speculated then, for the creation filter bank features suitable for discriminating

Sabellaria, the filter kernel size, $k \approx 3R_G$ and $\sigma \approx 1.5R_G$ may in general, provide good results.

In the next section, some further results from an extension of this investigation are presented and discussed.

6.3.4 Further results and considerations

In a practical application scenario, it is conceivable, that presented with a mosaic image on a computer screen, a human operator could manually identify potential target regions of Sabellaria (S) and non target regions ($\neg S$) then select a few training points to induce a supervised classification model for Sabellaria discrimination. This is of course, similar to the approach used in the previous chapter, for pockmark discrimination. To test the feasibility in this context, a number of composite textural images were constructed using image patches from different regions of the mosaic. An example of a composite is shown in figure 6.16 (a) and the classes are described in table 6.10. The corresponding regional class map is shown in figure 6.16 (b) with the white regions representing the target, S class.

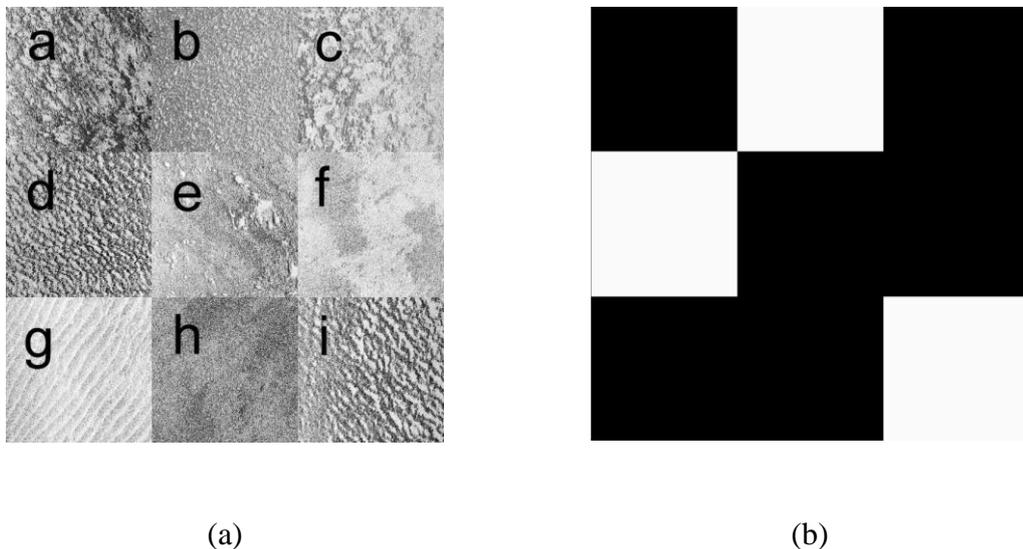


Figure 6.16 (a) Example of a composite image used in this test case. See table 6.10 for a description of the ground truthed textural classes. (b) Regional class map. The target S class is white.

Image key	Description of specific class	Generic (binary) class label
d	Sabellaria (moderate Sabellaria growth) with low-moderate elevation (28cm), patchy cover (up to approx. 75 %.)	<i>S</i>
b	Sabellaria (patchy Sabellaria aggregations) on a mixed substrate, with significant damage and broken tubes.	<i>S</i>
i	Sabellaria (patchy Sabellaria) on a mixed substrate, with significant damage and broken tubes.	<i>S</i>
e	Mixed coarse, shelly substrata with occasional cobbles and mudstone.	$\rightarrow S$
f	Sandy substrate with some gravel and cobbles.	$\rightarrow S$
a	Dense Mussel bed on sand, at least 75 % cover.	$\rightarrow S$
c	Rock outcrop (possibly chalk) with some overlying sand, sediments and epifauna.	$\rightarrow S$
g	Sand, Linear dunes, wavelength approx. $\lambda = 3$ m, with fine surface sediments, gravel and shell.	$\rightarrow S$
h	Lumps of clay and mudstone on mixed substrata.	$\rightarrow S$

Table 6.10 Descriptions of the ground truthed textural classes in the composite image (fig 6.16.)

Three of the textural regions are *S* class, the remaining 6 are from a variety of $\rightarrow S$ textural classes. Two of the *S* texture regions have intentionally been chosen to be similar (d, i) and another (b) which is different, as this reflects more realistically, the natural intra-class variability encountered. The problem here is more challenging due to the heterogeneous, $\rightarrow S$ textures from which the *S* class has to be discriminated.

A large number of experiments were again carried out using the parameter ranges identified in the previous sub section (table 6.9.) with some additional parameter values for σ and k , using a larger ($k = 25$, i.e., up to approximately 25 m^2 ground coverage) sized kernel. The fixed parameters/configurations used are; (1) half-octave frequency progression, (2) angular separation of $\pi/6$ (six orientations), (3) magnitude response output and (4) $\omega = 0.1$. Different kernel sizes ($k \geq 17$) were considered and as before, a circular Gaussian envelope was used in the filters ($\sigma_n \equiv \sigma_p$) and the affect of varying σ investigated.

Results discussed for the specific test case here were generated in a more expensive process, by convolving the filter kernels at each pixel location and using all the output channels without further dimensionality reduction, i.e., the filter bank is not rotationally invariant. Additionally, a 2-D Gaussian filter is applied in a post processing stage to each independent dimension of the pattern space (the channel sub-images) to smooth the magnitude output before model induction and subsequent classification. A Gaussian envelope with an sd approximately 1.5 times larger than the Gabor filter envelope, i.e 1.5σ was applied. In fact, overall, this experimental set up represents a more extreme end of the computational costs, compared to the harness in section 6.3.3. Training samples comprise multiple balanced sets of 30 instances selected from each generic class in the synthetic image. As in the case study of chapter 4, the BVM was used as the core learning algorithm, with the same default settings. As stated previously, the fact that a different base learner is used in the different harnesses is not relevant as no detailed comparisons of results were made between the harnesses, since the experimental objectives were different. See also, the evaluation section (6.5.1) where some of the implementation and coding issues are outlined.

Classification accuracy is measured using mean values of sensitivity (Se_μ) for the S class and specificity (Sp_μ), for the $\neg S$ class. Mean accuracy (Acc_μ) and error rates (ER_μ) are also used. The sd for each metric is abbreviated as; sensitivity (Se_σ), specificity (Sp_σ), accuracy (Acc_σ) and error rate (ER_σ). Regional classification is a function of lower (pixel)-level, binary classification outcomes in the textural class regions. A 50 % accuracy threshold criterion is used to assign the overall class label for the texture region.

Results in table 6.11 show the variation in classification accuracy at the pixel level, within the regions (a,...,i) with $\sigma \in \{5, 8, 10, 20\}$ and $k = 25$. The Sabellaria texture regions are indicated with an (S) symbol, e.g., b(S). Results in the Sabellaria texture rows represent sensitivity (the proportion of pixels correctly classified as belonging to the S class) and those in the other rows, specificity, i.e, the proportion of pixels in a $\rightarrow S$ region correctly classified as belonging to a $\rightarrow S$ class. The values in table 6.11 can thus be thought of as relative frequencies, an estimate of the probability that a specific texture region (a,...,i) is either S or $\rightarrow S$ class, e.g., for $\sigma = 10$, $p(a = \rightarrow S) = 0.74$ and $p(d(S) = S) = 0.84$.

Region	$\sigma = 5$	$\sigma = 8$	$\sigma = 10$	$\sigma = 20$
a	65.1	72.5	74.0	53.3
b(S)	79.0	81.9	72.6	63.1
c	68.2	72.7	76.2	80.7
d(S)	92.8	89.8	83.8	88.6
e	70.0	80.1	84.7	89.4
f	89.8	94.6	95.7	93.8
g	95.1	99.1	99.4	92.5
h	76.9	95.0	90.5	87.9
i(S)	79.5	76.0	68.1	76.1

Table 6.11 classification accuracy at the pixel level for the specific texture regions (a,...,i) as a function of σ , with $k = 25$. The region key is the same as in table 6.10 and the symbol (S) identifies regions, e.g., b(S), of the target, Sabellaria class.

The results of table 6.11 are shown graphically in figure 6.17 and the pixel-level binary classifications of the test composite in figure 6.18. The grey dashed curve connecting the lowest regional accuracies in fig 6.17 represents the maximum estimated certainty threshold at which all of the regions can be correctly assigned a generic S or $\rightarrow S$ class label. Using a 50 % accuracy threshold, the resultant generic class labelling of each region for the different values of σ is summarised in table 6.12.

Region	$\sigma = 5$	$\sigma = 8$	$\sigma = 10$	$\sigma = 20$	True class
a	$\neg S$	$\neg S$	$\neg S$	$\neg S$	$\neg S$
b(S)	S	S	S	S	S
c	$\neg S$	$\neg S$	$\neg S$	$\neg S$	$\neg S$
d(S)	S	S	S	S	S
e	$\neg S$	$\neg S$	$\neg S$	$\neg S$	$\neg S$
f	$\neg S$	$\neg S$	$\neg S$	$\neg S$	$\neg S$
g	$\neg S$	$\neg S$	$\neg S$	$\neg S$	$\neg S$
h	$\neg S$	$\neg S$	$\neg S$	$\neg S$	$\neg S$
i(S)	S	S	S	S	S

Table 6.12 Predicted generic regional class labels using a threshold of ≥ 50 % correctly labelled pixels in each texture region. The region key is the same as in table 6.10 and the symbol (S) identifies regions, e.g., b(S), of the target, Sabellaria class.

Using the 50% threshold, all of the regions are correctly assigned the generic class label, regardless of the value of σ , it is 100% accurate. Clearly, as was shown in the previous section, the classification accuracies are affected by changing σ , due to the response of the filters to the textures at different bandwidths. As can be seen in figure 6.17, the highest regional certainty threshold that could be applied, in order for all regions to be correctly labelled, has a peak near to $\sigma = 8$ at 72.5 %.

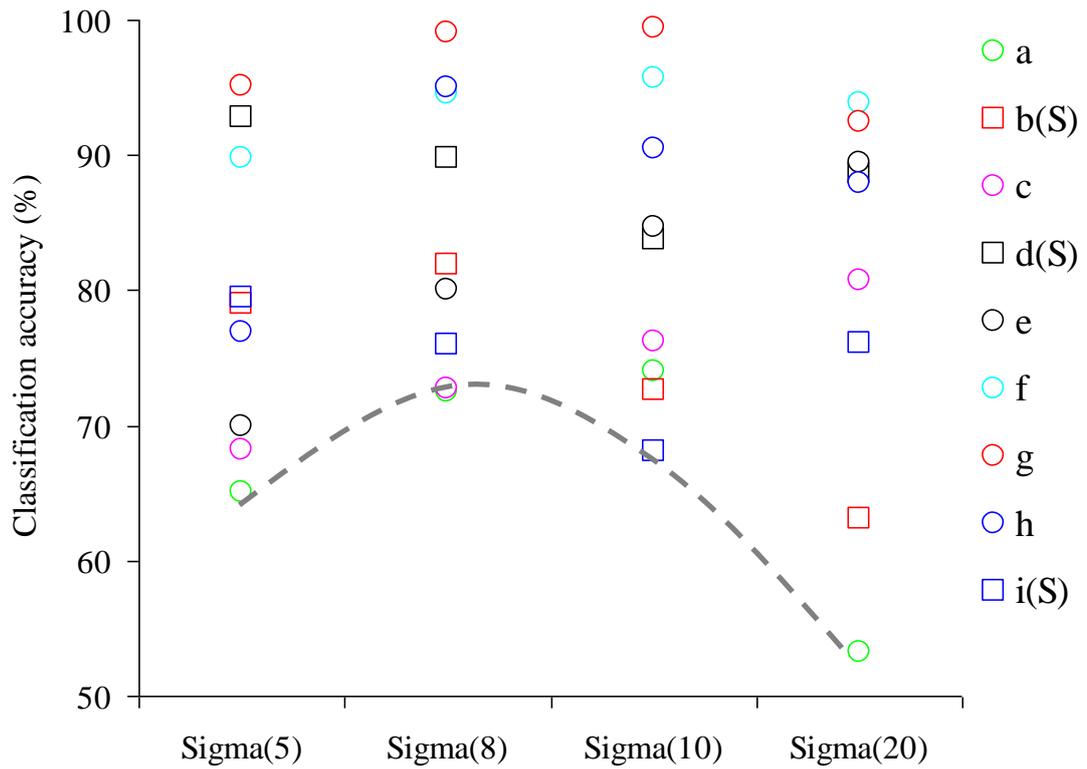


Figure 6.17 results of table 6.11 showing the variation in pixel-level classification accuracy with changing σ . The grey dashed curve represents the maximum certainty (threshold) with which the regional classification can be applied.

With a 50% certainty threshold, in this case, all other things being equal, the choice of σ is irrelevant to the final generic regional classification ($C_{\text{FINAL}} = 100\%$) since the chosen configuration and parameterisation of the filter bank will give rise to the correct labeling of all the regions as either S or $\neg S$ classes. However, with a 50% threshold, it may not be that certain if an individual region is correctly labeled, e.g. at $\sigma = 20$, $p(a = \neg S) = 0.53$, slightly better than a random guess. If a higher confidence is required in the results, e.g., $p(C_{\text{FINAL}} = 100\% \geq 0.70)$, this can only be achieved with $\sigma = 8$.

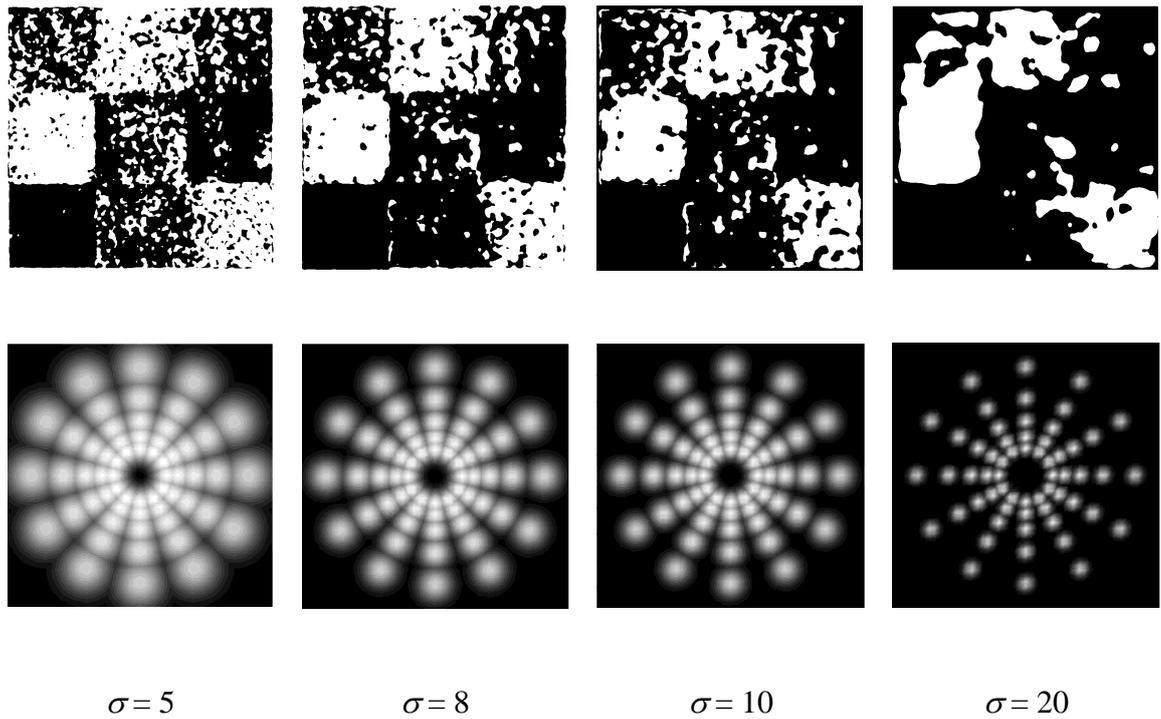


Figure 6.18 Visual results of the pixel level classification on the test synthetic image, for $\sigma \in \{5, 8, 10, 20\}$ and $k = 25$. The filter bank patterns in the wavenumber (u, v) space are also shown for comparison.

The final experiment concerns how the pixel-level classification accuracies are affected by variability in the classification models, induced with multiple training sets, sampled from different regions of the imagery. The textural classes and the filter bank parameters/configurations are as before, with $\sigma \in \{5, 8, 10\}$. Results are summarised in table 6.13. Clearly, the classification results from the filter bank features are resilient to data variability as the sd's for all metrics are less than 3.6 %. Highest sensitivity of 84.3 %, is achieved with $\sigma = 5$, but the best overall accuracy, 83.1 %, is obtained with $\sigma = 8$.

	Se_{μ}	Se_{σ}	Sp_{μ}	Sp_{σ}	Acc_{μ}	Acc_{σ}	ER_{μ}	ER_{σ}
$\sigma = 5$	84.3	3.2	79.1	3.6	80.9	1.5	19.1	1.5
$\sigma = 8$	83.4	2.9	82.9	3.1	83.1	1.5	16.9	1.5
$\sigma = 10$	77.0	2.9	84.0	3.1	81.7	1.4	18.3	1.4

Table 6.13 Mean and sd of pixel-level classification accuracy metrics, from different classification models induced with multiple training sets. $\sigma \in \{5, 8, 10, \}$ and $k = 25$.

6.3.5 Summary

Despite the importance of tuning to the success of a particular texture classification task, there is no unified approach to Gabor filter bank design (Bianconi and Fernández 2007). A heuristic approach was used in this section to investigate the efficacy of Gabor filter bank features for discriminating Sabellaria textures in sonar mosaic imagery. Various parameters and configurations were investigated on several discrimination tasks in different experimental spaces.

The research in section 6.3.3 facilitated the specification of a useful set of filter bank design parameters for discriminating Sabellaria textures from other seabed textural classes (table 6.9 is shown again below.)

Parameter (configuration)	Value (description)
ω_0	0.10-0.15
Octave ratio	2 (half-octave progression)
Kernel size, k	≥ 17
$\sigma_n, \sigma_p \{ \sigma_n \equiv \sigma_p \}$	~ 9 (or approx. $k/2$)
$\Delta\theta$	$\pi/6$ (six orientations)
Channel configuration	I_{Mag} (magnitude response)
Rotational configuration	Ψ^{max} (maximum response)

The overall classification process applied in this case was relatively efficient because (1) the filter bank was used as a down sampler in the feature creation process (2) rotational invariance (hence dimensionality reduction) was applied and (3) no post filtering was applied. This process and the configuration and parameters in table 6.9 typically gave rise to classification accuracies in the range of 75-80%.

In section 6.3.4 a more challenging task was considered, using similar texture classes, parameter ranges and configurations. Overall accuracies were higher than achieved in section 6.3.3, usually exceeding 80%, as shown in table 6.13. A major difference in the process though, for this improvement in accuracy, compared to the process used in

section 6.3.3 is the computational cost. The cost factors included (1) generating features on the kernel neighbourhood of each pixel, compared to downsampling (625 times more expensive for $k = 25$), (2) not applying rotational invariance (6 times more expensive, compared to using max or mean rotational invariance, due to their dimensionality reduction), (3) post filtering the sub-images with a Gaussian kernel, requiring another convolution operation on the kernel neighbourhood of each pixel on the sub-image produced by each channel of the filter bank.

Finally, it is speculated, for the creation filter bank features suitable for discriminating Sabellaria, the filter kernel size, $k \approx 3R_G$ and filter envelope sd $\sigma \approx 1.5R_G$ may in general, provide good results.

6.4 Other useful features

The results from both the waterfall and mosaic data using the Gabor filter bank are very promising and the researcher considers that filter banks are potentially the best method for Sabellaria texture discrimination. However, parameter tuning adds complexity to the feature creation process and cost is a further consideration. Certain low cost features with few or no parameters (e.g median, mean and standard deviation) already have a proven utility in sonar image processing. Parameter free features have a clear advantage in practical terms, as they do not require a saliency evaluation process prior to their application (strictly, all so-called parameter-free features have a single parameter, the kernel size.)

Three further types of features that have been considered for Sabellaria discrimination in this research work are; descriptive statistics, edge filters and lacunarity. Aside from the kernel size, the edge and lacunarity features have single threshold parameter. The purpose of this section is to outline some of properties of these features creation techniques and explain how they may potentially be useful for the task of Sabellaria (or other seabed texture) discrimination.

[1] Descriptive statistical features

Descriptive statistical features are simple, yet can also be some of the most useful features for classifying regions of sonar imagery. Information contained in the pixel intensity distributions in localised regions of the imagery underpins the process of discriminating between class regions, using descriptive statistics.

As pointed out earlier in chapters 2 and 3, there is sometimes a good correlation between sediment grain sizes and the mean backscatter intensity. It is also well established that hard surfaces such as rock are acoustically more reflective than say soft clay sediments. Therefore, it is possible to discriminate between different seabed regions using mean backscatter intensity alone. A clear benefit of using mean backscatter as a discriminatory feature is that since it can have a relationship to the physical character of the seabed, it facilitates a somewhat easier interpretation and understanding of the classification results in geological and habitat mapping context. Some commercial packages such as QTC Swatheview⁴ use mean backscatter intensity as one of the features in the feature vector.

The caveat with using simple statistics (and some other types of features) is that there will always be undesirable factors influencing the intensity of the sonar signal and the radiosity values of the mosaic image pixels. Usually the uncertainty in the radiosity value is far greater in sidescan than multibeam imagery. In the worst case, using mean backscatter as a feature may lead to a partitioning of regions with different user applied gain settings or seabed proximity to the transducer rather than any true differences in seabed properties. So the basic assumption when using these features is that the data have been prepared or calibrated in such a way to, as far as possible, suppress these undesirable radiometric artefacts.

A comparison of Sabellaria (fig 6.19 (d)) and gravel (fig 6.19 (e)) classes reveals some interesting differences. The Sabellaria class has a visually higher contrast, also shown by the intensity distributions in figure 6.20. The distribution for Sabellaria, is broader, whereas the lower contrast gravel class intensities are narrowly spread around the median value.

⁴ <http://www.questertangent.com/seabed-classification/seabed-classification-products/qtc-swathview-seabed-classification/> [accessed 06-07-2012]

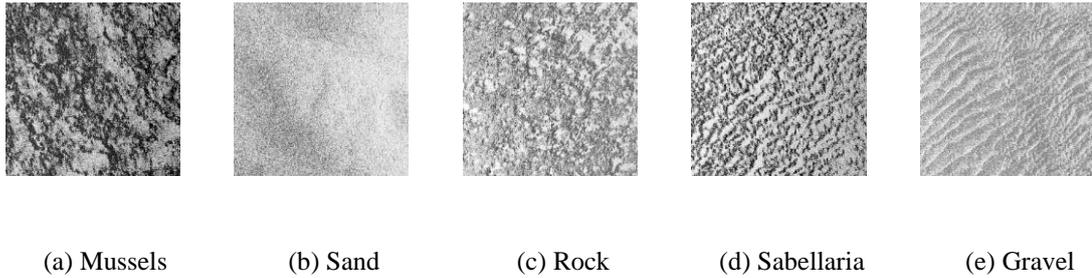


Figure 6.19 Mosaic regions showing some typical seabed types.

Mean intensity for the gravel is slightly higher than for Sabellaria, suggesting (in these samples) the gravel may have a comparatively higher acoustic reflectivity. Similar differences can be observed between the distributions for all of the classes. Clearly, there may also be some useful information about the local area of the seabed in the distribution shape itself.

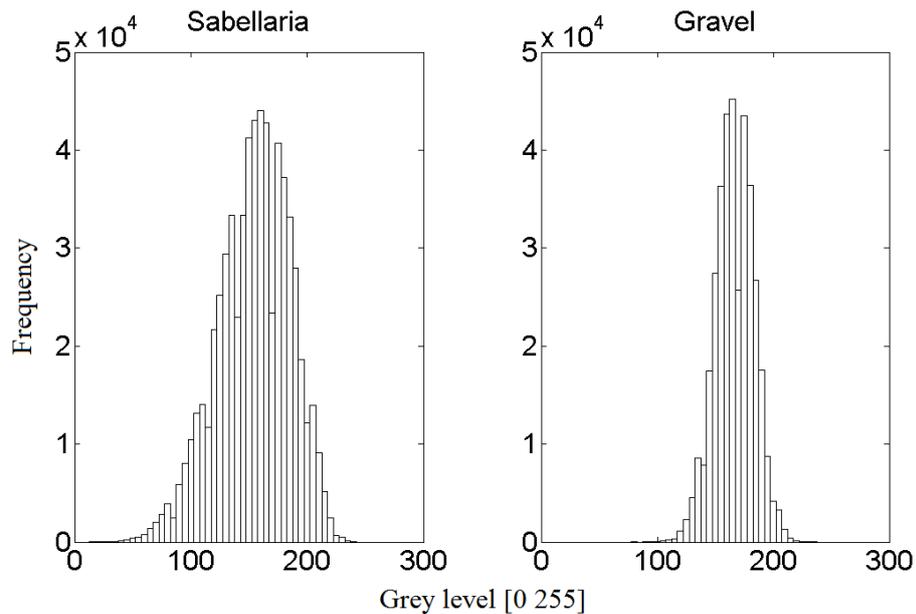


Figure 6.20 Intensity histograms of Sabellaria (shown in fig. 6.19 (d)) and Gravel (shown in fig. 6.19 (e)) classes.

In fact, estimating the parameters of a distribution model is another means of classifying a particular seabed class in calibrated sonar imagery, as described by de Moustier (2009) in Hughes-Clarke *et al.* (2009). There are many proposed distribution models though, such as Rayleigh-Rice, K -distribution, Log-Normal and Rayleigh Mixture, see for instance, Jackson and Richardson (2006), Dunlop (1997), Gensane (1989) and

Lyons and Abraham (1999). Histogram matching would be another, empirical approach, given class exemplars of intensity distributions, the similarity to the generated intensity histogram on the test sample could be measured and used to classify the test data.

Due to the lack of calibration and the noise and distortion usually present in qualitative sidescan mosaic imagery, there will be ambiguity in attempting to fit the image intensity distribution to a particular parametric distribution model. In this case, an empirical approach is preferable, to estimate the probability mass function. The (normalised) histogram bin values can then be used directly as individual feature values.

Stewart *et al.* (1994) found, in their analysis of sidescan data from the Juan de Fuca Ridge, that variation in mean area intensity together with textural variability is associated with different seafloor types. Johnson and Helferty (1990) also provide evidence of the power of a histogram to discriminate between regions of differing acoustic reflectivity. Outside of the sonar image classification domain, intensity histograms have been used in studies of medical imagery. For instance, Tweed and Miguet (2002) used a combination of histogram and textural features to detect regions of interest in mammograms.

[2] Edge filters

The response of a local pixel neighbourhood to an edge detector can be used to create features. Textured areas such as regions of rock, Mussels or Sabellaria have different quantities and spatial distributions of edges, compared to relatively smooth areas such as sands and sediments. Coverage of the theory of various edge detectors can be found in, for instance, Marr and Hildreth (1980) and Torre and Poggio (1986).

A Sobel mask (Sobel, 1990) was applied in the experimental investigations carried out for the thesis. It is a first derivative, directional edge detector and whilst easy to implement, it has the disadvantage of being sensitive to noise (Sharifi *et al.*, 2002.) Masking known noisy areas of the imagery is an assumed prerequisite to applying this method.

The directional gradient operator in the Sobel mask responds strongly to transition regions where there are large differences in the radiosity values of neighbouring pixels.

Detected edge pixels are highlighted by binarising the response of the image to the Sobel masks, at a predefined threshold parameter, t . Directional gradients, g_x and g_y are computed over the pixel neighbourhood, $n(x, y)$ by convolution with the Sobel masks S_x and S_y , respectively, i.e., $G_x = S_x * n(x, y)$ and $G_y = S_y * n(x, y)$. The strength of the detected edges in the neighbourhood, $G_n(x, y)$, is the magnitude of the directional derivatives, $G = (g_x^2 + g_y^2)^{0.5}$.

The resultant image is binarised, to obtain the binary image, $B_n(x, y)$ using the detection threshold parameter, $t > 0$ and the discrete unit Heaviside function,

$$H[B_n(x, y)] = \begin{cases} 0, & G_n(x, y) < t \\ 1, & G_n(x, y) \geq t \end{cases} \quad (6.5)$$

Changing t , produces different ratios of edge pixels to total pixels, r_E , in B_n . These ratios can be used as feature values. As t increases, r_E reduces at different rates, for different seabed classes.

Figure 6.21 shows the result of the first step in the feature creation process, applying the edge detector at various t values to the images in figure 6.19 to obtain B_n for each class. At the minimum value of (in this case) $t = 0.05$, in the top row, the SNR is low and even texturally smooth regions such as the sand, exhibit a large number of edge pixels. Increasing t to 0.15 (second row) has a filtering effect, with the more highly textured classes exhibiting a greater number of edge pixels. The filtering effect increases as t is increased further until the number of edge pixels approaches zero (i.e., $r_E \rightarrow 0$).

The next step in the feature creation process is to compute the ratio, r_E , of edge ('1' or 'white' pixels) to the total number of pixels in the neighbourhood, at the different threshold values. The plot in figure 6.22 shows how r_E varies with t for the five different classes (applied to the entire 250×250 pixel image sample.) For small t , ($t = 0.05$) it is clear that the low SNR does not produce a wide range of r_E values for the different classes. However, there appears to be an optimal domain for t that yields wide ranging, r_E values, potentially useful for discriminating between the classes. In this case values of t between 0.1 and 0.25 produce the greatest differences of r_E between the classes.

Some advantages of this feature creation approach are (1) it is rotationally invariant, (2) it is (mean) illumination invariant, as it is based on derivatives, (3) for relatively homogeneous textures, the results are consistent over a wide range of spatial scales, i.e., it is scale invariant, (4) the feature values are easily computed ratios and do not require normalisation. Examples of the results over different kernel sizes are shown in figure 6.23.

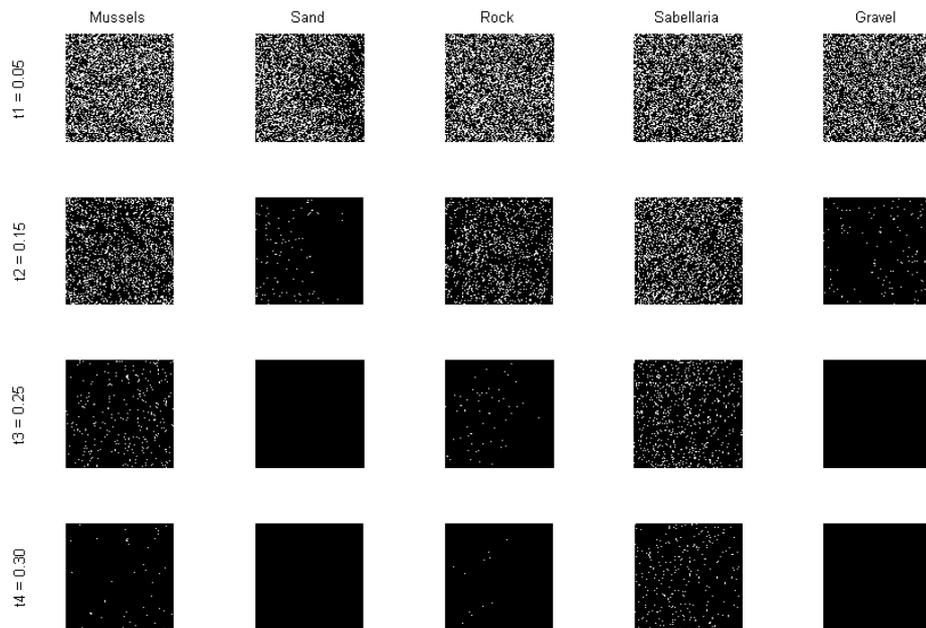


Figure 6.21 The result of applying a Sobel edge detection mask to the images in figure 6.19. The threshold value increases from top to bottom. Each column contains the class instances corresponding to the left to right layout in figure 6.19.

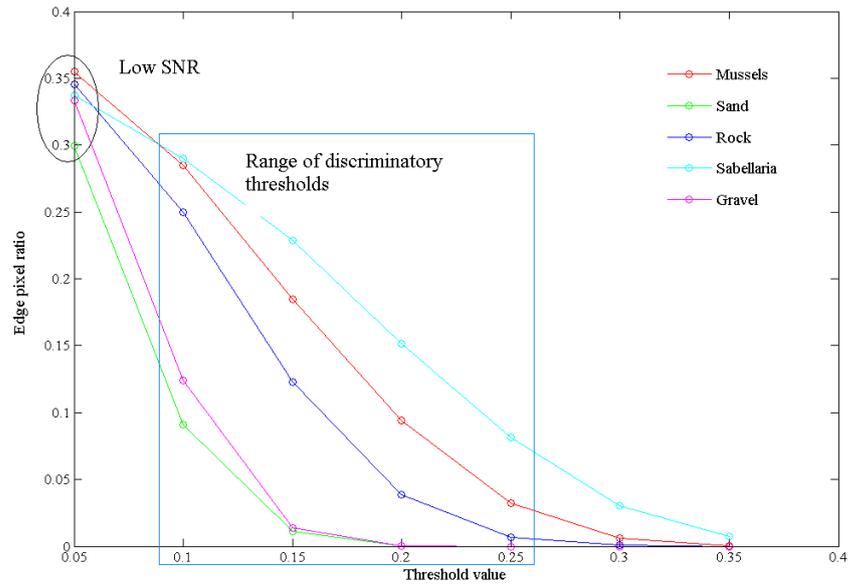


Figure 6.22 Variation in the ratio of edge pixels to total neighbourhood pixels, with threshold value for five different seabed classes. The ratio is computed on the entire 250×250 pixel image sample.

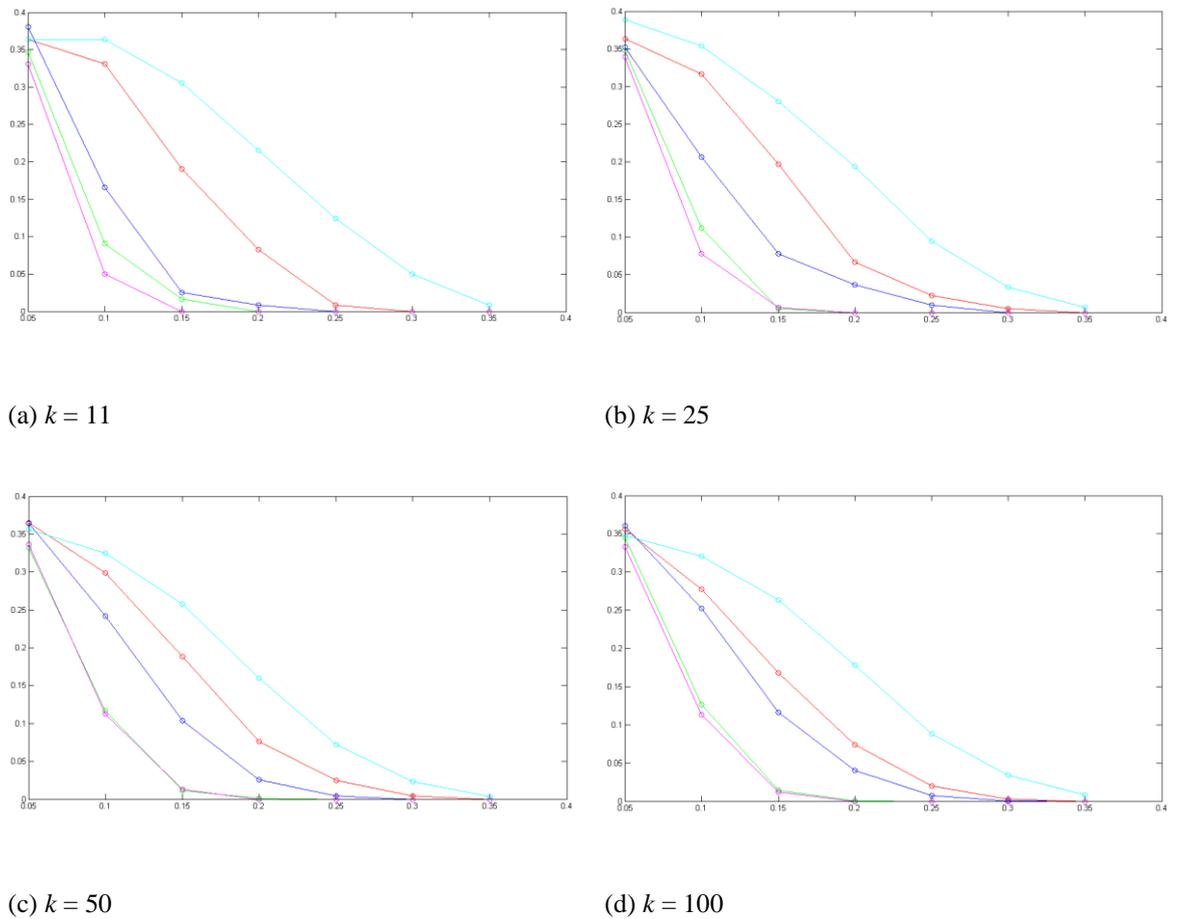


Figure 6.23 The profiles of the feature values are relatively consistent over an order of magnitude of kernel sizes.

[3] Lacunarity

Lacunarity measures the ‘gapiness’ or ‘hole-iness’ in a geometric structure (Kaye, 1989.) The concept of lacunarity, first introduced by Gefen *et al.* (1983), can be applied to the analysis of a binary image representing the distribution of targets of interest, against the background. In the case here, the targets of interest are taken as the edge pixels, generated as described in the preceding paragraphs. It is clear that the different seabed textures have different spatial arrangements of edge pixels, so the lacunarity of these arrangements could potentially be used as a discriminatory feature.

Lacunarity has been studied widely as a measure of natural textures, for instance, in satellite images of forest cover (Butson and King, 2006, Mahli and Román-Cuesta, 2008) and landscape textures (Plotnick *et al.*, 1993, McIntyre and Wiens, 2000.) According to Mahli and Román-Cuesta (2008), information about the forest structure, captured by lacunarity is difficult to interpret and the relationship of lacunarity to the real structures is not clear.

The natural distribution and clumpiness of certain types of forest canopy cover may in certain cases, resemble the distribution of natural targets, such as Sabellaria reefs or other habitats on the seabed. However, there are no published studies on using lacunarity analysis to discriminate between seabed textures or for analysing the distribution of natural targets such as reefs.

The commonly applied gliding box algorithm of Allain and Cloitre (1991) was implemented and is used to generate multi-resolution feature values within a specific kernel size. Some examples of the visual classification results using edge filter features and lacunarity features are presented in case study 4 (chapter 8.)

6.5 Evaluation, conclusions, recommendations and scope for further work

6.5.1 Evaluation

The aim of this study was to establish the tractability of a machine approach to the novel tasks of discriminating Sabellaria textures in sidescan waterfall and mosaic imagery. It was the first study of its kind to approach this important interpretative problem. The first objective was to identify the most promising feature creation methods on waterfall imagery and the second, to investigate one of these methods in more detail on mosaic imagery, identifying any salient, method specific feature configurations or parameterisations. Features created from the signal processing methods were generally found to be the best discriminators on the waterfall imagery, the top two being different configurations of a Gabor filter bank. Research work on the waterfall imagery features resulted in a peer reviewed conference publication. Configurations of the Gabor filter bank and filter parameters were considered in much greater detail for discriminating Sabellaria textures from other textural classes in mosaic imagery. Cost of the configuration (in terms of the number of feature channels created) was also considered as a constraint on the design. The aim and objectives have been successfully achieved, since discriminatory filter bank configurations and filter parameter ranges for the task were identified.

The main contribution and novelty of this work is in the task of machine discrimination of Sabellaria textures - it is the first study of its kind to address this problem. Identifying the Gabor filter bank as a suitable method for feature creation and further establishing a filter bank configuration and filter parameterisation is an important step in confirming the feasibility of automating the Sabellaria discrimination task. Relatively few studies have considered Gabor filter bank features for classification of sonar textures, so the work also contributes more generally to the domain specific application scope of filter bank features and the use of texture surrogates as a subspace for solving seabed classification problems.

Three different experimental harnesses were designed and implemented and these had variations in their architecture, according to the experimental objectives. The feature

creation and evaluation harnesses used in section 6.3.4 was developed primarily using C++ and comprised a collection of dynamic link libraries (dll's) with processes between the dll's controlled through a bespoke Visual Basic user interface. MS-DOS batch files were used for reading and writing intermediate data and for invoking other executables. It became apparent that this approach would be too inflexible for the experimental campaign as a whole, since, as work on the harness progressed and experiments were run, it was clear that there would be further experimental dimensions required in the enquiry. Facilitating these investigations computationally within the harness would have introduced unnecessary technical burdens especially in expanding the experimental space and in attempting to deploy heterogeneous software on the high performance computing (HPC) facilities at the UEA. However, the harness fully served its purpose in generating the results for section 6.3.4. A decision was made to revise the architecture and design and implement a new harness in Matlab. The Matlab harness was used extensively for results generation in section 6.3.3. Numerous function files were implemented and these could be deployed relatively easily on the HPC. The Matlab harness used for generating the results in section 6.2.3 was kindly made available during the collaborative aspect of this case study (section 6.2), by Professor Francesco Bianconi, of the University of Perugia, Italy.

6.5.2 Conclusions

An investigation on samples of waterfall imagery revealed that several types of feature creation methods could be useful for discriminating Sabellaria textures from other seabed texture types. The Gabor filter bank, DT-CWT, ILBP and BGC were some of the strongest performing methods in this respect.

The Gabor filter bank was chosen for further investigation regarding the discrimination of Sabellaria textures in mosaic imagery. The following parameter settings provided reasonable overall results on the data under test, using an isotropic Gaussian kernel:

Parameter (configuration)	Value (description)
ω	0.10-0.15
Octave ratio	2 (half-octave progression)
Kernel size, k	≥ 17
$\sigma_n, \sigma_p \{ \sigma_n, \equiv \sigma_p \}$	~ 9 (or approx. $k/2$)
$\Delta\theta$	$\pi/6$ (six orientations)
Channel configuration	I_{Mag} (magnitude response)
Rotational configuration	Ψ^{max} (maximum response)

The process was relatively inexpensive, as the filter bank was used to down sample the input imagery in the feature creation process. It was speculated, for the creation of filter bank features suitable for discriminating Sabellaria textures at a ground resolution in the image of, R_G pixels per metre, the filter kernel size, $k \approx 3R_G$ and filter envelope sd $\sigma \approx 1.5R_G$ may in general, provide good results.

A further experimental investigation using a more expensive process also yielded very good classification results. However, only a slight improvement in accuracy was achieved at considerable extra computational cost.

Evidence from the investigations presented in this chapter provides a strong indication that discrimination of Sabellaria textures is indeed tractable. A useful set of parameters and configurations were established, also taking into consideration the cost of generating the features. The aim and objectives for this case study have therefore been successfully achieved.

6.5.3 Recommendations

For discriminating Sabellaria target textures from other texture classes in sonar waterfall and mosaic imagery, the filter bank should be used and configured as specified in the conclusions. To achieve maximum efficiency, the filter bank should be applied as a downsampler in the feature generation process. Further efficiency improvements may be gained through a more efficient spatial domain implementation of the filter bank,

e.g., using the separability properties, as in Amayeh *et al.*, (2009). For maximum classification accuracy and resolution (at the cost of reduced computational efficiency) the filter bank features should be generated at each pixel location and a Gaussian post filter applied to the feature space sub-images. For ease of implementation, Matlab is recommended and the Naïve Bayes classifier from the Matlab library can be used as a core learning algorithm, since it can also output probabilistic support (certainty) for the classes, if desired. It is also recommended that a number of different experiments are designed to evaluate the features.

6.5.4 Scope for further work

- There are already established optical imaging benchmarking databases, such as Outex⁵ and CURET⁶ and in principle, there is no reason why a similar benchmarking database for sonar imagery cannot be established. Perhaps one of the single most important means of advancing the study of seabed texture discrimination and classification would be to build a database of ground-truthed sonar images. The database could contain imagery of different sediments, bedforms, habitats, biogenic and geological structures and so on. Imagery could be captured by a variety of industry-standard instruments under various environmental conditions, operating frequencies and resolutions. The database would then be made freely available to researchers around the world.
- It would be very useful to evaluate the Gabor filter bank features (and others too) on larger sized, real-world imagery, capturing wider coverage of Sabellaria colonies, their natural variability, distributions and boundaries with the surroundings. In order for this to be meaningful, good quality data are required, with (preferably dense) ground-truth and multiple human interpretations.

In carrying out this work, it became clear that there had been few detailed studies concerning methods for evaluating features and feature parameters in the context of sonar image texture classification tasks. Consequently, the case study in chapter 7 investigates the problem of feature evaluation and the robustness of feature evaluation methods.

⁵ <http://www.outex oulu.fi/> [accessed 07-07-2012].

⁶ <http://www.cs.columbia.edu/CAVE/software/curet/> [accessed 07-07-2012].

Novel consensus approaches

Chapter 7

to the reliable ranking of features

for seabed imagery classification

The work presented in this chapter resulted in the publication of a paper, “Novel consensus approaches to the reliable ranking of features for seabed imagery classification,” by Harrison *et al.* (2012).

Contents

7.1 Introduction

7.2 Related work

7.2.1 Domain specific feature selection

7.2.2 Robustness of feature selection methods

7.3 Consensus feature ranking methods

7.3.1 Nonparametric distribution estimation

7.3.2 Individual distance measures

7.3.3 Consensus ranking of features

7.3.4 Reliability weighting of permutations

7.4 Application context

7.4.1 GLCM feature creation

7.4.2 Experimental data

7.5 Experimental results and discussion

7.5.1 Investigating saliency and classification accuracy

7.5.2 Defining a baseline rank

7.5.3 Investigating task robustness

7.6 Evaluation, conclusions, recommendations and scope for further work

7.6.1 Evaluation

7.6.2 Conclusions

7.6.3 Recommendations

7.6.4 Scope for further work

7.1 Introduction

The previous chapter considered various feature creation methods for the discrimination of textures in mosaic and waterfall imagery. In doing this work, it became clear that there had not been much research concerning methods for evaluating and selecting features for machine learning applications in the sonar image classification domain or proposals of new methods for this purpose. Further, the more generic concept of robustness of a feature evaluation method, whilst important, has received very little attention in the published literature. The aim of this chapter then, is to propose a novel framework for evaluating distance measures, for the purpose of evaluating and ranking parametric features on sonar imagery. It is more abstract than the previous chapter in the sense that the focus is on the method used to evaluate the features rather than the features themselves. The objectives are to establish which individual distance measures are most reliable for feature evaluation and ranking and to determine if there is any advantage in using a committee of distance measures for this purpose. Methods are compared using the correlation of estimated saliency to classification accuracy and the correlation of a feature ranking to a baseline ranking, together with the variability in these properties.

A fundamental process in machine learning for a specific data-driven task is the induction of a model from data in a chosen feature space. Selecting a small set of discriminatory features to construct the feature space improves computational efficiency and often results in a more accurate and simpler model for capturing the intrinsic properties of the problem. However, estimating the saliency (i.e. discriminative potential) of features, ranking them and selecting a subset for model generation is a non-trivial task.

The main paradigms for feature ranking and selection are the filter-based, wrapper-based and embedded approaches, as described in, for instance, Guyon and Elisseeff (2003), and Liu Motoda (2008). It is well-known that wrapper-based approaches are model-dependent and computationally expensive (Kohavi and John 1997). Whilst they can in certain cases (e.g Patrinos *et al.* 2010) produce better rankings and selections, their high computational cost is a limiting factor in their applications. Further, as the objective function is based on minimising the error-rate of the classifier, they can lead to model over fitting (Kohavi and Summerfield 1995.)

Embedded feature selection is an integral sub-component of a dedicated classifier system. There are many embedded variants, a well-known example being Random Forest (RF) (Breiman, 2001.) The RF algorithm selects multiple random subsets of features that are used to induce different decision tree (DT) models, to construct a DT ensemble.

In contrast, model-independent filters do not require any inductive learning algorithm and are consequently more efficient. Filters are usually based on correlation measures (sometimes called dependence or association measures), information (or uncertainty) measures and distance (separability, discrimination or divergence) measures. Distance measures are simple and effective methods for quantifying the class separability of individual features. They are preferred in real world applications where the features and their parameterisations need to be evaluated quickly and the size of the parameter space is large, such as in classifying acoustic imagery of the seabed.

There are though, dozens of distance measures to choose from, as in Cha (2007) yet no universal criteria for selection of a particular measure for a given type of data or application. Due to their definitions, different measures generate different numerical values of distance between estimated class probability density functions (pdf). After assigning rank scores to the features, the score permutations from different measures often do not coincide. Additionally, the ranking of features produced by any individual measure may be unstable if small changes in the data sampled for feature evaluation produces different rankings. This creates difficulty, for example, when selecting the “*top-n*” features, if combinations of the n -components differ, it is not clear which will produce the most discriminatory feature space.

These problems are amplified when data are noisy (which is always the case with sonar imagery), features are parameter-dependent and their saliency is sensitive to changes in the parameter space. In a case like this, it is imperative to be able to evaluate and select the best parameters and features, accurately and consistently. A reliable and efficient evaluation and selection approach is thus needed which is robust across a range of parameters and domain classification tasks and stable in the sense that the results are unaffected by small changes in the sampled data.

This chapter presents two novel, filter-based methods; an unweighted consensus and a weighted consensus, for estimating the saliency of individual features and ranking them. The consensus results and those of the constituent individual distance measures are compared by considering measured saliency in relation to classification accuracy. Correlation of candidate rank score permutations with baseline rankings established on validation data is used as a means of evaluating the robustness to changes in parameters and classification tasks. The methods are tested using Grey Level Co-occurrence Matrix (GLCM) features (Haralick *et al.* 1973) created on sonar mosaic and waterfall imagery of the seabed and for further comparison, on a selection of textures from the Outex database¹. GLCM features are probably the most widely used textural feature generation methods in the sonar image processing domain and can be found in several sonar image classification packages, such as; Swathview², TexAn (Blondel *et al.* 1998), SeaClass³ and SonarWiz 5⁴.

The remainder of this chapter is organised as follows. In section 7.2, related work is outlined. A conceptual framework for consensus ranking is defined in section 7.3. Section 7.4 describes the application context, including the GLCM features and data used in this study. Section 7.5 covers experimental methods and a discussion of the results. Finally, an evaluation, conclusions, recommendations and suggestions for further work are presented in section 7.6.

¹ University of OULU, Outex Texture Database, <http://www.outex oulu.fi/> [accessed 05-08-2012]

² Quester Tangent Corporation, <http://www.questertangent.com> [accessed 18-06-2012]

³ Triton Imaging Inc. <http://www.tritonimaginginc.com/site/content/products/modules/seaclass> [accessed 18-06-2012]

⁴ Chesapeake Technology Inc. <http://www.chesapeakeotech.com> [accessed 18-06-2012]

7.2 Related work

As the application interest is in machine segmentation and classification of targets in sonar imagery of the seabed, the focus is on some related work in this domain. The more recent topic of feature selection robustness is also considered.

7.2.1 Domain specific feature selection

In one of the leading commercial software packages for seabed classification, Swathview, Preston (2009) applies a variety of kernels to generate 29 features (including GLCM) from the sonar imagery. Principal Components Analysis (PCA) is used to convert the high-dimensional feature space into a three-dimensional component space. Whilst PCA is a widely used technique for extracting meta-features, i.e. principal components, it has three shortfalls;

- Parameter dependent individual features are not evaluated explicitly for class separability, as PCA *per se* is not designed for this purpose.
- Even though an end result of applying PCA is reduction in dimensionality, it still requires the creation of numerous features from the entire input space of a typically large image. This is computationally expensive in time and storage and inefficient too, as sub-images of the input space are generated regardless of the usefulness of the feature kernel.
- The semantics of the original individual features derived from the input imagery is not propagated through to the final classification results. In the case of seabed class maps, the meaning of the class types and their spatial distribution in relationship to the component space and the physical character (ground truth) of the seabed may not always be obvious to an expert interpreter.

Where PCA seeks to maximise the variance of components in a linear space, Curvilinear Components Analysis (CCA) is a local topology preserving, non-linear method. It is still subject to the aforementioned shortfalls of PCA though. Lanaaya *et al.* (2005a) proposed supervised CCA for seabed imagery feature extraction. In their experiments on a database of real sonar imagery, a 15-dimensional space of wavelet

features was reduced to five CCA components. It was found that supervised CCA improved the computational time of the classification tasks, compared to unsupervised CCA and the original, unreduced feature space. However, this increase in computational efficiency was at the expense of accuracy, with the global accuracy of the supervised CCA method being several percent less than the other two methods. They point out a possible reason for this is the highly overlapped classes and the imbalance of different class instances in their database.

In a related paper, Lanaaya et al. (2005b) used a Genetic Algorithm (GA) in a wrapper approach, with accuracy output from a Support Vector Machine (SVM) as the fitness function, to select subsets from 63 wavelet features. They found that using the GA for subset selection led to an improvement in accuracy, compared to no feature selection. However, using the GA wrapper is a time consuming process in the learning stage.

Blondel and Gómez Sichi (2009) use a static, two-dimensional feature space comprising two GLCM features, entropy and homogeneity. Systematic parameter evaluation is carried out to optimise the co-occurrence matrix, using image data from training zones. Parameters are chosen in a contextual (visual) assessment, by considering separations of class clusters in the feature space. The semantics of the features are preserved in this method.

In yet another approach, where sonar imagery was used as one of the test cases, Karoui *et al.* (2008) devised a textural similarity measure based on a weighted sum of Kullback-Leibler divergence measures of distances between the feature distributions.

The Kullback-Leibler divergence is one of numerous distance measures, many of which are defined in Cha (2007). Since induction of a classification model is unnecessary, filter-based approaches, such as these are computationally efficient, compared to say wrapper-centred feature evaluation and selection processes. Furthermore, unlike feature extraction methods such as PCA and CCA, distance measures facilitate an intelligent choice of individual features and their parameterisations, to construct the feature space. The features may be evaluated by investigating their discriminatory potential on small patches of the input image containing the targets of interest. Evaluation and validation patches are selected by an expert human interpreter. So, when generating patterns on large data sets, it is only necessary to apply those pre-selected kernels to generate a few,

salient features. This is clearly more efficient than applying numerous kernels, many of which may create ineffective features.

Several empirical studies have considered distance measures for saliency estimation, in a variety of contexts, for instance; colour and texture (Rubner *et al.* 2001) name-matching (Cohen *et al.* 2003), medical image registration (Penney *et al.* 1998) and signal processing (Basseville, 1989). Yet, aside from Harrison *et al.* (2011) and Karoui *et al.* (2008), there have been few studies addressing the use of distance measures, for feature evaluation with respect to sonar imagery.

7.2.2 Robustness of feature selection methods

Harrison *et al.* (2011) demonstrated the influence of different feature parameterisations and classification tasks on disagreements in rank score permutations derived from different distance measures. Faced with ambiguities of rank, it is not a straightforward matter to determine which is most useful and reliable under the conditions of the parameters, features and classification task on which the ranking is based.

Until quite recently, despite its fundamental importance, the concept of stability of a feature selection method has not received a great deal of attention (Kalousis *et al.*, 2007, Křížek *et al.*, 2007, Saeys *et al.*, 2008.) Stability of a feature selection algorithm can be defined as the robustness of the results or “feature preferences” to small perturbations in the data samples used to evaluate the features (Kalousis *et al.*, 2007.)

Kalousis *et al.* (2007) proposed the first framework to measure the stability of feature selection algorithms, applied to problems in high dimensional spaces. The measure of stability used will depend on how the feature selection algorithm represents the results of the feature evaluation, which, could be a weighting score, a ranking or a feature subset. In the case of ranking, Spearman's rank correlation coefficient was used to evaluate the rank stability (see also, Saeys *et al.*, 2008.) Kalousis *et al.* (2007) found that of the different feature selection methods investigated, no individual method was consistently more stable than the others over the problem space considered, although for some types of problem a given method may be more stable than others. Regarding the issue of stability and classification performance, the point is made that if the algorithm

consistently selects the same features there will be greater confidence in the selection and in the corresponding classification performance.

The importance of considering robustness together with classification performance is echoed by Saeys *et al.* (2008) and as Křížek *et al.* (2007) point out, the stability of the selected features may not be connected to the quality of the features, i.e. a stable feature set is not much use if the features are uninformative or generate inaccurate classification models. Whilst the notion of stability is clearly important to the feature selection method, it cannot be considered independently of feature saliency when making a decision about which feature selection method to use.

Křížek *et al.* (2007) propose an entropy based method for assessing the stability of feature selection methods. The approach is motivated by the fact that the objective functions for feature selection use a random sample of data. The stability of the feature selection algorithms may therefore be evaluated by considering the generated probability distributions of selected feature subsets, since different feature subsets are selected with a certain probability when randomly sampled input data is used. Entropy quantifies the randomness (or uncertainty) of the system states, with a purely random selection of features having maximum entropy and a perfectly stable selection of features on different random samples having zero entropy.

Saeys *et al.* (2008) devised an ensemble approach to improve the stability of feature selection. Ensemble methods as in Scherer *et al.* (2009) and fusion algorithms, e.g. in Baraque (2011) have demonstrated effectiveness in producing reliable solutions across a range of applications, e.g. Pal (2007). In their study, Saeys *et al.* (2008) used two filters: Symmetric Uncertainty (univariate) and the RELIEF algorithm (multivariate) (Kononenko, 1994), together with two embedded methods: RF and recursive feature elimination in a linear Support Vector Machine. For each of the four selection methods, a homogeneous ensemble of feature selection methods was induced using bagging, with 10 sub-samples of data. The output of the feature selector is aggregated into a consensus feature ranking, by weighted voting. This contrasts with the approach presented in this case study, where the consensus rank is produced by aggregating the independent ranks of different feature selectors. The ranks are based on the average values output from the individual feature selector using multiple random samples. This will have the effect of suppressing the influence of small changes in the data on the output rank, circumventing

some of the stability issues. We can then consider the robustness of the method over different feature parameterisations and classification tasks.

7.3 Consensus feature ranking methods

The distance between estimated univariate probability density functions (pdf) of feature components generated on different classes, provides a measure of the classification error probability (Webb, 1999, Duda *et al.*, 2001.) Disjoint distributions are indicative of highly discriminative (i.e. salient) features, compared to those with class distributions that are closer together.

Motivated by the inherent variability of qualitative sonar imagery, the key idea is to fuse diverse information captured by multiple measures in a consensus approach. The goal is to achieve an efficient means of generating a more representative and robust ranking. A theoretical framework presented in this section defines the most general case of consensus ranking with m independent distance measures. Important stages in the process include:

1. balanced, multiple random sampling of the input space,
2. scaling the individual feature values to $[0, 1]$ across the classes and features,
3. estimating the normalised, non-parametric probability distributions for all features and classes,
4. computing the median pair-wise distances between all distributions (based on the multiple random samples,)
5. assigning an integral (or continuous) consensus rank score to each feature using equally weighted outputs from each distance measurement,
6. reliability weighting of rankings by considering the robustness to different feature parameterisations and classification tasks.

7.3.1 Non-parametric distribution estimation

A non-parametric approach to representing the feature distributions of the classes is preferred over a model-based approach, mainly due to the potential ambiguity of fitted models (particularly with qualitative sonar imagery).

For a set of n specific classes, $c_s = \{c_1, c_2, \dots, c_n\}$, non-parametric class conditional probability distributions $\hat{P}(\mathbf{x} | c_s)$ for each component, x_f of the feature vector, $\mathbf{x} \in \mathbb{R}^D$ (where D is the dimension of the feature space) are estimated using histograms.

The scaled feature domain, $x_f = [0, 1]$ is divided into equal width intervals, i and the distributions are normalised: $\int_{x_f} p \cdot dx_f \equiv \int_{x_f} q \cdot dx_f \equiv 1$. Pair-wise distances between the component distributions for specific classes are computed if the distances between a generic target T and non-target $\neg T$ distribution is required ($T, \neg T \in \{c_1, c_2, \dots, c_n\}$, $T \cap \neg T = \emptyset$.) Probability distribution estimates for the generic classes are then, $p = p(x_f | T)$ and $q = q(x_f | \neg T)$ respectively.

Once computed the set of $\frac{1}{2} nD(n-1)$ histograms is re-used by each distance measure. Histograms are recomputed for each specific parameter combination from the feature parameter space (see subsection 7.4.1). The number of bins is fixed at 30 for the results reported in this chapter.

7.3.2 Individual distance measures

Without loss of generality, six commonly used parameter-free distance measures (defined in table 7.1) are chosen for this test case, five of them from the different families of measures as categorised by Cha (2007). These are; Minkowski family: Euclidean L_2 (Eu) , L_1 absolute difference family: Canberra (Ca), Fidelity family: Bhattacharyya (Bh) (Bhattacharyya, 1943) Chi-squared family: Chi-Squared (CS), Shannon's entropy family: Kullback-Leibler divergence (KL) (Kullback and Leibler 1951) finally, the Kolmogorov-Smirnov (KS) distance (Geman *et al.* 1990).

$$d_{\text{Eu}}(p, q) = \sqrt{\left\{ \sum_{i=1}^n (p_i(x_f) - q_i(x_f))^2 \right\}}$$

$$d_{\text{Ca}}(p, q) = \sum_{i=1}^n \frac{|p_i(x_f) - q_i(x_f)|}{p_i(x_f) + q_i(x_f)}$$

$$d_{\text{Bh}}(p, q) = \arccos \left\{ \sum_{i=1}^n \sqrt{p_i(x_f) q_i(x_f)} \right\}$$

$$d_{\text{CS}}(p, q) = \sum_{i=1}^n \frac{(p_i(x_f) - q_i(x_f))^2}{p_i(x_f) + q_i(x_f)}$$

$$d_{\text{KS}}(p, q) = \max \{ |\text{EDF}_p(x_f) - \text{EDF}_q(x_f)| \}$$

$$d_{\text{KL}}(p, q) = \sum_{i=1}^n p_i(x_f) \ln(p_i(x_f)) + \sum_{i=1}^n q_i(x_f) \ln(q_i(x_f)) -$$

$$\sum_{i=1}^{i=n} p_i(x_f) \ln(p_i(x_f) + q_i(x_f)) -$$

$$\sum_{i=1}^{i=n} q_i(x_f) \ln(p_i(x_f) + q_i(x_f)) + 2 \ln(2)$$

Table 7.1 Definitions of the candidate distance measures used in the test case.

Euclidean distance is a distance between vectors but the concept is transferrable to feature distributions (histograms), since the probabilities (frequency coded bin values) are vector components. Euclidean distance and other measures are affected by the range of feature values, hence the need to normalise both the feature values and the histograms. Although it is widely employed in a variety of applications, Aksoy and Haralick (2000) found that it is was significantly outperformed by likelihood based measures on image retrieval tasks. Foote (1997) found it was outperformed by a cosine distance in retrieving simple sounds.

In common with the other measures, the Canberra measure has diverse usage. Emran and Ye (2001) used a modified Canberra measure to establish the similarity of an observed event sequence to normal or attack sequences, for the generation of computer intrusion warnings. Their approach performed well in the special case where attack and normal events were widely separated but less satisfactorily in other cases. In Androutsos *et al.* (1998), the measure was compared against others in an application to colour image retrieval. Best retrieval rates were achieved with the Minkowski L_1 norm and L_2 norm (Euclidean) distance measures. The Canberra measure resulted in many erroneous retrievals. Its retrieval rate was comparable to the Minkowski L_∞ measure.

Bhattacharyya's measure is a measure of angular separation between the histograms, since the probabilities represent direction cosines (Bhattacharyya, 1943). It is applicable to any data sample, regardless of the underlying distribution (Thacker *et al.* 1997).

Puzicha *et al.* (1997) proposed the Chi-squared statistic as a similarity measure for texture segmentation and image retrieval tasks. It can also be used as an approximation to the Kullback-Leibler divergence (Vasconcelos and Lipmann, 2000). Bugatti *et al.* (2008) evaluated the Chi-Squared and other measures in the context of similarity queries applied to medical image data sets. They found that it outperformed the other methods under test, with respect to precision and recall, only in the case when a feature vector comprising Zernike moments was used. With other feature types the Canberra distance produced a better performance.

The Kolmogorov-Smirnov distance, in contrast to the other distance measures here uses an estimated distribution function (EDF), instead of an estimated pdf. It was first proposed as a distance measure in an image segmentation context by Geman *et al.* (1990.)

Kullback-Leibler divergence is an information theoretic method, and differs from the others in that it measures probabilistic uncertainty between the feature distributions. An extended symmetric version from Petrou and Sevilla (2006) is implemented in this study.

7.3.3 Consensus ranking of features

Each of the M distance measures, \mathbf{d}_m , $m \in [1,2,\dots,M]$ generates a set of real values dependent on the component of the feature vector, x_f , $f \in [1,2,\dots,D]$ and specific class distribution comparison, $(p, q)_w$, $w \in [1,2,\dots,\eta]$, where $m, f, w \in \mathbb{Z}^+$ and say, $\eta = |T| - |T^-|$. Thus, for every pair-wise class combination we have a matrix, $\mathbf{d}_m(\mathbf{x}, (p, q)_w)$ of M sets of computed distance values for each feature component, given by,

$$\mathbf{d}_m(\mathbf{x}, (p, q)_w) = \begin{pmatrix} d_{11}(x_1) & \dots & d_{1D}(x_D) \\ \vdots & \ddots & \vdots \\ d_{M1}(x_1) & \dots & d_{MD}(x_D) \end{pmatrix} \quad (7.1)$$

Note that for the purposes of this work, the values are the median distance values computed from the histograms based on multiple random samples, since the median is less affected by outliers. The output of a ranking function is the matrix, $\mathbf{r}(\mathbf{x}, (p, q)_w)$ (matrix 7.2), of rank scores (weights), \mathbf{r}_* for each feature component as determined using the ranks derived from each distance measure and specific class distribution comparison.

$$\mathbf{r}(\mathbf{d}_m, \mathbf{x}, (p, q)_w) = \begin{pmatrix} x_{11}(r_*) & \dots & x_{1D}(r_*) \\ \vdots & \ddots & \vdots \\ x_{M1}(r_*) & \dots & x_{MD}(r_*) \end{pmatrix} \quad (7.2)$$

The top-ranked feature component is assigned an integral rank score equal to the cardinality of the feature set (i.e. D). Descending positions are assigned scores, progressively decreased by one. In the event of a tie for any rank position, all tied features are allocated equal rank scores of the tied position. Elements in the rows of matrix 7.2 form a specific permutation of rank scores for the feature vector as derived from the corresponding distance measure. Each feature component has a rank score, in a

three-dimensional version of matrix 7.2 (the third dimension being the pairwise combinations of classification tasks).

By summing the rank scores for feature components over all distance measures and all pair-wise combinations of specific classes, a consensus rank score for each component, $\mathbf{r}_{\text{co}}(x_f)$ in terms of its capacity to discriminate between the generic classes is obtained, from,

$$\mathbf{r}_{\text{co}}(x_f) = \sum_{m=1}^M \sum_{w=1}^{\eta} x_{m,f}(r^*, (p, q)_w) \quad (7.3)$$

In the case of a single binary classification task, this is equivalent to a column-wise summation of the elements in matrix 7.2. The consensus represents a general case for any number of components, distance measures and pair-wise class comparisons.

Of course, it is not essential to assign an integral rank score. Since, the distance measures are continuous valued functions over bounded ranges, clearly, the method can easily be modified to generate a normalised mean distance measure, for any component and pairwise class comparison,

$$\bar{\mathbf{d}} = \frac{1}{M} \sum_{m=1}^M \frac{\mathbf{d}_m(x_f, (p, q)_w) - \min(f_{\mathbf{d}_m})}{\max(f_{\mathbf{d}_m}) - \min(f_{\mathbf{d}_m})} \quad (7.4)$$

where, $\max(f_{\mathbf{d}_m})$ and $\min(f_{\mathbf{d}_m})$ are the functional maxima and minima, respectively, of the specific distance measure. The output from eq. 7.4 can be used to produce a continuous valued ranking of the feature components. Continuous valued output may be preferable to integral rank scores when applying certain feature selection methods, since rank does not provide a measure of the magnitude of the distances between the feature vector components.

Computationally, the multiple (independent) distance measure calculations are embarrassingly parallel as the results are generated independently, from the common set of input distributions. There is no need for communication of intermediate results

between individual measures. The process can easily be implemented in parallel if justified by the data volume, feature parameter space and number of dichotomies.

7.3.4 Reliability weighting of permutations

In the case of no ties, there are $D!$ possible permutations of the elements (rank scores) in the rows of matrix. Eq. 7.3 produces a consensus rank score on the premise that each distance measure has an equally valid rank permutation to contribute to the consensus. In reality, the performance or suitability of an individual distance measure varies according to the feature parameterisation and classification task. Therefore, the specific permutation of rank scores derived from one measure may be less robust than the others and should be treated differently.

Now, consider a case where the robustness is not known in advance. As the most reliable individuals (i.e. the most robust ranks over the problem space) are desired for contributing to the consensus, a quality control process is introduced to evaluate the individuals.

It uses a feedback mechanism to compare the rank permutation instances in \mathbf{r} (matrix 7.2) with baseline permutations \mathbf{r}_b (see subsection 7.2) using Spearman's rank correlation coefficient, ρ .

A diagonal weighting matrix, $\mathbf{W}(\mathbf{d}_m, (p, q)_w)$ is constructed, with entries corresponding to the rank correlation coefficient, ρ_m where,

$$\rho_m(\mathbf{r}, \mathbf{r}_b) = 1 - 6 \sum_i \frac{(\mathbf{r}_i - \mathbf{r}_{b(i)})^2}{D(D^2 - 1)} \quad (7.5)$$

The matrix of component rank scores, weighted for reliability and rounded to the nearest integer, $\mathbf{rw}(\mathbf{d}_m, (p, q)_w)$ is,

$$\mathbf{r}_W = \begin{pmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_D \end{pmatrix} \begin{pmatrix} x_{11}(r_*) & \cdots & x_{1D}(r_*) \\ \vdots & \ddots & \vdots \\ x_{M1}(r_*) & \cdots & x_{MD}(r_*) \end{pmatrix} \quad (7.6)$$

further, by applying a binary threshold, say,

$$\rho_m = \begin{cases} 1, & \rho_m \geq 0.5 \\ 0, & \rho_m < 0.5 \end{cases} \quad (7.7)$$

rankings from \mathbf{r} can be accepted (1) or rejected (0), if the robustness measure for the ranking is above or below the threshold, respectively. After determining the weightings, the consensus rank score for a feature component, weighted for robustness, $\mathbf{r}_\gamma(x_f)$ is recomputed by,

$$\mathbf{r}_\gamma(x_f) = \sum_{m=1}^N \sum_{w=1}^n (\lceil \gamma_{m,f} \rceil, (p, q)_w) \quad (7.8)$$

where, $N = M$ if no sequences are rejected and $\lceil \gamma_{m,f} \rceil \in \mathbf{r}_w$

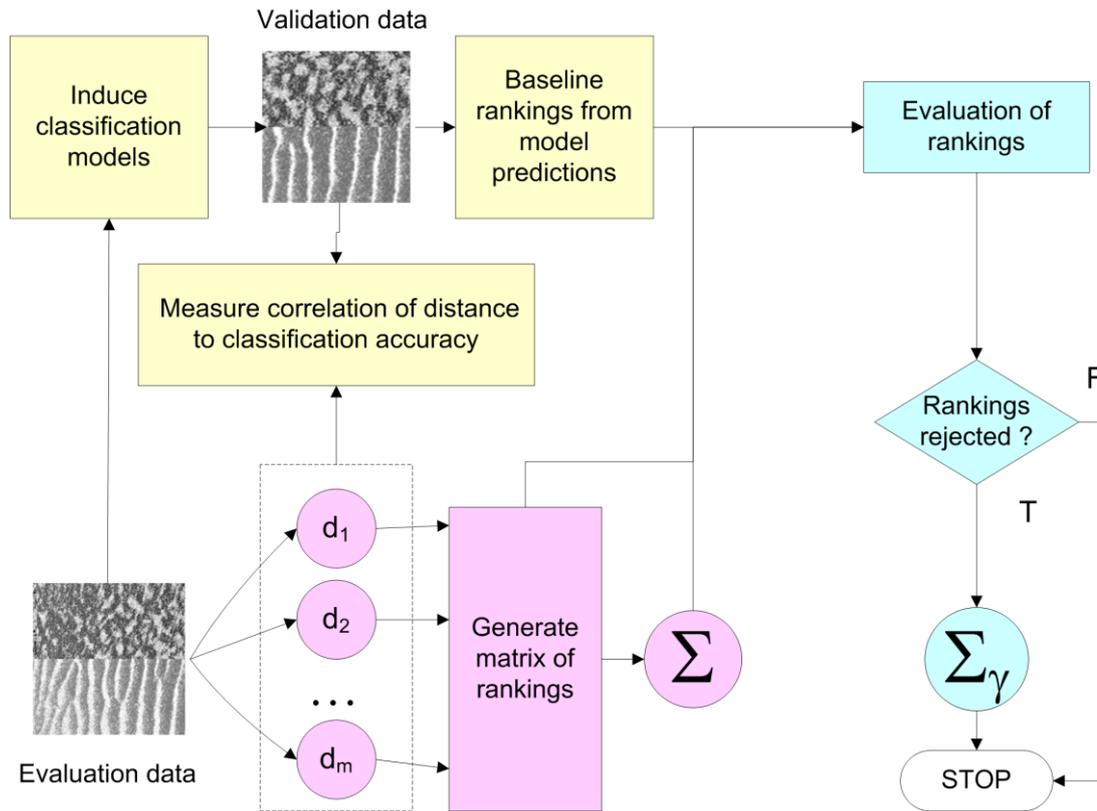


Figure 7.1. Greatly simplified representation of the experimental harness. The three main conceptual blocks of the processes in the harness are colour coded according to the tasks performed, such as estimating feature saliency and feature rank (mauve symbols), evaluating the feature evaluation methods and their ranking robustness (yellow and blue symbols). See text for an explanation.

Figure 7.1 represents the three conceptual processing blocks in the experimental harness, corresponding to feature evaluation and ranking (mauve coloured symbols), validation processes (yellow), ranking evaluation and quality control (blue). The filter sub-process is represented by the mauve symbols. Globally, the experimental harness is a wrapper for evaluating the filter methods.

The process begins by deriving a set of parametric features from the image data instance (evaluation data) under consideration. Matrix 7.2 is constructed (generate matrix of rankings) using the output from the distance measure channels and the consensus rank computed by applying eq. 7.3. The independent validation sub-process is used to generate baseline ranks from separate validation data and for assessing the correlation of the measured saliency of the features with the classification accuracy from multiple models. The sub-process for evaluation of ranking robustness and weighting of the

distance measure ranks, according to their reliability, uses information generated in the validation and feature evaluation sub-processes.

7.4 Application Context

Identifying and mapping interesting regions in seabed imagery is a subjective and time consuming task for a human analyst and an immensely challenging task for any feature-based machine learning system. Much seabed classification research has focused on using sonar image texture as a discriminative subspace. Since GLCM features are probably the most widely used in the domain, they are applied in the test cases here.

7.4.1 GLCM feature creation

Devised as a means of textural analysis by Haralick *et al.* (1973) a co-occurrence matrix is an estimation of the joint probability distribution of combinations of quantised intensity value pairs at the ends of a sampling vector, d . The sampling vector is applied at multiple orientations, $\theta \in \Theta$, to pixel neighbourhoods in the input image, bounded by a kernel, of size k . The size of k determines the final resolution of the segmentation or class map, whereas $\|d\|$ is the length scale at which the texture is analysed.

Although there are many others, five features (or textural indices), are derived from the co-occurrence matrices in our test case; Angular Inverse Difference Moment (AIDM), Angular Second Moment (ASM), Contrast (CON), Correlation (COR) and Entropy (ENT). Our choice of derived features follows those used by Reed and Hussong (1989) in their classification of SeaMARC II sonar imagery.

The discrete valued parameter space $\mathcal{A} = \{\lambda_1, \lambda_2, \dots, \lambda_\tau\}$ yields $\prod_{j=1}^{\tau} |\lambda_j|$ unique instances of the full feature vector, \mathbf{x} . Matrices are generated at four symmetric pairs of orientations, so the dimension of \mathbf{x} is initially 20. The parameter and configuration subset for which results are presented in this paper is summarised in table 7.2.

Parameter	Value
Kernel dimension, k (pixels)	{5, 11, 17}
Quantisation level, Q (bits)	3
Sampling orientation	Θ_0, Θ_1
Inter-pixel distance, $\ d_0\ $	$f(k, S_p)$
Configuration	
$(x_f)_\mu = \frac{1}{ \theta } \sum_{k=1}^{ \theta } (x_f)_\theta$	$D = \mathbf{x} $
Sampling pattern, $S_p(\theta)$	Octagonal

Table 7.2 Summary of co-occurrence matrix parameters and configurations

It would be feasible to work directly with the five features derived at four orientations and several different kernel sizes, treating each parameterisation of the features as a separate attribute. However, this is not really necessary as \mathbf{x} would soon grow in dimension greatly, containing hundreds of feature components, many of which would be highly correlated. The end product of this would be a high-dimensional feature space containing numerous redundant or irrelevant feature components. PCA could be applied to reduce the dimensionality but this is not what is desired, as outlined in the earlier argument in subsection 7.2.1.

Mean components are computed over Θ , to create the five-dimensional, parameter dependent, rotationally invariant vectors, $\mathbf{x}_\mu = (\text{CON}(A), \text{COR}(A), \text{ASM}(A), \text{AIDM}(A), \text{ENT}(A))$. Obviously, the removal of directional dependence of the kernels usefully serves to reduce dimensionality. Directionally dependent features can be useful for discriminating anisotropic textures but this is not of concern here. An additional advantage is that it tends to suppress some of the undesirable variability captured by the features, due to the effects imparted by sensor platform motion on the imagery.

The quantisation level is fixed at $Q = 3$ bits and an octagonal sampling pattern is applied. Defining the inter-pixel distances at a particular analysis scale, $\alpha_p = \{d_0, d_1\}$ where, d_0 is orthogonal to the sides of the (square) pixel and $d_1 = \lceil \sin(\frac{\pi}{4})d_0 \rceil$, diagonal

to the pixel, a set of sampling vectors for the octagonal pattern at any scale of analysis is,

$$\begin{cases} d_0 \{ \cos(\theta_0)(1,0)^T + \sin(\theta_0)(0,1)^T \}, & \Theta_0 \\ d_1 \lceil \cos(\theta_1)(1,0)^T \rceil + d_1 \lceil \sin(\theta_1)(0,1)^T \rceil, & \Theta_1 \end{cases}$$

$$\Theta_0 = \left\{ 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2} \right\} \text{ and}$$

$$\Theta_1 = \left\{ \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4} \right\}$$

(7.9)

An octagonal sampling pattern is implemented at these kernel sizes and scales of analysis as it ensures the scales in the Θ_0 and Θ_1 directions are approximately equal. This contrasts to a square sampling pattern in which the Θ_1 length scales are a factor of $\sqrt{2}$ greater than in the Θ_0 directions. The octagonal pattern also helps to ensure directional uniformity in the quantity of samples entered in the joint distribution. Exploration of the parameter space is restricted to the variables, k and d_0 , in this investigation, keeping all others fixed. Feature elements of \mathbf{x} , in addition to being rotationally invariant, are always derived at homogeneous scales of analysis and kernel sizes.

7.4.2 Experimental data

Sidescan sonar imagery contains useful textural information indicating the nature and extent of different physical and morphological regimes on the seabed (Blondel 2007, Johnson and Helferty 1990.) Of particular concern are protected habitats, such as reefs formed by colonies of the tube-building worm, *Sabellaria Spinulosa* (Sabellaria)(Birchall, 2007, Hendrick and Foster-Smith, 2006, Limpenny *et al.* 2010,.) The test cases focus on Sabellaria as the target texture in sonar mosaic imagery and attempt to discriminate it from other specific seabed types. In addition, some consideration is given to Sabellaria textures in waterfall imagery. Photographic images of a selection of natural textures from the Outex database are also used to make wider comparisons. Samples of the imagery are shown in figure 7.2. Mosaic imagery sized for the kernel, $k = 5$ is shown in the top row, waterfall data ($k = 11$) in the middle row and Outex

textures ($k = 11$), in the bottom row. The data types and textural classes are summarised in table 3. Abbreviations, e.g. Sab., Ro., are used in the next section.

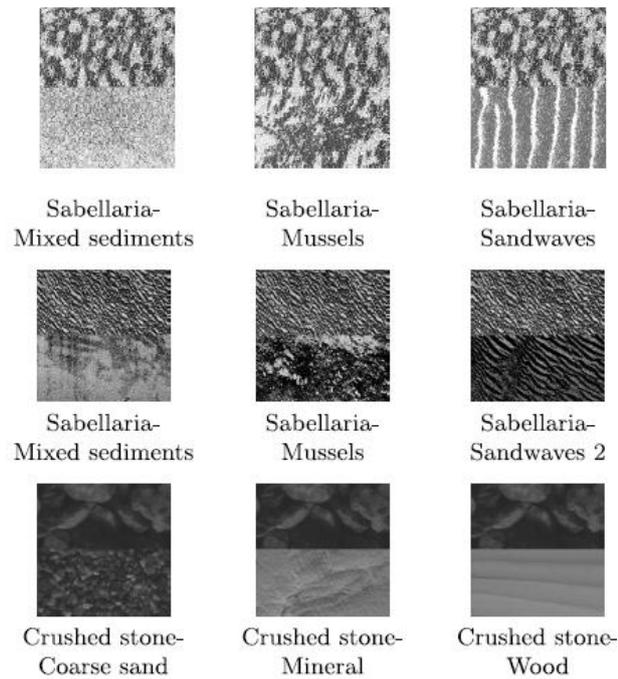


Figure 7.2. Examples of the different textural classes used in the test case.

Mosaic	Waterfall	Outex
Sabellaria (Sab.)	Sabellaria (Sab.)	Crushed stone (CSt.)
Mixed sediments (Mix.)	Mixed sediments (Mix.)	Coarse sand (CSa.)
Mussels (Mu.)	Mussels (Mu.)	Fine sand (FSa.)
Sandwaves (Sw.)	Sand (Sa.)	Flour (Fl.)
Rock (Ro.)	Sandwaves 1 (Sw1.)	Mineral (Mnl.)
Sand (Sa.)	Sandwaves 2 (Sw2.)	Wood (Wo.)

Table 7.3. Summary of all classes and datasets used in this study together with their abbreviations.

7.5 Experimental results and discussion

The purpose of these experiments is to investigate the efficacy of our consensus approach to feature saliency estimation and ranking. The main application interest is the

sonar mosaic imagery so the presentation and analysis of results concentrates more on these data.

A set of quantitative properties that a useful method for evaluating and ranking features on sonar mosaic imagery (or any other data) should possess is defined as;

1. Distance estimates correlate well (in a linear, least squares regression) with the median accuracy of multiple classification models applied to validation data.
2. Low variability of property (1) over different feature parameterisations and classification tasks.
3. Good rank correlation of the derived feature rankings with median baseline rankings.
4. Low variability in property (3) over a range of feature parameterisations and classification tasks.

A Naïve Bayes classifier is used as the core learning algorithm in the wrapper for determining the classification accuracies and establishing the baseline rankings. Multiple models are induced for each feature parameter setting and classification task. Median classification accuracies are then computed from the class predictions of the models on a set of validation patterns. To clarify, the purpose of the wrapper is as a means of validation of the filter method within the experimental harness. In practice, to configure the feature evaluation and ranking process to work in the context of the same classification objectives on standardised, calibrated data sets, this validation process would need to be performed once on representative samples of the data. Following the validation, the filter may then potentially be applied to other similar data with the same classification objectives, without further validation.

7.5.1 Investigating saliency and classification accuracy

Properties (1) and (2) concern the relationship of distances output from individual measures and the consensus, to the median classification accuracy from multiple models. Figure 7.3 shows the distances (horizontal-axis) between individual features

over a range of classification tasks and parameters, plotted against classification accuracy, for each feature evaluation method applied to the mosaic imagery. Consensus (Co) saliency in this case, is the output from equation 7.4. Sabellaria is the target class and the markers represent the texture classification tasks.

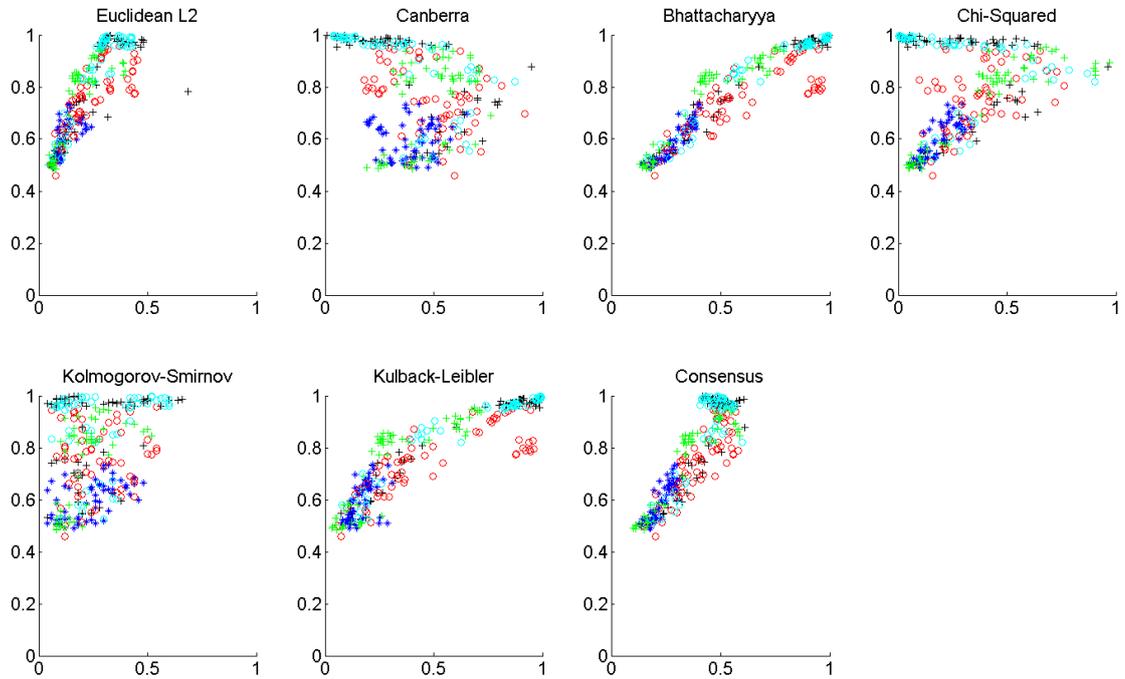


Figure 7.3. Scatter plots showing measured saliency (horizontal-axis) vs. median classification accuracy for individual features, over multiple classification tasks and parameter settings on the sonar mosaic data. Marker key: + Sabellaria-Mixed sediments, * Sabellaria-Mussels, o Sabellaria-Sandwaves, + Sabellaria-Rock, o Sabellaria-Sand. }

According to figure 7.3 there are varying degrees of scatter amongst the different classes and feature evaluation methods. Eu, Bh, KL and Co are, overall, more correlated to the median classification accuracy than Ca, CS and KS. It is also clear that Eu, Bh and KL have outlying values, compared to Co. In order to quantify the relationship of estimated, relative feature saliency to classification accuracy, the R^2 statistic is used as a measure of goodness of fit, in a linear, least squares regression. The fit is better as $R^2 \rightarrow 1$. Consistently high values of R^2 are desired over different classification tasks, i.e, a relatively low standard deviation of R^2 . The results for each of the data sets are summarised in table 7.4.

Mosaic	Eu	Ca	Bh	CS	KS	KL	Co
Sab-Sw	0.55	0.15	0.90	0.00	0.07	0.84	0.78
Sab-Mix	0.44	0.10	0.79	0.54	0.28	0.40	0.77
Sab-Sa	0.55	0.01	0.66	0.26	0.05	0.57	0.74
Sab-Ro	0.70	0.35	0.82	0.72	0.41	0.66	0.83
Sab-Mus	0.75	0.13	0.86	0.00	0.09	0.79	0.76
μ	0.60	0.15	0.81	0.30	0.18	0.65	0.78
σ	0.13	0.13	0.09	0.32	0.16	0.18	0.03
Waterfall							
Sab-Mix	0.24	0.01	0.55	0.15	0.13	0.45	0.62
Sab-Mu	0.53	0.03	0.49	0.49	0.61	0.46	0.54
Sab-Sa	0.71	0.00	0.79	0.05	0.04	0.73	0.80
Sab-Sw1	0.19	0.12	0.45	0.27	0.60	0.36	0.46
Sab-Sw2	0.44	0.16	0.30	0.27	0.49	0.33	0.36
μ	0.42	0.06	0.52	0.25	0.37	0.47	0.56
σ	0.21	0.07	0.18	0.16	0.27	0.16	0.17
Outex							
CSt-Csa	0.10	0.25	0.56	0.31	0.14	0.37	0.57
CSt-FSa	0.36	0.10	0.67	0.43	0.10	0.55	0.63
CSt-Fl	0.36	0.08	0.90	0.01	0.19	0.83	0.68
CSt-Mnl	0.18	0.41	0.69	0.32	0.30	0.51	0.63
CSt-Wo	0.11	0.14	0.22	0.12	0.17	0.10	0.32
μ	0.22	0.20	0.61	0.24	0.18	0.47	0.57
σ	0.13	0.14	0.25	0.17	0.08	0.27	0.14

Table 7.4 R^2 values for specific classification tasks and distance measures

Figure 7.4 shows a plot of the mean, μ (horizontal axis) against the standard deviation, σ (vertical axis) of R^2 , over all tasks and data sets. From top to bottom; (a) sonar

mosaic, (b) sonar waterfall, (c) Outex. Horizontal and vertical broken lines in each plot correspond to the mean values of μ and σ , respectively, over all distance measures and classification tasks on the particular data set.

Markers further to the right of the plots and closer to the horizontal axis correspond to methods which satisfy, properties (1) and (2) more strongly. Since, correlation to classification accuracy is better than the mean for the group and variability over the tasks and parameters lower than the mean variability. Assigning a score of one point to a method with a mean R^2 greater than the group mean and one point for a mean standard deviation less than the group mean, the results over all data sets and tasks are conveniently summarised in table 7.5. Co is the only method that satisfies properties (1) and (2) over all tasks on the three different sets of data.

On the mosaic data (our main interest), it is clear that Bh is the strongest individual measure over this range of parameters and classification tasks. It has a mean R^2 value of 0.81, slightly higher than Co (0.78). However, the key advantage of Co is the 67% lower spread of values ($\sigma = 0.03$) over the classification tasks on these data, c.f. the best individual, Bh, with $\sigma = 0.09$. Low variability in the relationship between saliency and classification accuracy over different classification tasks is most desirable in practical, industrial applications.

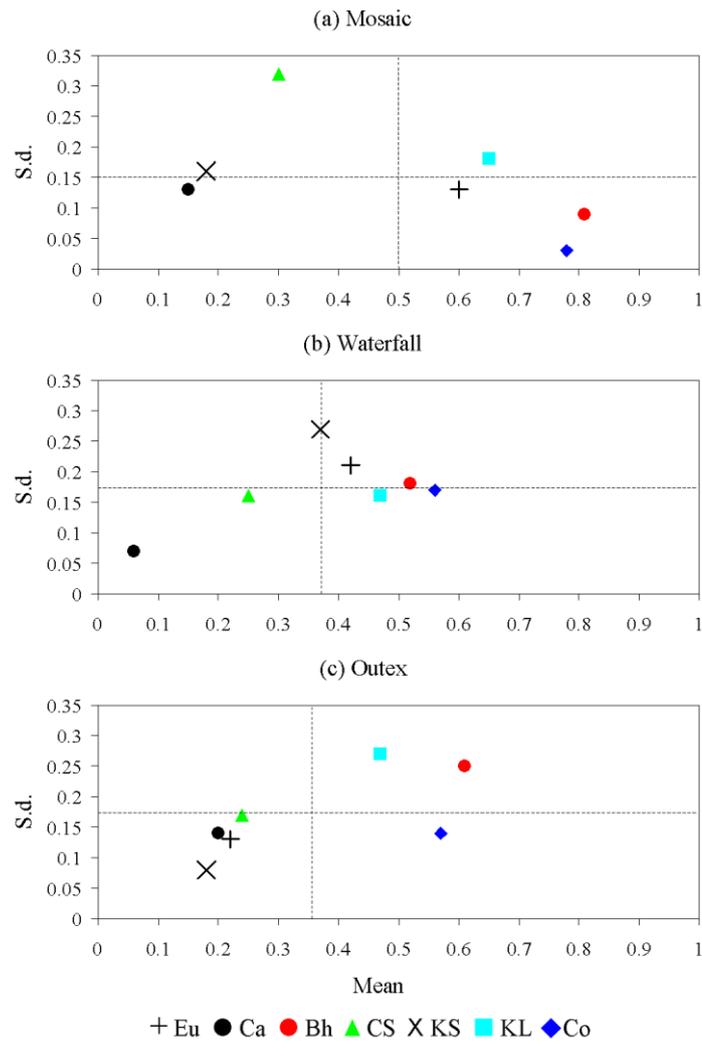


Figure 7.4. Plots showing the mean (horizontal axis) and standard deviation of the R^2 statistic over all tasks and parameters for the different datasets.

All data	Eu	Ca	Bh	CS	KS	KL	Co
Property (i)	2	0	3	0	0	3	3
Property (ii)	2	3	1	2	1	1	3

Table 7.5 Properties (i) and (ii): Number of times the properties are satisfied for all data sets.

Clearly though, the measured saliency must also correlate to the classification accuracy, regardless of whether the accuracy itself is relatively high or low. Co is the overall

winner with respect to these qualities, considered over all the data sets, closely followed by the individuals, Bh and KL. It is interesting to note that the overall mean correlation to classification accuracy is approximately 25% higher on the mosaic imagery, compared to the other two.

7.5.2 Defining a baseline rank

Before properties (3) and (4) concerning the rank correlation and robustness can be considered, it is first necessary to define a baseline ranking against which all other computed ranks can be compared. The approach used is the same for all data sets, so we explain this procedure in the context of the sonar mosaic data.

In a supervised learning scenario, a human operator may manually select small evaluation and validation regions from the sonar imagery. The evaluation data is used to estimate the feature saliency and induce the classification models. It is assumed the underlying class distributions are similar in the evaluation and validation data sets and the validation data set can be defined as a non-universal, class-specific benchmark (i.e. the benchmark is valid on one data set only.)

A baseline ranking is defined, derived from the median classification accuracy of multiple predictions on the validation data, using an exhaustive search of the feature subsets on the different classification tasks, over specified parameter ranges.

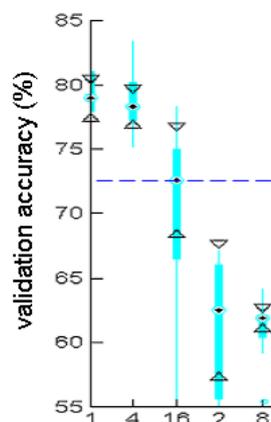


Figure 7.5. Classification accuracy on validation data, using multiple models for individual features. The median values are used to establish the baseline ranking.

Figure 7.5 shows the results used to determine a baseline ranking for the individual features (subsets of $D = 1$), on the Sabellaria-sandwaves classification task, with parameter settings, $k = 17$, $\|d_0\| = 13$. Validation accuracy is on the vertical axis and the numbers on the horizontal axis correspond to binary codes for the individual features. Small black triangles show the 95% confidence interval for the location of the median validation accuracy, with the median indicated by a black dot. The horizontal broken line represents overall median accuracy for the subset bin. Basing the rank on median accuracies, in descending order, the individual features are, (1) ENT, (4) ASM, (16) CON, (2) AIDM and finally, (8) COR. In accordance with the earlier definition, this is used for the baseline rank score permutation of the individual feature components for the prescribed classification task and parameter settings, i.e. $\mathbf{r}_b(\mathbf{x}) = (3, 1, 4, 2, 5)$.

It is clear from figure 7.5 though, that some of the individual features produce classification models that are more stable than others to variability in the sampled data (all other things being equal.) ENT and ASM models are both stable as there is a relatively small spread between the 95% confidence limits. They are also salient as they yield high classification accuracies. COR is the most stable but least salient. CON and AIDM are relatively less stable than the others, with a wide spread of accuracies between the confidence limits and in some cases, relatively poor models are induced.

If the confidence limits of classification models induced from the individual features have overlapping ranges of accuracy, other baseline rank permutations are possible. It is far more likely that the top two rank positions held by ENT and ASM could be interchanged but impossible (in this test case) for the rank of ENT and AIDM to be interchanged as there is a vertical “gap” between the lower and upper limits of the confidence intervals. Interestingly, whilst AIDM is relatively less stable, it is unlikely to affect the rank stability as much as the two most salient feature models with a relatively higher stability. Obviously, the baseline ranking has some uncertainty, dependent on model and rank stability. A rigorous, quantitative treatment of these issues could form the subject of future work.

7.5.3 Investigating task robustness

Having defined a method for establishing baseline rankings attention is turned to using Spearman’s rank correlation coefficient, ρ , to compare the rankings derived from the

distance measures, with the baselines, over a variety of parameters and classification tasks. The purpose is to assess the robustness of the rankings against properties (3) and (4). Table 7.6 contains detailed results for the values of ρ , computed over a sample of classification tasks and parameter settings for the sonar mosaic imagery. Consensus rankings are generated in unweighted (Co) and weighted (Co+) modes of operation, i.e., from eq. 7.3 and 7.8, respectively. In the weighted mode a binary threshold is applied, as defined in eq. 7.7.

Differences, $\Delta\rho$, between the correlation coefficients for unweighted and weighted cases and the number of distance measure rankings rejected, N , for each task/parameterisation, are given in the last two columns of the table. The final row of the table contains the relative frequency of rejections for each method over the sample of tasks and parameterisations. The mean and standard deviation, ρ_μ , ρ_σ , for the methods over all tasks and all data sets are summarised in table 7.7 and shown graphically in figure 6. Values of ρ_μ and ρ_σ , are shown on the horizontal and vertical axes, respectively. The vertical and horizontal broken lines on each plot indicate the overall mean values for ρ_μ and ρ_σ for the specific data set. Methods with high values of ρ_μ (property (3)) and low values of ρ_σ (property (4)) are preferable as they reflect good correlation of the estimated ranking with the baseline and a relatively high robustness to the different classification tasks and feature parameterisations. These methods are indicated by markers towards the bottom-right of the plots.

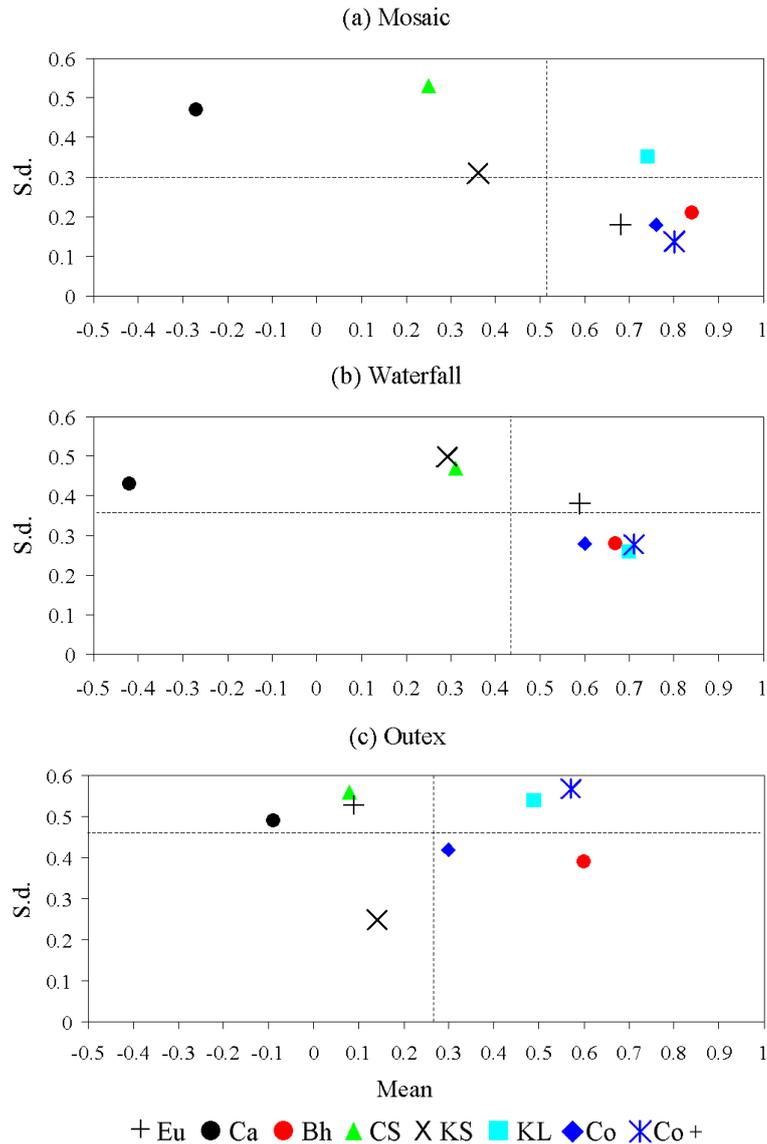


Figure 7.6. Mean and standard deviation of Spearman's rank correlation coefficient.

Considering the results on the mosaic imagery, as an individual, Bh has performed well, on average, with $\rho_\mu = 0.84$. The unweighted committee has a slightly lower value ($\rho_\mu = 0.76$) but the variability in the results is 15% better (lower), $\rho_\sigma = 0.18$ for Co, c.f. Bh, $\rho_\sigma = 0.21$. Using the quality control mechanism, with the binary threshold to accept or reject individual rank sequences improves ρ_μ slightly, by increasing the value from 0.76 for Co to 0.80 for Co+. Importantly, variability is reduced further still, by another 22%, from $\rho_\sigma = 0.18$ for Co to $\rho_\sigma = 0.14$ for Co+ (i.e. 34% lower variability than Bh), demonstrating an improved robustness on the sonar mosaic tasks.

In two cases, as shown in table 7.6, Bh performed unsatisfactorily, $\rho = 0.3$, on the Sabellaria-mussels and Sabellaria-sand classification tasks. Using the same parameters, it performed well on other tasks such as Sabellaria-sandwaves. On the other hand, using Co+, none of the rankings has a ρ - value less than 0.6, a clear indication of good ranking conformity over the tasks.

Task	(k, d)	Eu	Ca	Bh	CS	KS	KL	Co	Co+	$\Delta\rho$	N
Sab-Mix	11,9	0.6	-0.6	1.0	0.1	0.2	1.0	0.7	1.0	+0.3	3
	11,7	0.3	-0.1	0.9	-0.2	-0.1	1.0	0.6	0.9	+0.3	4
	17,15	0.6	-0.6	0.9	-0.3	-0.1	0.9	0.4	0.7	+0.3	3
	17,13	0.3	-0.4	0.7	-0.2	-0.3	0.7	0.5	0.7	+0.2	4
Sab-Mu	11,9	0.8	-0.2	0.9	0.8	0.8	-0.3	0.9	0.8	-0.1	2
	11,7	0.7	0.4	0.3	0.6	0.0	0.6	0.6	0.6	0.0	3
	17,15	0.6	0.3	0.9	0.7	0.8	0.6	0.6	0.7	+0.1	1
	17,13	0.6	-0.3	0.9	0.7	0.8	0.0	0.7	0.6	-0.1	2
Sab-Sw	11,9	0.9	-0.5	1.0	0.7	0.3	1.0	0.9	0.9	0.0	2
	11,7	0.9	-0.6	1.0	0.9	0.5	1.0	0.9	0.9	0.0	1
	17,15	1.0	-0.9	0.9	-0.1	0.5	0.8	1.0	0.9	-0.1	2
	17,13	0.9	-0.9	1.0	0.0	0.3	1.0	0.9	1.0	+0.1	3
Sab-Ro	11,9	0.7	0.1	1.0	0.9	0.3	0.9	1.0	1.0	0.0	2
	11,7	0.7	0.3	0.9	0.7	0.6	0.6	1.0	0.7	-0.3	1
	17,15	0.7	0.5	0.9	0.7	0.5	0.9	0.7	0.7	0.0	0
	17,13	0.7	0.5	0.9	0.3	0.5	0.9	0.7	0.9	+0.2	1
Sab-Sa	11,9	0.7	-0.7	0.7	-0.9	0.5	0.9	0.7	0.6	-0.1	2
	11,7	0.7	-0.3	0.3	0.3	0.5	0.5	0.8	0.9	+0.2	3
	17,15	0.6	-0.6	0.7	-0.3	0.5	0.9	1.0	0.7	-0.3	2
	17,13	0.6	-0.7	0.9	-0.5	0.1	0.9	0.6	0.7	+0.1	3
Task mean	ρ_μ	0.68	-0.27	0.84	0.25	0.36	0.74	0.76	0.80		
Task s.d	ρ_σ	0.18	0.47	0.21	0.53	0.31	0.35	0.18	0.14		
Rejections	-	0.05	0.40	0.05	0.25	0.20	0.05	-	-		

Table 7.6. Rank score permutation evaluation - detailed results from the sonar mosaic imagery, showing Spearman's rank correlation coefficients for the different methods over a range of classification tasks and parameters.

Mosaic	Eu	Ca	Bh	CS	KS	KL	Co	Co+
ρ_μ	0.68	-0.27	0.84	0.25	0.36	0.74	0.76	0.80
ρ_σ	0.18	0.47	0.21	0.53	0.31	0.35	0.18	0.14
Waterfall								
ρ_μ	0.59	-0.42	0.67	0.31	0.29	0.70	0.60	0.71
ρ_σ	0.38	0.43	0.28	0.47	0.50	0.26	0.28	0.28
Outex								
ρ_μ	0.09	-0.09	0.60	0.08	0.14	0.49	0.30	0.57
ρ_σ	0.53	0.49	0.39	0.56	0.25	0.54	0.42	0.57

Table 7.7 Mean, ρ_μ and standard deviation, ρ_σ , of the spearman's rank correlation coefficient for all feature parameterisations and classification tasks over all data sets.

Compared to Co, $\Delta\rho$ indicates an improvement in the correlation of Co+ in 45% of the test cases. There was no change in 25% of cases and a reduced correlation in 30%. It is not immediately obvious if there is any pattern in the improvement or reduction in correlation, with respect to the number, N , of rankings rejected.

By storing quality control information, performance profiles of the methods can be built up progressively. The usefulness of Ca within the committee may be questionable on this task, as the proportion of rejections due to this measure (0.4) is much higher than the others. CS has the greatest variability (lowest robustness), which, according to the distribution of the rejection instances in table 7.6 may be more unstable with respect to classification task than changing feature parameters. Eu, Bh and KL have far lower frequencies of rejection that are not correlated to each other. If so desired, rankings from each measure could be weighted, according to their reliability history, and form the consensus from the weighted contributions instead of using a binary select/reject criteria. As an illustrative example, first removing Ca, the weighting would be 0.29 Eu + 0.29 Bh + 0.07 CS + 0.06 KS + 0.29 KL for the mosaic imagery, based on the rejections in table 7.6.

All data	Eu	Ca	Bh	CS	KS	KL	Co	Co+
Property (iii)	2	0	3	0	3	3	3	3
Property (iv)	1	0	3	0	1	1	3	2

Table 7.8 Properties (iii) and (iv): Number of times the properties are satisfied for all data sets.

Referring to figure 7.6 and using the same approach as in section 7.1, one point is assigned to a method if it exceeds the overall mean of ρ_μ (property (3)) and one point if it has a standard deviation lower than the overall mean of ρ_σ (property (4)). Points allocated are summarised in table 7.8. The joint winners are Co and Bh considered over all feature parameterisations, classification tasks and data sets. Co+ and KL also demonstrated good overall performances.

The consensus approaches can improve the rank correlation and reduce variability in the ranking (hence improve the robustness) over multiple feature parameterisations and classification tasks. An improvement in robustness is particularly evident with Co+ on the sonar mosaic imagery. Clearly though, Bh and KL also have good overall performance as individuals.

Taking into consideration the total points awarded to each method for all of the properties (1)-(4), then the consensus approach, Co is the overall winner as its performance, on average, exceeded the group with respect to these properties, in all test cases. In other words, Co has demonstrated a relatively robust ranking of features and a measure of saliency that consistently correlates well with the achieved classification accuracy.

7.6 Evaluation, conclusions, recommendations and scope for further work

7.6.1 Evaluation

This case study was the most abstract of the four and concerned an investigation into the robustness of distance measures and novel distance measure committees for feature evaluation and ranking on sonar imagery.

The objectives were to establish which individual distance measure(s) are most reliable for feature evaluation and ranking on sonar imagery and to assess if there is any advantage in using a committee of distance measures for consensus evaluation and ranking on this task. A complex experimental (computational) framework was designed and developed for the task. The detailed results analyses indicated, under the prescribed conditions of the experimental sub-space, that a consensus feature evaluation and ranking approach improves the correlation of measured feature saliency to classification accuracy and the correlation of feature rankings to a baseline rank. In addition, under the same constraints, the variability of these properties could be reduced, over a range of feature parameters, classification tasks and data sets, i.e. the feature evaluation and ranking robustness were improved. Bhattacharyya's measure and the Kullback-Leibler divergence were identified as the best individual performers in the investigation. The objectives have therefore been successfully met.

A key advantage of the approach is that it facilitates an intelligent selection of a few individual discriminatory features and their parameters (including scales of analysis) to induce a classification model. Semantics of the features are propagated, so that classification results can possibly be related to the observed properties of the data and ground truth in a more meaningful way, compared to feature extraction methods such as PCA, as used by Preston (2009), where a large number of features are recombined with different weightings. It is also efficient, as the features are evaluated on relatively small, representative evaluation areas of the input imagery. This allows decisions to be made about the features before they are created on the entire (much larger) input space. There has only been one other study on the use of distance measures applied to sonar imagery feature evaluation, by Karoui *et al.* (2008). The application of a committee approach is

therefore the first of its kind in the domain. In addition to contributing as an application to the sonar imaging domain then, this work also contributes to the feature evaluation and selection domain. It is only recently (since about 2007) that issues of feature stability and robustness have been considered and only a handful of papers have so far been published on this important aspect of machine learning. Perhaps the most closely related work is by Saeys *et al.* (2008). However, the main difference is they used bagging and a homogeneous committee of feature selectors (i.e., the same feature selector, with re-sampled data), compared to an aggregation of results from different feature selectors in the thesis case study. An outcome of the research in this case study was a peer-reviewed publication in a leading computing science journal.

The achievements of this work could be extended by considering a notion of “diversity” in the committee of feature selectors and investigating the connection between diversity and the stability/robustness of the feature evaluation and ranking.

The design and implementation of the experimental harness was very challenging and its scope extends beyond the condensed subset of experimental dimensions and results presented in this case study. For instance, within the harness, it is possible to generate all feature subset combinations and evaluate the methods in relationship to feature combinations over different dimensions of feature space. As an example, the evaluation and ranking of any pairs of features could be considered. The harness is implemented as a suite of Matlab files and was developed *ab initio* by the researcher. Results generated by the harness required further organisation, processing and analysis. A number of further Matlab scripts were written to implement these processes, facilitating the generation of the final results and their visual representations.

7.6.2 Conclusions

The correlation between the estimated saliency of the features and classification accuracy was investigated, as was the robustness of ranking over different GLCM feature parameterisations and classification tasks on three data sets. On the sonar mosaic imagery, compared to the best individual, the consensus achieved a 67% reduction in the variability of the R^2 goodness of fit measure of saliency correlation to median classification accuracy. Considering the rank permutations, the unweighted consensus approach led to a 15% reduction in the variability of the correlation of rank score

permutations, relative to the best individual. This reduction was improved to 34% using ranking selection according to a binary threshold, in a weighting mechanism. A reduction in variability is evidence of an improvement in robustness over the prescribed feature parameterisations and different classification tasks. Considering properties (1)-(4), as evaluated over all tasks, parameters and data sets, the consensus approach is the best method overall, demonstrating a relatively robust ranking of features and a measure of saliency that consistently correlates well with the achieved classification accuracy. The Bhattacharyya distance measure and Kullback-Leibler divergence are the best individuals.

The aim of this chapter was to propose a novel framework for evaluating distance measures, for the purpose of evaluating and ranking parametric features on sonar imagery. The aim has been successfully achieved, since the framework has been proposed and implemented as an experimental harness. The objectives were to establish which individual distance measures are most reliable for feature evaluation and ranking and to determine if there is any advantage in using a committee of distance measures for this purpose. From the work carried out in the chapter and the conclusions, it has been demonstrated that the objectives have also been successfully achieved.

7.6.3 Recommendations

The investigations carried out showed that the Bhattacharyya distance measure and Kullback-Leibler divergence are the best individual distance measures for feature evaluation on sonar imagery. These individual methods are recommended, as the measured saliency of the features correlated consistently well with the achieved classification accuracy. It is also possible and potentially useful to calibrate the measured saliency with an estimate of likely classification accuracy, as in Harrison *et al.* (2011). For a more robust (but computationally expensive) method of evaluation, the consensus of multiple distance measures is recommended. Further, it is also recommended that in general, holistic approaches are taken towards the feature evaluation process, as in this case study, case study 1 (chapter 4), where a diverse range of methods were applied for the assessment of individuals and combinations of different features. It is clear that robustness of the methods by which the saliency and reliability of an individual feature or subset of features is measured is of fundamental importance

to the design of a meaningful, discriminatory feature space for a particular classification task.

7.6.4 Scope for further work

- A wider range of classification tasks on seabed imagery captured by different sidescan sonar instruments under varying conditions could be considered.
- A rigorous treatment of rank stability in the feature selectors and model stability at the validation stage would usefully provide a means of quantifying uncertainty in the estimated robustness of the feature selectors over the space of parameters and classification tasks.
- GLCM features are the most well known textural features in the sonar image classification domain, yet there are other very useful parametric features, such as the Gabor filter banks that could also be used as test cases.
- A notion of diversity in feature selectors could be considered and importantly, how to define and measure it. It would then be possible to build different committees of feature selectors and investigate if there is any connection between the diversity of the committee and the robustness of the features and feature parameters selected for a particular classification task.

So far in the thesis, much of the experimental work has involved using small sections of sonar mosaic imagery. This is common practice in the domain and many published studies use this approach. However, discrimination of interpretative targets in a much larger, real-world sidescan sonar mosaic image presents a number of significant technical challenges, some of which are addressed in the following chapter.

Unsupervised classification of sonar imagery

Chapter 8

Contents

8.1 Introduction

8.1.1 Important note

8.2 Previous work

*8.2.1 Problem background and solution outline**8.2.2 Sidescan sonar image classification**8.2.3 Ensemble/fusion approaches*

8.3 Proposed feature-distributed process and methods

*8.3.1 Sampling strategies**8.3.2 Image block processing scheme**8.3.3 Feature creation**8.3.4 Model induction**8.3.5 Independent classification channels**8.3.6 Post processing of output blocks**8.3.7 Fusion of classifications**8.3.8 Assigning class labels*

8.4 Experimental work

*8.4.1 Experimental aim**8.4.2 Data*

8.5 Results and discussion

*8.5.1 Inspection of the specimen classification**8.5.2 Single channel supervised and unsupervised approaches**8.5.3 Unsupervised fusion approach**8.5.4 Overall discussion*

8.6 Evaluation, conclusions, recommendations and scope for further work

*8.6.1 Evaluation**8.6.2 Conclusions**8.6.3 Recommendations**8.6.3 Scope for further work*

8.1 Introduction

The research presented in this chapter concerns a “demonstration of concept” case study. It focuses on a context limited exploratory investigation into the feasibility of a novel process and methods for the fully automated classification of subjective targets in real-world sonar mosaic imagery. The main underpinnings of the case study are;

- **Motivation.** In an industrial processing context, the real-world mosaic images are large, ranging from about 10 MB to 100’s of MB. The goal is to fully- or semi-automate the manual interpretative process, achieving a plausible classification outcome, similar to what a human might produce but in a fraction of the time, i.e., in minutes as opposed to days.
- **Synthesis.** Many of the ideas and methods, such as feature creation and evaluation, investigated during the course of the research are brought together as components of a coherent system, to achieve the goal.
- **Design.** Dealing with real-world imagery presents a number of additional challenges, particularly in standardising the unsupervised process across multiple image blocks. A novel hybrid process using heterogeneous machine learning algorithms is designed to surmount this challenge. A heuristic for fusing multiple unsupervised feature channels (models) is also developed. The efficacy of the unsupervised process is demonstrated in the case study by comparison of results with manual and fully-automated supervised classifications.

The overall aim of this work then is to devise a novel process for classifying subjective targets in sidescan mosaic imagery. The first objective is to devise a novel, hybrid unsupervised approach, using heterogeneous machine learning algorithms and a heuristic for fusing unsupervised classifications. The second objective is to evaluate the unsupervised process by comparison with a fully automated, supervised machine approach and a specimen class map of sediment types and therefore demonstrate the plausibility of the process and its desirable properties for the task.

Research in this chapter has an empirical, quantitative component but compared to the previous chapters, it is more qualitative and descriptive in nature. As pointed out earlier, using mixed methods provides a more holistic view. There is no exact “answer” to be achieved but through experimentation with the process and methods it is possible to eliminate ineffective components and arrive at an effective system configuration for achieving a reasonable classification outcome in the application context.

The remainder of the chapter is organised as follows. Section 8.2 outlines the context of the problem and describes some related work. In section 8.3 the overall process, and novel components are proposed. The experimental campaign is described in section 8.4. A selection of results are analysed and discussed in section 8.5. Finally, conclusions and scope for further work are presented in section 8.6.

8.1.1 Important note

The application scenario in this case study concerns sediment discrimination in sidescan sonar mosaic imagery. It is a contingent application, using resources that were available for a demonstration of the process, instead of the intended Sabellaria target discrimination. An assumption is made for the purposes of this case study, that intensity and texture in the mosaic image have an unknown relationship to the distribution of sediment targets. Sidescan imagery can be used for sediment grain size classification, as in Collier and Brown (2005) although an angular dependency method using MBES backscatter is often employed for this purpose, as in Hughes Clarke *et al.* (1997) and Fonseca *et al.* (2009). In the application considered here, despite not being as originally intended some curious, serendipitous findings arose and it has therefore been successful in this respect. It also demonstrates the versatility of the process and intimates the potential for transferability, regarding various target discrimination applications.

8.2 Previous work

8.2.1. Problem background and solution outline

Sidescan sonar is one of the most widely used acoustic remote sensing platforms in marine environmental surveys. Mosaic imagery, produced from the processed swath of

backscatter signals, contains tonal (intensity) and textural information related to the roughness characteristics and spatial distribution of different regions on the seabed. See, Blondel (2007), Fish and Carr (2001) and Le Bas and Huvenne (2009). It has proven utility for high-resolution imaging of biogenic structures (Birchall, 2007, Degraer *et al.* 2008, Hendrick and Foster-Smith 2007, Limpenny *et al.* 2010.) Natural features on a variety of spatial scales, including sand dunes and rock outcrops are often distinguishable in the mosaic. Whilst most of the image information content is textural (Blondel and Gómez-Sichi 2009), distributions of well-sorted sediment particle sizes, under certain conditions, are strongly correlated to sidescan sonar backscatter intensity (Collier and Brown, 2005) and hence to pixel radiosity values in the mosaic image. Thus, features such as descriptive statistics and local intensity histograms can be used to represent and discriminate between surface sediments. Sidescan sonar imagery acquisition is usually accompanied by a ground truth sampling campaign, typically comprising collection of sediment grab samples for particle size analysis (PSA) and photographic or video capture of benthic organisms and consolidated structures.

Due to their qualitative nature, mosaic images are often analysed manually, by human interpreters. In fact, the human eye is generally accepted as being the best delineator of the regions in a sidescan image (Limpenny *et al.*, 2010). During the interpretation process, the extents of relatively homogeneous regions of textures and intensities in the image are analysed, together with information derived from the ground truth. The aim of this manual interpretative process is usually to create a seabed class map, by interpolating between (or extrapolating from) known ground truth points and the corresponding regional image properties associated with these points. Ground truth partially validates the interpretation, although as discussed in chapter 4, section 4, due to the localisation of the ground truth, the classes and their boundaries are always uncertain and ambiguous, regardless of the processing stages and whether the map was created in a human or a machine process. There may also be also a semantic gap between the true meaning of a descriptive label, the local “real” attributes of the seabed the label is describing and the representation on the class map. It is possible to represent any number of characteristics as classes on a single map but five or six classes are usually sufficient for depicting the main substrate types (gravel, sand, mud), together with areas of say, outcropping and biogenic structures or similar regions of interest, depending on the classification objective. The classification objective greatly

affects the outcome of the process. For instance, objective ground truth representations such as grain size distributions and their parameters may be used to represent the ground truth in different ways, according to different classification schemes.

Manually producing a class map is a time consuming and expensive task, requiring the effort of at least one human interpreter over many days. It is also well known that the process is subjective and inconsistent, since different human analysts yield differing interpretations (Blondel 2007, Johnson and Helferty 1990, Martin *et al.* 2006) However, qualitative image capture using the comparatively low-cost, high resolution sidescan sonar and its subsequent human interpretation remains an important, established practice in the marine surveying industry and is likely to continue for the foreseeable future. It would be beneficial then, to devise a means of improving the efficiency and reliability of the industrial human processing pipeline by making use of machine learning technology. The idea of automating this process is not new. Czarnecki (1979) published one of the earliest investigations into the application of pattern recognition techniques to the classification of sidescan sonar imagery. Since then, other supervised and unsupervised feature based approaches have been proposed, such as, Blondel *et al.* (1998) (TexAn), Kalcic and Bibee (2004), Preston *et al.* (2004) (QTC), Samiee and Rad (2008), and Sun and Shim (2008). To date though, there are still no universally accepted methods or procedures. It remains a formidable problem and presents a number of technical challenges. Relatively few of the proposed approaches have matured into established sonar image processing implementations. Of these, perhaps the best known are from QTC (commercial processing software), and the TexAn package (academic processing software).

A generic problem with supervised learning approaches in any domain is the potential introduction of label noise during the sampling and training process, giving rise to an erroneous classification model, as in (Lawrence and Scholkopf 2001). Selection of sonar image training regions can be through a qualitative visual inspection, as in Blondel and Gómez Sichi (2009) or by using the ground truth points as seed locations (one of the approaches taken in this case study). Visual inspection and selection is very effective when there are a few distinct and homogeneous textural classes such as smooth sediments and rock. Small representative regions of the classes can be chosen from the image and used for feature (and feature parameter) evaluation and validation,

as in TexAn. This approach was also used for the feature evaluation investigations in the previous two chapters of the thesis. The reason this approach works well is because the regions of interest have been pre-identified using expert knowledge. It is then a case of evaluating the most discriminatory features for the classification objective and inducing a machine model for classifying the targets in the entire image. The manual selection of visual training regions may not work as effectively though, when there are several classes without clear spatial demarcation, for instance, transitions from one type of smooth sediment to another with a continuous intensity gradient.

Using ground truth points as seed locations facilitates a fully automated, supervised training and classification process, as no manual intervention is required. Due to positioning uncertainties though, the image region sampled may not correspond to the class sampled at the true ground truth location. The chance of introducing label noise is high, in this case, as positional uncertainty can exceed 35% of the water depth, as documented in the survey of Collier and Brown (2005). If the seed locations for training the classifier are not representative of the class in terms of intensity and texture, then an erroneous model will be induced and classification results may be poor.

Unsupervised (or clustering) approaches, as in Preston *et al.* (2004) and Preston (2009) do not depend on a priori knowledge, consequently, label noise and variability in human selected training regions is not a concern. At the end of an unsupervised process, meaningful labels can be assigned to the partitioned regions (provisional classes) by human inspection of the segmentation in relation to the observed regions in the mosaic imagery and consideration of any ground truth points lying within the segmented regions. It is possible to automate this process and a procedure for doing so is also presented in the case study. There are three key problems to address when implementing an unsupervised approach;

- how to choose an appropriate number of classes (clusters), either specified by a human user or determined automatically in some way,
- how to select the best clustering model for the number of classes and

- how to assign ground truth class labels to the unsupervised solution.

A further consideration, when dealing with the unsupervised classification of large, real-world sonar mosaics is the processing efficiency, which can be improved by dividing the image into numerous independent blocks. Block processing constrains the quantity of pattern instances created and processed in memory to a manageable size. An important, highly relevant practical problem though, is that because the blocks are independent, without careful treatment, inconsistent class partitioning and non-corresponding labeling of patterns can arise. The problem is addressed in this work by implementing a novel pre-clustering process that assigns provisional class labels to training patterns. The best solution set is selected through an internal validation process. This pattern set is subsequently used to induce a global, Naïve Bayes classification model (see, for example, Duda *et al.*, 2001, p 13.) Thus, the same unsupervised classification model is applied on each block and patterns created on the blocks are range normalised to the pattern set used to induce the global model. This hybrid approach ensures consistent results on each of the independent image blocks within the separate feature spaces and probabilistic support is provided by the classifier, for each classified pattern on the individual blocks.

Different unsupervised models can be induced in independent feature spaces, known as feature-distributed clustering (Hadjitodorov *et al.* 2006.) However, the predicted provisional class labels output from each feature channel will not correspond – an inherent problem with unsupervised ensembles (Strehl and Ghosh 2002). In this work, the unsupervised output channels are fused together by considering relative frequencies of label permutations in the joint distributions of spatially co-located class labels, together with their numerical support (posterior probabilities.) Permutations with the “*top-n*” relative frequencies are selected and the corresponding sets of points are relabeled, in their spatial locations to produce a provisional, *n*-class map. Finally, meaningful class labels are assigned to the provisional classes, based on the conditional probability of occurrence of a particular ground truth class type within the spatial coverage of the provisional class region. The proposed unsupervised process is evaluated by comparison with a supervised classification and a manual interpretation of the imagery and ground truth data.

8.2.2 Sidescan sonar image classification

There have been several proposed approaches to sonar image classification published recently, employing a wide variety of feature creation techniques and machine learning methods. Although as yet, there are no definitive review works, some of the approaches are covered in Simard and Stepnowski (2007), Hughes Clarke *et al.* (2009) and Schumann *et al.* (2010). An outline of some related previous work is recapitulated here, as much has already been covered in earlier chapters.

Karoui *et al.* (2008) state that a number of studies have indicated a fusion of features from different feature creation methods could improve texture characterisation. They investigated a fusion of Gabor filter bank responses and co-occurrence distributions to characterise the texture of sidescan sonar images. Two other recent studies where Gabor filter banks were applied are those by Samiee and Rad (2008) and Sun and Shim (2008). The feature creation process of Samiee and Rad (2008) used the filter banks to generate a sub-image from each channel followed by morphological closure. Whilst their method is robust to noise and the data acquisition processes, their results were reported on small, clean image regions with distinct textural contrasts and relatively well-defined textural boundaries. Sun and Shim (2008) combined Gabor filter bank features with a model-based feature, the fuzzy fractal dimension (FFD) in a hybrid fusion method. Using a Multi-Layer Perceptron (MLP) classifier, they found overall, the classification accuracy was higher with the fusion approach, compared to using the Gabor filter bank or FFD features in isolation. For more comprehensive coverage of the Gabor filter bank, refer to chapter 6 of the thesis.

Standard, k -Means clustering is a popular unsupervised technique with researchers in the domain. Whilst k -Means is quite straightforward, it does have some limitations. It cannot handle non-globular clusters or clusters of different sizes and densities and performance is also significantly affected by outliers (Tan *et al.*, 2006.) Blondel and Gómez Sichi (2009) claim that more recent applications have tended to use the k -Means algorithm. Their tests indicated that a simple Euclidean metric was sufficient for the classification application in their research. However, the user needs to have an idea of how many different clusters are expected in the data, as this must be specified before

running the k -Means algorithm. Unlike a SOM, k -Means cannot determine the number of clusters automatically.

8.2.3 Ensemble/fusion approaches

An ensemble or committee comprises a collection of machine-learning models and decision fusion strategy for combining the model predictions (Dietterich, 2000). See Sinha *et al.* (2008) for a recent survey of decision fusion methods. Multiple models are created from different samples of training data or within different feature spaces and can be supervised (classifier ensemble) or unsupervised (cluster ensemble). Although supervised ensembles are in common use and have many real-world applications (Pal 2007), they have received very little attention within the seabed classification domain.

Unsupervised (cluster) ensembles (Strehl and Ghosh, 2002) are generally more difficult to implement than supervised ensembles due to the problem of class label correspondence. It is not possible to use a simple voting mechanism, as in the case of a supervised ensemble, since provisional class labels output from each channel are not associated to the same class. So, in addition to the problems of determining the number of clusters, the best clustering solution and assigning meaningful labels to the provisional classes, it is also necessary to have a strategy for aligning the labels and combining the label vectors.

In the context of sidescan sonar imagery, there have been few ensemble based approaches proposed. Martin *et al.* (2004) and Martin (2005) studied the fusion of classification results based on models created with different types of textural features was investigated. Generated patterns were input to a committee of MLP's, with each committee member associated to a discrete feature input channel. Channel outputs were combined in two decision fusion strategies based on evidence theory and distances. Sun and Shim (2008) also used an MLP in their hybrid fusion process, however, the work of Sun and Shim (2008), Martin *et al.* (2004) and Martin (2005) was limited to classifications on small patches of sonar imagery of relatively distinct classes. Dartnell and Gardener (2004) used an unsupervised ensemble technique for seabed classification based on a fusion real-world MBES bathymetry and backscatter imagery. A rule-based approach was implemented with a hierarchical DT classifier to combine four

unsupervised classifications in the ERDAS Imagine¹ software package. The rules were based on the known classes in the ground truthed areas.

The approach used in this case study assumes nothing about the relationship of the clusterings to the ground truth until after the clusterings have been combined. It is treated as a cluster ensemble problem, as defined by Strehl and Gosh (2002).

8.3 Proposed feature-distributed process and methods

The proposed, hybrid, feature-distributed, unsupervised ensemble approach presented here is the first of its kind in the context of interpretative sonar image classification on real-world mosaics. The hybrid process is applied only in the unsupervised case and regardless of whether a single or multi-channel (ensemble) approach is used.

It is evaluated by comparison with a supervised machine classification and a specimen classification (class map.) The supervised and unsupervised approaches require different processing methods. Eight of the more important stages and corresponding methods are specified in table 8.1 and described in the following subsections. The abbreviation GT is used for ground truth and *S*, *T*, *E*, *L* for Statistical, Texture, Edge detection and Lacunarity feature channels, respectively.

¹ <http://geospatial.intergraph.com/products/ERDASIMAGINE/ERDASIMAGINE/Details.aspx> [accessed 05-12-2012]

Stage	Supervised	Unsupervised
1. Image sampling	GT sample station locations, 8- <i>NN</i> kernel neighbourhoods.	Multiple, random locations, 8- <i>NN</i> kernel neighbourhoods.
2. Block processing	Sequential, non-overlapping blocks.	Sequential, non-overlapping blocks.
3. Feature creation	Multi-channel $\{S, T, E, L\}$	Multi-channel $\{S, T, E, L\}$
4. Model induction (Naïve Bayes)	Labelled samples from GT stations.	Pre-clustering. Induce probabilistic classification model with clustered patterns and cluster labels.
5. Independent classification channels	Predict class label and posterior probability for each pattern instance.	Predict provisional class label and posterior probability for each pattern instance.
6. Post processing (smoothing)	Output blocks smoothed using categorical distribution of <i>k-NN</i> in intermediate map space.	Output blocks smoothed using categorical distribution of <i>k-NN</i> in intermediate map space.
7. Fusion (DsI-DeO)	Weighted voting	“Top- <i>n</i> ”, relative frequencies in joint distributions.
8. Assign class labels	Not applicable	Conditional probability of GT class occurrence in spatial coverage of provisional class.

Table 8.1. Summary of main processing stages and methods for supervised and unsupervised ensembles.

8.3.1 Sampling strategies

Before any image sampling can take place, a binary mask is created to flag areas of imagery that should not be processed, such as no-data, nadir² lines or any areas of poor quality.

In the supervised approach, the geographic coordinates of the Ground truth (GT) physical sampling stations are first converted to local image coordinates. The image coordinates assigned to each GT location depend on image resolution and the size of the feature creation kernel, since, for efficiency, the image is down-sampled during the feature creation process. The eight nearest neighbours³ (8-*NN*) of the GT image neighbourhood covered by the kernel are also sampled, providing a maximum of 9 seed patterns at each location. Assumptions made are (1) the neighbouring image patches represent a homogeneous class and are thus assigned the same class labels as the GT point and (2) the true location of the GT point is within the area bounded by the 8-*NN*. Image regions that have been masked are not sampled.

The unsupervised approach uses multiple (5) random samplings of arbitrary image locations with the number of non-replacement coordinate pairs, $q_i \in \{25, 50, 100, 200\}$. Again the central location and its 8-*NN* are sampled. No assumptions are made about the class of the central location or the neighbourhoods under the feature creation kernel. Re-sampling is required in order to carry out the internal validation process to select the number of clusters and the best clustering solution on each feature channel.

8.3.2 Image block processing scheme

Block processing, see for instance, Lalgudi *et al.* (2000), reduces computational memory overheads and processing time by reading image blocks from disk into memory instead of the entire image. The algorithm implemented divides the image into non-overlapping blocks that are processed sequentially. Block size is constrained by

² The nadir is the point on the seabed vertically beneath the transducer. It traces out an imaginary line as transducer moves.

³ Nearest neighbours (*NN*) in this case study refers to neighbouring points in a fixed arrangement in a regularly spaced grid, rather than an unordered collection of neighbouring points in a feature space.

two prescribed parameters, firstly, the number of pattern instances to be processed on each block, N_p , and secondly the feature creation kernel size, k . For an input image of size m rows \times n columns, unless m and n are exactly divisible by $k\sqrt{N_p}$, zero padding at the image edges is required to produce a set of identically sized blocks covering the entire mosaic image. The binary mask is also delineated congruently to the mosaic image and each mask block is read into memory at the same time as the corresponding image block so that patterns are not generated on masked regions. At the end of processing, the blocks of class labels and class posteriors output by each feature channel are concatenated.

8.3.3 Feature creation

Although any number of independent feature channels can be implemented for model induction and class prediction, four were initially considered in the preliminary experimental investigation; statistical (S), textural (T), Edge (E) and lacunarity (L). In the detailed experimental investigation presented in section (8.5.2), just two of these, S and T , are selected for more in-depth consideration. The reason for this is the preliminary investigation into the results from the feature channels led to a rejection of the E and L channels for this specific task. Using just two channels in the detailed analysis also simplifies the processing and post-processing stages and the exposition of the analysis.

Statistical features (S)

The simplest (and possibly most effective and efficient) features for discriminating regions are statistics based on backscatter intensity. A basic (and rather naive) assumption is that lighter and darker regions of the imagery correspond to areas with different backscatter intensities and thus different material properties or sediment types. Combinations of statistical features, such as mean, median and standard deviation can be used together to partition backscatter imagery, in relationship to the substrate grain sizes, as in Collier and Brown (2005). A two-dimensional feature vector is applied in the test case, comprising, local mean and local standard deviation at kernel sizes (resolutions), $k \in \{5, 11, 15\}$ pixels.

Textural features (T)

Based on the promising results obtained in some of the previous work (chapter 6), a Gabor filter bank (GFB) is designed and implemented to generate the textural features in this test case. Task dependent tuning is required as there are several parameters and configurations as well as numerous different types of features that can be derived. See, for instance, Bianconi and Fernández (2007), Grigorescu *et al.* (2002) and Clausi and Jernigan (2000). The feature derived here is the magnitude response. Rotational invariance is implemented by selecting the directional channel at a specific centre frequency, with the maximum response. Radial separation of the filters is set at $\pi/6$ radians, i.e., the number of orientations, $|\Theta| = 6$. The size of the computational kernel varies according to the specified resolutions. Initial centre frequency is, $f_0 = 0.15$. The half-octave frequency progression yields five band-pass centre frequencies, $|\mathbf{F}| = 5$ and hence a five-dimensional feature vector. An isotropic Gaussian envelope is used, with standard deviation, σ , proportional to the computational kernel resolution, $\sigma = (k-1)/2$.

Edge features (E)

An edge detection filter is applied, as described in section 6.4, using a Sobel mask. Thresholds of 0.04, 0.07 and 0.10 are used and therefore three features are generated.

Lacunarity features (L)

The commonly applied gliding box algorithm of Allain and Cloitre (1991) is implemented, to generate multi-resolution feature values within a specific kernel size. The number of lacunarity features generated, β , satisfies $\{\beta : \max(2^\beta) < k\}$. In the test case, the maximum number of features is three as the maximum kernel size is $k = 15$.

Feature group	Parameters/configuration	Nr. of features
Statistics (S)	Local mean and local standard deviation.	2
Texture (T)	$ \Theta = 6$, $ F = 5$, $\sigma = (k-1)/2$, $f_0 = 0.15$, max. RI.	5
Edge (E)	Thresholds; 0.04, 0.07, 0.10.	3
Lacunarity (L)	Binarisation threshold; 0.04.	2 or 3

Table 8.2 Summary of feature channel parameters

8.3.4 Model induction

Model induction for the supervised case is straightforward. Image samples automatically selected around the GT points are used to create training patterns for the different feature channels. The pattern instances and their associated class labels are then used to induce independent Naïve Bayes classification models in each channel. All feature values are range-normalised on [0 1] using the extrema of values in the training pattern instances.

Some ingenuity is required for creation of the unsupervised models. As pointed out earlier, some studies on unsupervised classification (and cluster ensemble approaches) use relatively small image samples, and apply the unsupervised approach to a global pattern space. However, real-world sonar images are large, typically 10 MB – 100's of MB and hence the use of a block processing scheme here. Each block is independent, so it is not possible to apply an unsupervised approach to each block without applying some form of global constraints. Otherwise, pattern partitioning and normalisation would be inconsistent, not to mention the label correspondence problem that would arise. To address this problem, a pre-clustering process is used to identify potential classes, based on multiple random sampling of patterns derived from the global image space. The best solution, established in an internal validation process is then used to induce a global, probabilistic classification model in each feature channel. A number of steps are involved and these are summarised in figure 8.1.

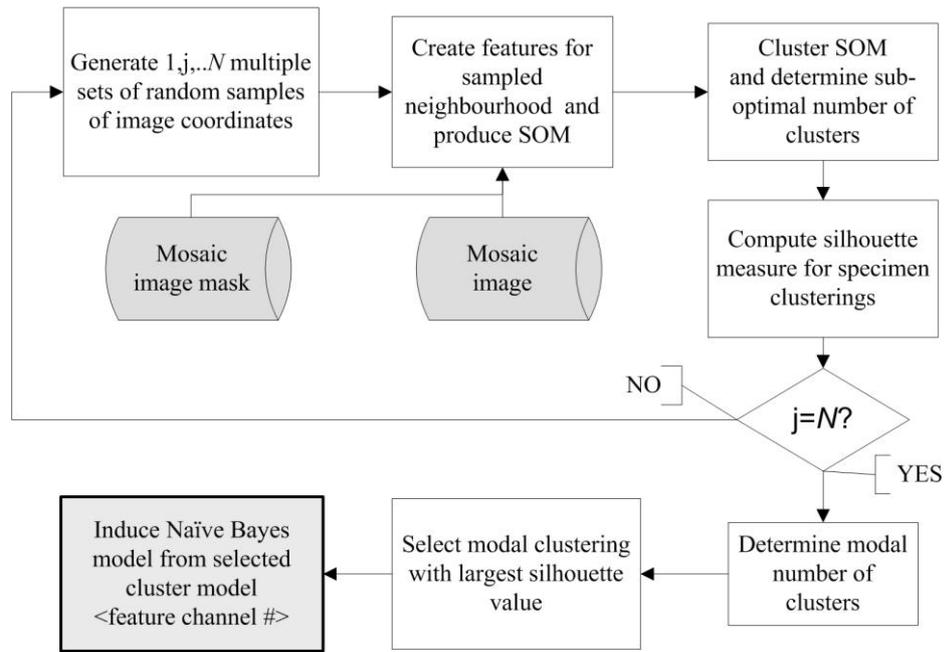


Figure 8.1. Simplified unsupervised model induction process.

Step 1. N samplings (n_j) of Q pairs (q_i) of random image coordinates are generated, subject to masking constraints.

Step 2. Pattern instances are created for each feature channel from the image neighbourhoods sampled at each set $n_j(q_i)$ of image coordinates.

Step 3. Each $n_j(q_i)$ is converted into a SOM for the feature channel, which is then clustered.

Step 4. The modal number of clusters generated over multiple clustering runs $N(q_i)$ from step 3, is determined and this is used to define the number of provisional classes present in the data set.

Step 5. From the pattern sets $n_j(q_i)$ corresponding to the modal number of clusters determined in step 4, the set that yields the clustering with the best mean silhouette measure (Rousseeuw 1986) from separate k -means clusterings in the original feature space is selected. This is the optimal cluster model in terms of i)

number of clusters, ii) number of data samples and iii) the specific set of random samples used to induce the patterns and the clustering, from the multiple re-sampling. The pattern instances in this set are assigned provisional class labels $\{1, \dots, 6\}$ in each feature channel, according to the cluster to which each pattern belongs.

Step 6. Naïve Bayes models are induced on each channel as in the supervised case, using the pattern instances and provisional class labels identified in step 5.

This validation process is rapid and takes less than one minute to induce and select the unsupervised classification models, with $N = 5$, $q_i = 100$. The end results are global, probabilistic classification models, optimised on the sample space. The unsupervised models can now be applied consistently, for pattern label prediction on each independent image block. The internal validation process helps to ensure that the best model is chosen with respect to the number of clusters and the best clustering solution for the identified number of clusters. The result is channel specific, since the number of clusters and the best solution will depend on the feature kernel that has generated the patterns.

8.3.5 Independent classification channels

The class prediction process is identical for the supervised and unsupervised approaches. Although, using the supervised approach, class labeling is consistent across the feature channels. Each feature channel has its own independent classification model that assigns a class label and posterior probability to the pattern stream generated from the image blocks. The pattern values created on the blocks are range normalised to the selected model pattern values. Values falling outside of the range of the model are clipped to 0 or 1. The output streams of class labels and posteriors are rearranged into rectangular grids, corresponding to the current image block. This requires sequencing with the mask flags to ensure that the classified points, their class labels and posteriors are assigned to the correct spatial locations in the down sampled output blocks.

8.3.6 Post processing of output blocks

Following block concatenation, a filtering stage is applied to the intermediate class map, subject to masking constraints, to produce a smoother result. This process involves determining the relative frequency of occurrence of class categories at each map point and its 24-*NN*. The modal class label is assigned to the point at the centre of the moving neighbourhood window. In the case of a multi-modal distribution the label assigned is that of the class with the greatest mean posterior probability on the local neighbourhood. This process is the same for the supervised and unsupervised approaches.

A weighted posterior probability, p is computed at each filtered point, for the assigned class label, ω^* . It is calculated as the product of the relative frequency of ω^* , in the 24-*NN* empirical distribution of the categorical classes and the mean posterior of ω^* over the neighbourhood. Since, if the 24-*NN* is mostly class ω^* and the mean posterior of ω^* is high, the probability that the class label assigned at the point is ω^* must also be high.

8.3.7 Fusion of classifications

Combining classifications is a form of decision fusion and the fusion methods concern the combination of individual beliefs into a consensus belief (Sinha *et al.* 2008). There are many applications and techniques (Mitchel, 2007, Sinha *et al.* 2008). In the unsupervised and supervised strategies applied in this work, the fusion model is: “soft” decision input \rightarrow decision output (DsI-DeO) according to the scheme of Dasarathy (1994.) Support, in the form of class posteriors is available on each input channel, hence the soft decision input. Input channels are streams from the post-processed blocks, described in 8.3.6. The decision output also has a support of class posteriors, further modified according to the inputs.

1. Supervised decision fusion strategy

In the supervised case, since class labels have already been assigned, the channel labels correspond to the same class but in different feature spaces. The supervised fusion strategy uses a weighted voting scheme, where the weight is the posterior probability of

the (filtered) class label instance. In the two-channel test case considered, (for example, after channel pre-selection of S and T), the posterior probability of a class label from intermediate map (channel), S , is p_S and from map T , p_T . If the class labels from collocated points in each map are the same, the decision fusion process assigns the common class label and the mean of the posteriors. If the class labels differ, the class label with the greatest posterior, $p^* = \max\{p_S, p_T\}$ is assigned as the output class, with a new posterior, $p' = p^* / (p_S + p_T)$. The modified posterior, p' sensibly reflects the increased uncertainty that the 'correct' class label has been selected, when the beliefs differ. This strategy is easily extended to several channels.

2. Unsupervised decision fusion strategy

Each input channel is similar to the supervised case in that it is a label stream with posterior support. There are three complications though, (1) the association between the channel input labels and the class is unknown, (2) the labeling in each channel does not correspond to the same entity and (3) the cardinality of the label sets on each channel is unequal. It is assumed that the ground truth is not available at this stage and can only be used after the decision fusion process has been applied to the intermediate maps. Further, it is assumed that no further information about the partitioning algorithm or the pattern instances is available, i.e., it is a cluster ensemble problem, according to the definition of Strehl and Ghosh (2002). A heuristic fusion strategy is applied as follows;

- (1) collocated points in each intermediate map (label vector, $\lambda^{(i)}$) are randomly sampled,
- (2) the joint distribution of the class label permutations of the samples is created,
- (3) the top- n permutations are selected, based on their relative frequencies in the joint distribution, subject to i) n does not exceed six and ii) permutations with a relative frequency less than 0.05 are not considered to be relevant,
- (4) a new map is created by relabeling all of the spatial locations according to the top- n selected label permutations of the intermediate maps and

- (5) a new posterior is computed as the mean of the class label posteriors in the permutation at the spatial location, since the point now has a probability of a particular permutation occurring. All points occupied by permutations with a relative frequency less than 0.05 are set to zero (the mask pixel value).

3. Channel selection

In the case of v -independent feature sets, there are 2^v-1 possible combinations of class map. Normalised Mutual Information (NMI) is one metric that may be used for pairwise comparisons of the intermediate clustering solutions from the channels, if some form of channel pre-selection is desired. NMI is defined as,

$$NMI(\Omega, C) = \frac{-2 \sum_k \sum_j \left\{ \frac{|\omega_k \cap c_j|}{N} \log_2 \left(\frac{N |\omega_k \cap c_j|}{|\omega_k| |c_j|} \right) \right\}}{\left\{ \sum_k \left(\frac{|\omega_k|}{N} \log_2 \left(\frac{|\omega_k|}{N} \right) \right) + \sum_j \left(\frac{|c_j|}{N} \log_2 \left(\frac{|c_j|}{N} \right) \right) \right\}}$$

(8.1)

Where Ω , C are two specimen clusterings to be compared, k , j are the number of clusters in Ω , C , respectively and N is the sample size. The NMI is a widely used information theoretic measure and estimates the reduction in uncertainty of channel C through knowledge of channel Ω . See for instance, Fred and Jain (2002), Vinh *et al.* (2010) and Kuncheva and Hadjitodorov (2004). NMI is defined on $[0, 1]$ with a zero value reflecting randomness (i.e. channel C contains no information about channel Ω) and one, identical (if channel Ω is known channel C is also known.)

8.3.8 Assigning class labels

The final stage in the unsupervised process is the assignment of semantic class labels to the provisional class labels of the top- n permutations. Given any provisional class, C_p , the ground truth class, C_g with the highest conditional probability of occurrence, $\max\{p(C_g | C_p)\}$ at the ground truth sampling points, within the spatial region of C_p , is assigned as the final class label for the region. For example, if a particular provisional class region, A , contains 7 instances of ground truth class X and one instance of ground

truth class Y , region A is assigned the semantic label of ground truth class X . All of the posteriors for the points in the region are then multiplied by the relative frequency of occurrence of X in region A to reflect the uncertainty of the final label assignment.

8.4 Experimental work

8.4.1 Experimental aim

The aim of the experimental campaign is to generate visual and numerical results for evaluating the proposed unsupervised approach against a supervised process and a manual classification. The experimental test case makes use of publically available data from the United States Geological Survey (USGS), Open-File Report 99-396⁴, collected from the Inner Shelf, off the coast of Sarasota, Florida.

The objective is to create maps of the seabed sediment type distributions (specifically defined by the mean grain size distributions), with the supervised and unsupervised machine learning processes. An implicit assumption is that the features derived from the mosaic imagery will capture some useful information concerning the sediment distributions. Three types of multi-class, machine classifications are generated in the experiments, with a variety of configurations (the focus is on the statistical and textural feature channels after initial channel pre-selection).

1. Supervised, single channel: statistical and textural feature channels are considered individually.
2. Unsupervised, single channel: statistical and textural feature channels are considered individually.
3. Unsupervised, multi channel (feature-distributed ensemble): decision level fusion of feature channels.

⁴ <http://pubs.usgs.gov/of/1999/of99-396/> [accessed 01-06-2011]

8.4.2 Data

Data used in the experiments comprises a sidescan sonar mosaic tiff image (fig. 8.2(a)), the details of 152 ground truth sampling stations and a specimen class map of the sediment distributions (fig. 8.2(b).) The mosaic image is 4063×4617 pixels and has an anamorphic ground scale of 4.000 pixels per metre horizontally and 4.047 pixels per metre vertically. All ground truth stations can be classified according to ranges of the mean grain size (ϕ) determined from the station sediment sample PSA. The class labels and corresponding ranges of ϕ are given in table 2, together with the colour legend for fig. 8.2(b). All subsequent maps and charts use the same colour coding to represent the classes.

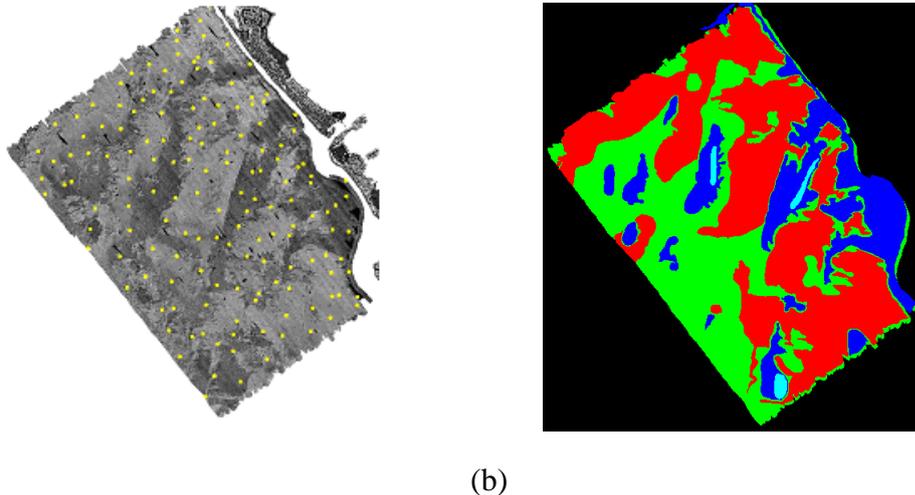


Figure 8.2. (a) Sonar mosaic showing locations of the 152 sampling stations (b) specimen class map, showing the estimated distribution of four different sediment grain size classes (see table 2 for legend.)

There are four supervised (S) classes, 1-4, in the table. Up to six unsupervised (US) provisional classes, 1-6 are assigned and the class labels correspond to the colour legend in table 8.3 but are unconnected to the four grain size classes, until the ground truth labels have been assigned at the end of the unsupervised process. Magenta and yellow can be assigned to a provisional (unsupervised) class at an intermediate processing stage but never to a ground truth class.

US	S	Mean ϕ	Description	
1	1	< 1.0	Very coarse sand and gravel	
2	2	1.0 – 2.0	Medium sand	
3	3	2.0 – 3.0	Fine sand	
4	4	> 3.0	Very fine sand	
5	-	-	-	
6	-	-	-	

Table 8.3. Sediment type definitions and colour legend for the class map in figure 8.2(b) and for all subsequent maps and charts. Magenta and yellow can only be assigned in the unsupervised case and never to one of the four ground truth class.

8.5 Results and discussion

8.5.1 Inspection of the specimen classification

A preliminary inspection of the ground truth class distributions and the specimen class map of fig. 8.2(b) provide some interesting insights. Figure 8.3 (a) shows the relative frequency of the four sediment classes recorded at the ground truth sampling stations, C_g . The samples are predominantly of class 1 (very coarse sand and gravel). Comparing this to the relative areas of the corresponding regions in the class map, C_R , the greatest regional coverage is class 1 (0.43), followed by class 2 (0.38). The proportions of the areas mapped as a particular class type are not the same as the relative frequencies of the ground truth class types, particularly so for classes 2 and 3.

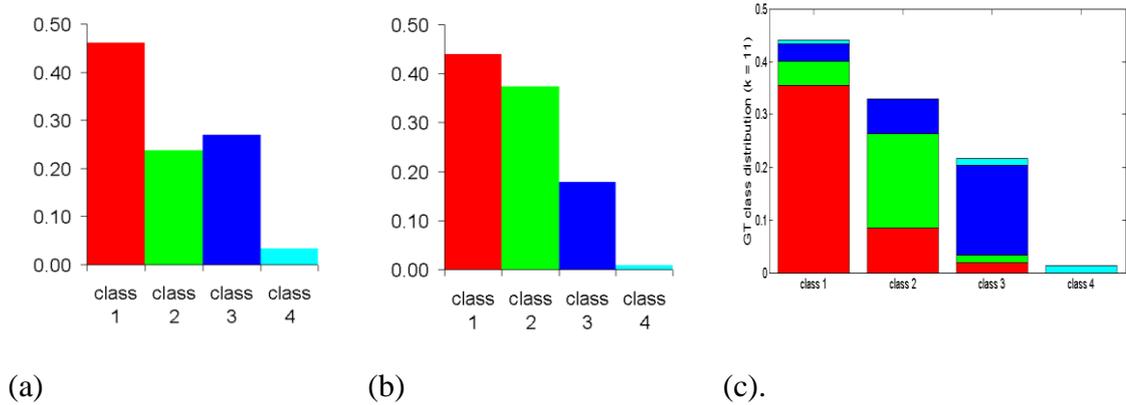


Figure 8.3. (a) Relative frequency of sediment class occurrences at sampling stations, C_g . (b) Relative areal coverage of the sediment class regions, C_R , determined from the class map in fig. 2(b). (c) Relative frequency of occurrence of a ground truth sediment class within the mapped sediment class region, $C_g \cap C_R$.

Class 2 occupies a relatively large area of the map, compared to the relative frequency of occurrence of the class type at the ground truth stations. Class 1, on the other hand, has a relative frequency of ground truth occurrences comparable to the relative areal coverage. Figure 8.3 (c) shows the relative frequency of occurrence of ground truth classes, within the corresponding classified regions in the class map, i.e., $C_g \cap C_R$. This demonstrates that the regions in the class map are impure. The region for class 2 has the highest impurity as it contains a lower, relative proportion of corresponding class 2 ground truth points. The class 4 region is 100% pure, it contains no ground truth points which are not of class 4, although Class 1 and Class 3 regions both contain some class 3 ground truth points. Whilst the regions do predominantly contain the associated ground truth class, they are really an inhomogeneous mixture of the different ground truth class types. Clearly there is some uncertainty here in the boundaries of the classes, their spatial coverage and the association with the ground truth. Classification accuracy per se, is not that important in this type of problem, since the true classes and their spatial distributions are unknown and unknowable. Thus it is important to appreciate, that in making any comparisons, one uncertain quantity is being compared to another uncertain quantity.

8.5.2 Single channel supervised and unsupervised approaches

In these experiments, specimen class maps are produced by the supervised and unsupervised machine approaches described previously, using the feature channels independently and various properties described and compared. Several preliminary experiments were run before commencing an in-depth treatment on a more focused experimental sub-space and progressing through the process and methods. In the preliminary experiments, it was clear that the unsupervised method was generally producing a greater number of (non-sparse) partitions, over the different feature spaces and parameters. The supervised approach had a tendency to produce predominantly binary classifications, sometimes with a dominant class and a sparser second class, as exemplified by the classification using Edge filters at a resolution of $k = 11$, in figure 8.4. By comparison, four potential classes were partitioned with the same feature channel, at the same resolution in the unsupervised approach.

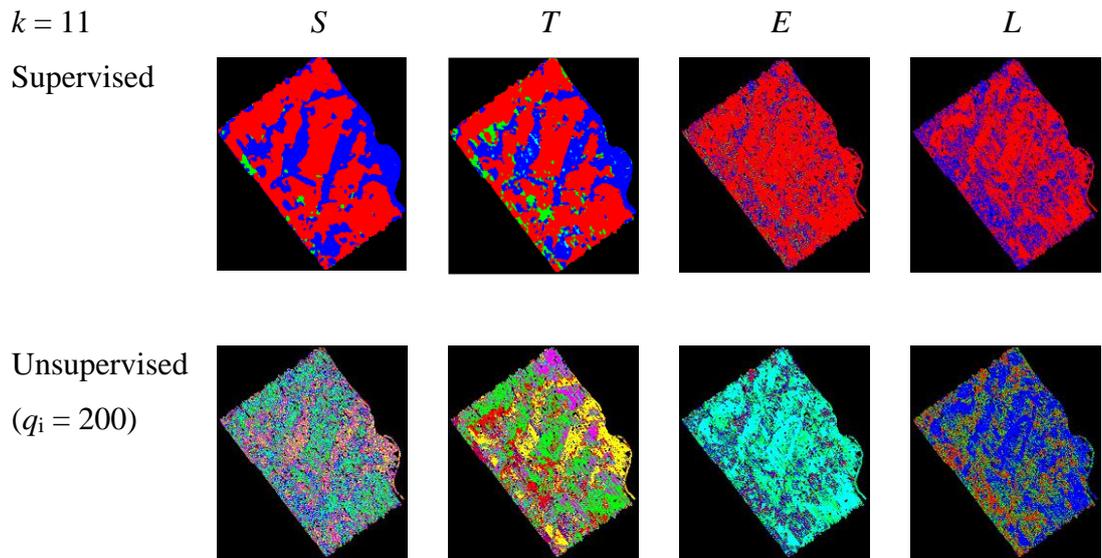


Figure 8.4 Examples of the intermediate class maps and partitioning produced respectively by the supervised and unsupervised approaches on the individual, *S*, *T*, *E* and *L* feature channels. The class colour codes in the supervised cases correspond to the legend of table 8.3. Colours in the unsupervised cases relate to different (unknown) and generally non-corresponding classes.

	<i>S</i>	<i>T</i>	<i>E</i>	<i>L</i>
<i>S</i>	1.00	0.14	0.05	0.05
<i>T</i>		1.00	0.08	0.07
<i>E</i>			1.00	0.33
<i>L</i>				1.00

Table 8.4 Pairwise NMI for the unsupervised feature channels with $q_i = 200$, $k = 11$.

The pairwise similarities of output maps were compared quantitatively using NMI. The NMI values for the unsupervised maps in figure 8.4 are summarised in table 8.4. According to this metric, the maps produced by the edge and lacunarity features are the most similar pair (NMI = 0.33). As was mentioned earlier, NMI could potentially be used as one means of assessing which channels to accept or reject, as part of the unsupervised fusion process. Further consideration of channel selection metrics could form the focus of future work but in this case, the two channels with the lowest mean pairwise NMI are selected, these are *S* (mean NMI = 0.08) and *T* (mean NMI = 0.097.) The remainder of section 8.5 is an exposition of a more detailed investigation and focuses on the process and results obtained using the two selected feature channels.

Figure 8.5 shows the post-processed map results, with features generated at the three different kernel sizes, $k \in \{5, 11, 15\}$. Unsupervised models were created with randomly generated sampling points, $q_i \in \{25, 50, 100, 200\}$. Only the results for $q_i = 100$ are shown.

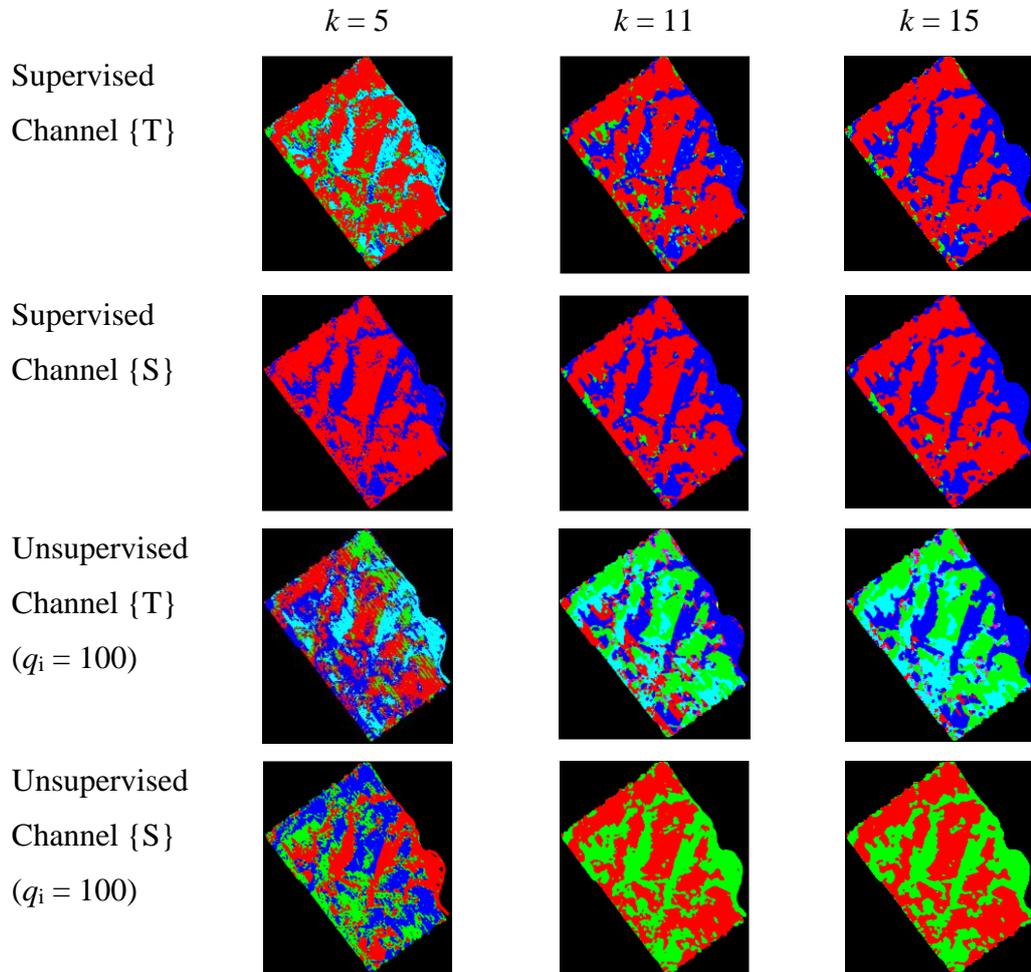


Figure 8.5. Specimen class maps produced using supervised and unsupervised approaches on two independent channels (see text for a full explanation.)

The top two rows in figure 8.5 show the results of the supervised classifications, using all the ground truth points to induce the classification models. In the supervised cases, the colour coding corresponds directly to the class legend in table 8.3 and the class map in fig. 8.2(b). With the exception of the texture channel at a scale of $k = 5$, the coverage of sediment has been predicted as predominantly class 1 and class 3. There is a gross classification error using the textural channel at $k = 5$, since the region which broadly corresponds to class 3 has been incorrectly classified as minority class 4. At this stage, the colour coding of the unsupervised maps represents provisional classes and has no

correspondence to the true class types.

The top rows of figures 8.6 and 8.7 show the relative frequency of ground truth class points within the predicted class regions for the S and T -channels, respectively, in the specimen class maps. The lower rows show the distributions of class posteriors over the predicted regions. The left three columns show the supervised approach with $k \in \{5, 11, 15\}$. Results for the unsupervised approach ($q_i = 100$) are on the right. Comparing figures 8.6 and 8.7, it is clear that in all cases, the textural features yield a greater number of predicted classes in both the supervised and unsupervised approaches.

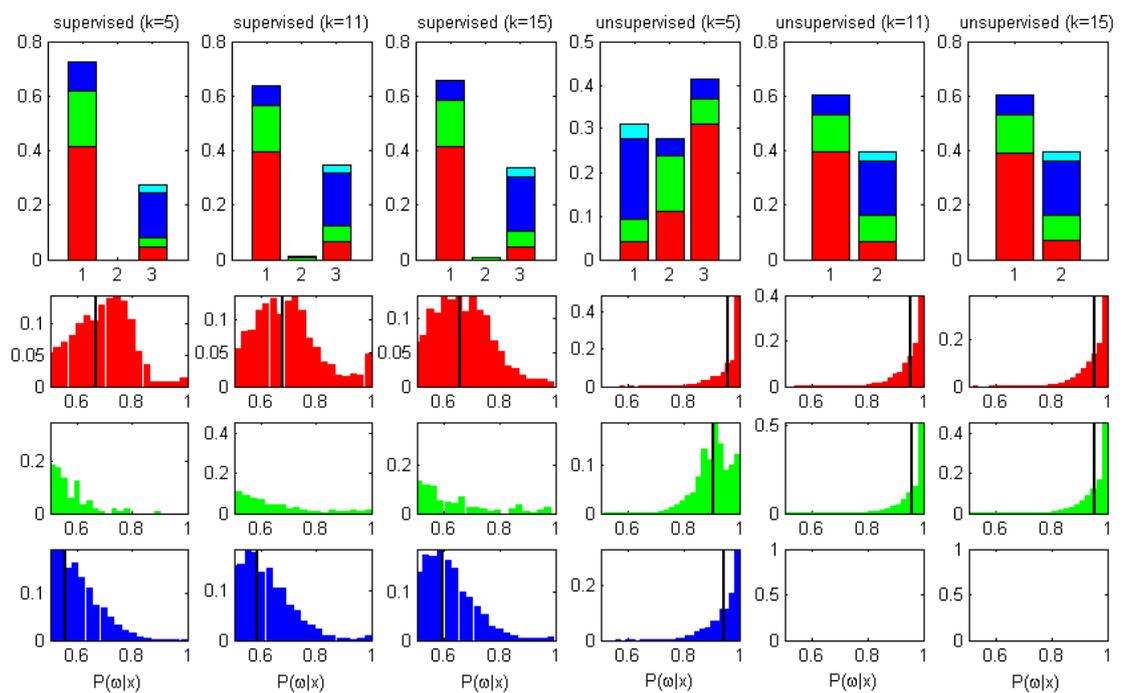


Figure 8.6. S -channel results. Top row: Relative frequency of occurrence of a ground truth sediment class within the predicted sediment class region, $C_g \cap C_R$. Lower three rows: distribution of class posteriors over the predicted areas in the machine maps. Left three columns show the supervised approach with $k \in \{5, 11, 15\}$. Results for the unsupervised approach ($q_i = 100$) are on the right.

An important difference, between the supervised and unsupervised approaches is the distribution of posteriors for the predicted classes. In all cases, the mean of the distribution (vertical black line) is greater for the unsupervised results, being more peaked, with modal posterior values close to unity and hence a much higher certainty in the identification of the provisional classes, relative to the supervised case. The supervised models have generally been successful at predicting the regions of class 1

and class 3. However, it is clear there is a tendency for the supervised models to misclassify class 2 as class 1. Interestingly, class 2 was the least pure in the manual classification and has lower mean posterior values over the classified region, compared to the other classes.

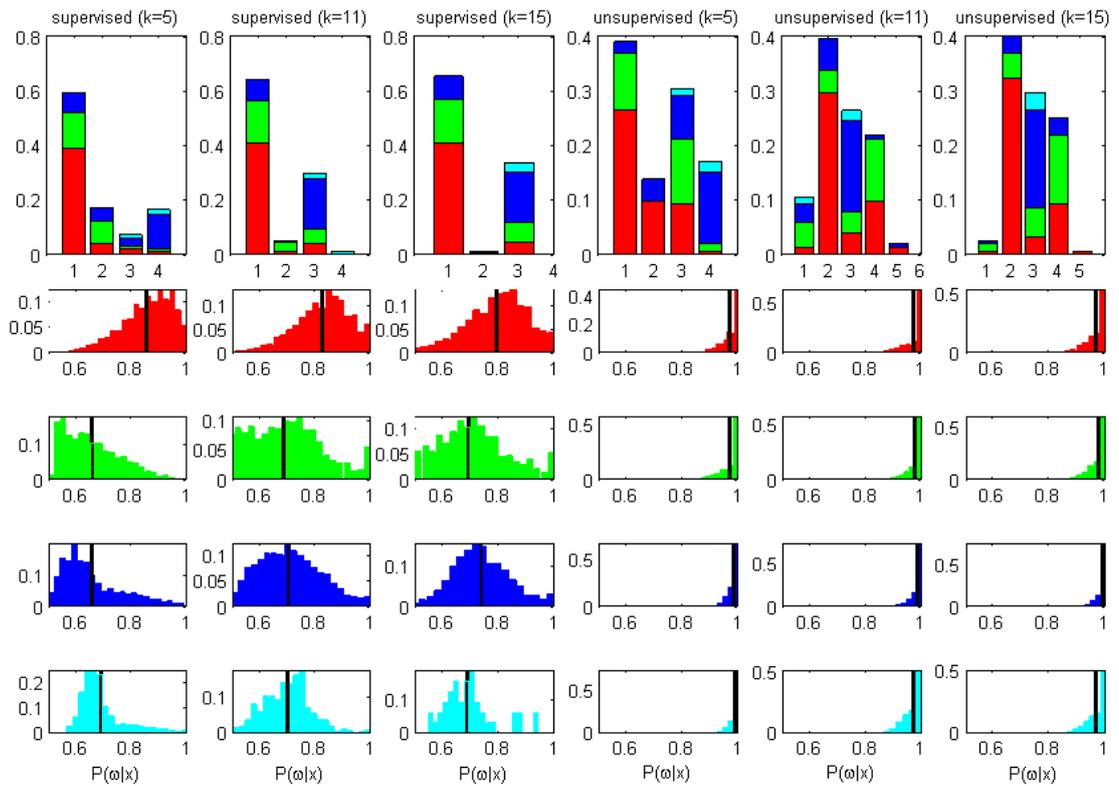


Figure 8.7. *T*-channel results. Top row: Relative frequency of occurrence of a ground truth sediment class within the predicted sediment class region, $C_g \cap C_R$. Lower four rows: distribution of class posteriors over the predicted areas in the machine maps. Left three columns show the supervised approach with $k \in \{5, 11, 15\}$. Results for the unsupervised approach ($q_i = 100$) are on the right.

Turning to the unsupervised approach, the provisional class regions are first relabeled, as described in subsection 8.3.8. As an illustrative example, consider the ground truth distributions in the provisional classes shown for the unsupervised, *T*-channel at $k = 15$ (top-right corner of figure 8.7.) Five provisional classes have been identified. Using the conditional probabilities of the ground truth classes within the provisional class regions, it is clear that provisional class 2 strongly corresponds to ground truth class 1, as the largest proportion of bar 2 in the figure is red. Provisional class 3 is labeled as class 3 (blue) and provisional class 4, as class 2 (green.) Provisional classes 1 and 5 contain mostly ground truth class 2 and 1, respectively, so all regions in these provisional classes are relabeled according to the ground truth class. Applying this process in each

of the test cases produces the (unsupervised) class maps, with the colour coded regions now corresponding to the class legend of table 8.3. The resulting specimen class maps are shown in figure 8.8.

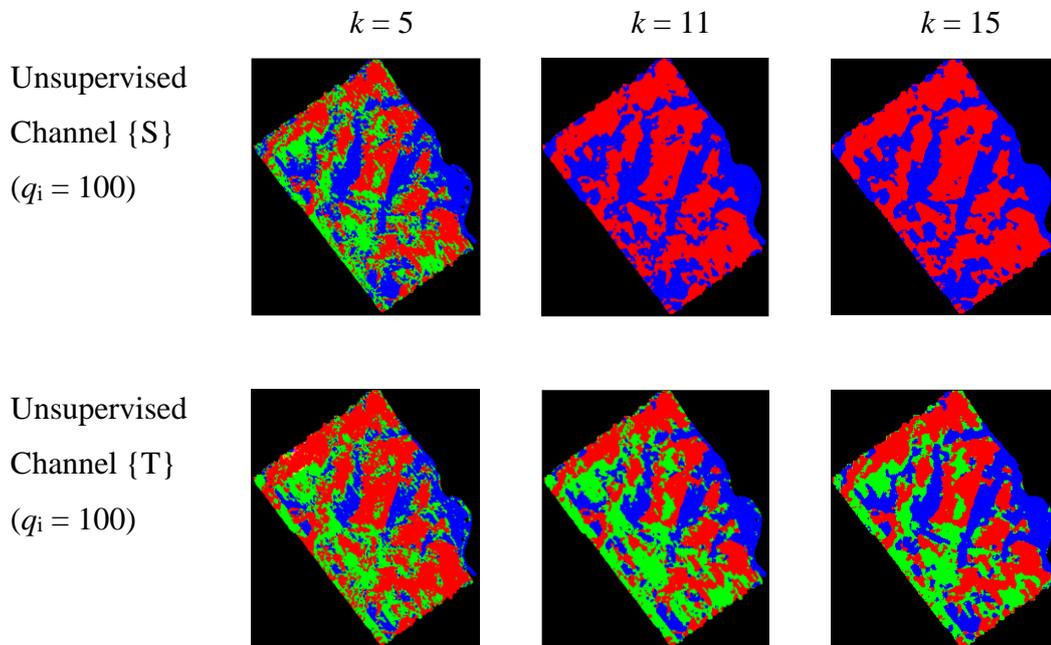


Figure 8.8 Specimen unsupervised class maps from fig. 8.5, relabeled so that the colour codes of the classified regions correspond to the ground truth class legend of table 8.3.

Visually, from figure 8.8, it is immediately apparent the unsupervised textural channel has been more successful at discriminating class 2 (green), at all resolutions. The statistical feature channel has discriminated class 2 only at the highest resolution ($k = 5$). Minority class 4 has not been identified in any of the cases. With corresponding class labels in all of the specimen maps, it is now possible to make a direct comparison between the supervised and unsupervised machine maps and the specimen map. The similarity of the maps is related to the number of collocated, identically classified points. It can be quantified as the sum of the diagonal elements in the 4×4 confusion matrix, divided by the sum of all entries in the matrix. This measure ranges from 0 (no similarity) to 1 (identical). The results for the test cases are summarised in table 8.5 and shown graphically in figure 8.9.

Channel	Supervised			Unsupervised		
	$k=5$	$k=11$	$k=15$	$k=5$	$k=11$	$k=15$
	Similarity					
<i>S</i>	0.52	0.54	0.54	0.56	0.56	0.56
<i>T</i>	0.50	0.57	0.55	0.61	0.59	0.59

Table 8.5. Similarity between the supervised and unsupervised class maps and the specimen class map.

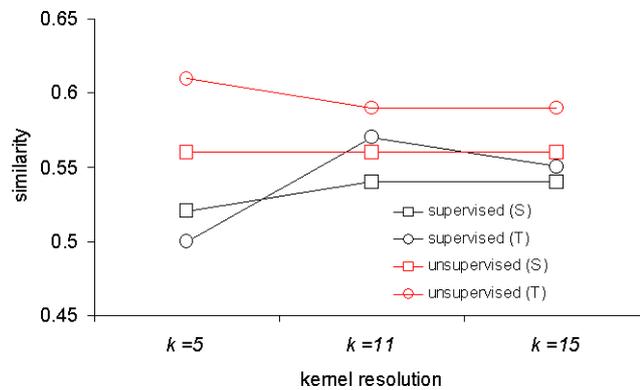
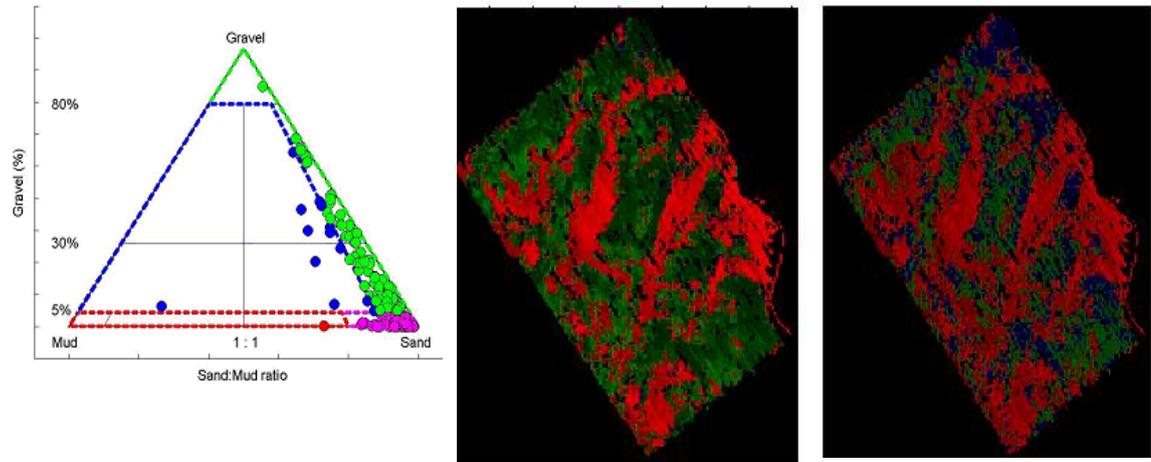


Figure 8.9. Similarity of the machine produced class maps to the human produced class maps for the experimental test cases.

Quantitatively, the unsupervised approaches produced class maps more similar to the manual approach (specimen map) than the supervised approach. The result is not really surprising, considering that in the supervised cases, class 2 was often misclassified as class 1 or class 3. A most likely explanation for this is label noise and class heterogeneity in the training samples. Results are of course dependent on the features used, for instance using local intensity histogram features will produce a different outcome to the descriptive statistical features applied in the test case. Another important factor that influences the results is the classification objective. In this case, it is mapping the sediment distribution according to estimated mean grain size. If there were a different classification objective, such as mapping the sediment distributions according to the UK SeaMap classification scheme (Long, 2006), there would be a different classification outcome. Since, in the UK SeaMap scheme, the ground truth data points are assigned categorical class labels related to the mixture ratio of the three basic

sediment types, sand, gravel and mud, from the PSA. The class labels are assigned according to the region occupied in the modified (simplified) Folk Trigon (ternary diagram), shown in figure 8.10 (a).



(a) UK SeaMap sediment trigon (b) *S*-channel (2-classes) (c) *T*-Channel (3-classes)

Figure 8.10. In the UK SeaMap classification objective, the ground truth points are assigned categorical class labels according to the predefined class region occupied by the points in the ternary diagram.

Figure 8.10 (b) shows the supervised UK SeaMap classification using the *S* channel and 8.10 (c), the *T*-channel. The results of this work are not discussed further here, as there is no means of making any comparison – there is no manually produced classification, compliant with the UK SeaMap classes for these data. It is curious though, that with this different classification objective, the filter bank features have again identified more sediment classes than the statistical features.

8.5.3 Unsupervised fusion approach

In the test case here, there is little point in combining the supervised channels, since in the case of $k = 5$, there is a gross classification error on the *T*-channel and at $k = 11$ and $k = 15$, the output maps from the *T* and *S* channels are similar.

For the unsupervised approach, there is more variability in the results from the channels, as is clearly shown in figure 8.5. The unsupervised results from the *S* and *T* channels are combined, as described in the procedural steps 1-5 in section 8.3.7, item 2. The case for

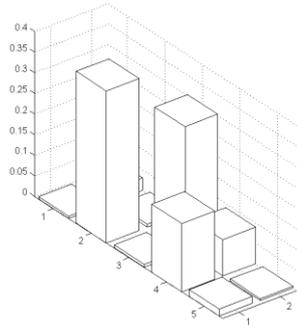
$k = 15$ is described here.

Firstly, the joint distribution of a collocated sample of points from each independent channel is constructed. For n -channels, the distribution will be n -dimensional. As two channels are used in the test case, the sample distribution is two-dimensional, as shown in figure 8.11 (a). The number of bins in the distribution depends on the number of provisional classes identified in each feature space. There are 10 bins in this example as five provisional classes were identified in the T -space and two in the S -space. Each bin corresponds to a label permutation. Therefore, the heights of the bars in the distribution correspond to the relative frequency of occurrence of a particular permutation of labels in the collocated sample space. In other words, it is an empirical estimation of the probability that a randomly sampled channel location will have a particular permutation of labels. If the relative frequency of occurrence is less than 0.05, the permutation bin is considered to be non-relevant. The top- n bins with the highest relative frequencies are selected as the provisional classes. In this case, four classes are chosen, as there are four ground truth classes. All collocated channel points are now relabeled, according to the permutation of the channel labels. Those with a non-relevant permutation are assigned a class label of '0', which corresponds to a masked location. The support assigned to the re-labeled permutations is the mean posterior of the labels at the collocated points. The resultant provisional class map is shown in figure 8.11 (d). To assign the most appropriate class labels to the provisionally labeled classes, the distribution of the ground truth classes within the regions of coverage of the provisional classes is again considered. As should be expected, the provisional classes are impure, shown by the distribution at the top of figure 8.11 (b). The four sub-figures show the distribution of support, i.e., the posteriors for each provisional class region. The 'green' and 'cyan' provisional classes have a lower mean certainty, in comparison to the 'red' and 'blue' provisional classes. Using the relative frequencies of occurrence of the ground truth classes in the provisional class regions, the conditional probabilities of occurrence of a particular ground truth class within a provisional class region are estimated. The results for all ground truth classes and provisional classes are given in table 8.6.

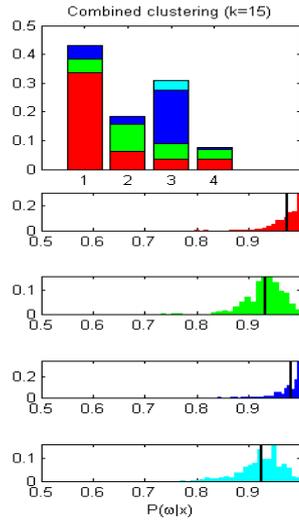
		Provisional class (C)			
		c_1	c_2	c_3	c_4
Ground truth	g_1	0.77	0.34	0.11	0.45
Class (G)	g_2	0.11	0.52	0.18	0.45
	g_3	0.11	0.15	0.60	0.09
	g_4	0.00	0.00	0.11	0.00

Table 8.6. Conditional probabilities of finding a ground truth point of a particular class, within a provisional class region.

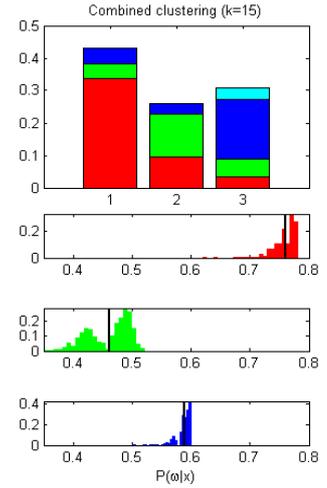
The provisional class region is assigned the label of the ground truth class with the maximum conditional probability. Considering table 8.6 and the distribution at the top of figure 8.11 (b); Provisional class 1 is labeled ‘class 1’ since, $p(g_1|c_1) = 0.77$, provisional class 2 is labeled ‘class 2’ ($p(g_2|c_2) = 0.52$) and provisional class 3 is labeled ‘class 3’ ($p(g_3|c_3) = 0.60$). The final provisional class 4, is a mixture of three ground truth classes which are not distinct from the others. So the most sensible decision is to merge this provisional class with another class. However, as the provisional class 4 is bimodal, a tie-break is needed to decide if it should be merged with class 1 or 2. A larger proportion of class 2 is distributed amongst the other classes, compared to class 1, so it is assigned to class 2. The resultant class map is shown in figure 8.11 (e). The support of the resultant class map is again modified because of the uncertainty in the assignment of the ground truth labels to the provisional class regions (as each region contained a mixture of ground truth points with an associated conditional probability). The posterior supports for each class region are multiplied by the conditional probability of finding the ground truth class within the classified region. This has the effect of shifting the distribution of conditionals left along the axis, as shown in figure 8.11 (c). The shape of the distribution of posteriors is unaltered, except in the case for class 2, which is now a bimodal, additive mixture of the distribution of posteriors for provisional class 2 and 4. Figure 8.11 (f) shows the final class map, with the colour intensity of the classified regions proportional to the posterior probabilities of the classified points in each class region.



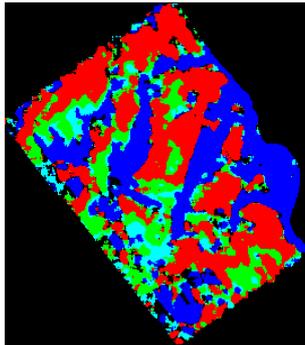
(a) Joint distribution of co-located label permutations.



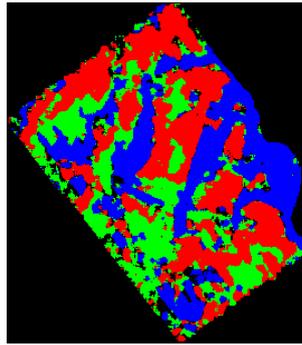
(b) GT label distributions within the spatial distribution of the provisional classes



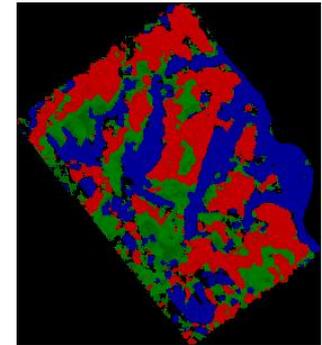
(c) GT label distributions after relabeling points in cluster 4 as class 2



(d) Class map showing distribution of the 'top-4' permutations



(e) Class map after relabeling and reassignment of ground truth labels.

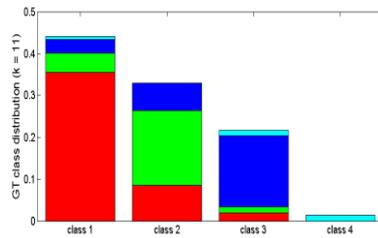


(f) Class map with colour intensities proportional to the class posterior probabilities.

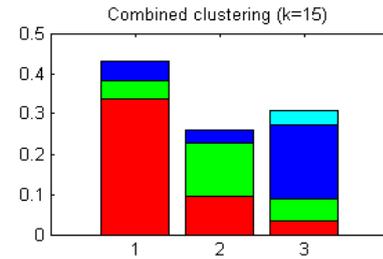
Figure 8.11. Results of applying the unsupervised fusion strategy and of relabeling the classes according to the most likely ground truth classes. See text for a full description.

For a further visual comparison, the results of the manual classification and the unsupervised ensemble are shown side-by-side in figure 8.12. Figure 8.12 (a) and (b) show the distribution of ground truth classes within the classified regions. The visual comparison between the manually produced and machine maps in figure 8.12 (c) and (d) is plausible, with bulk areas of class 2 (green) to the left and areas of class 1 (red)

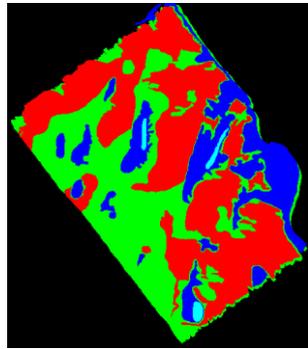
that appear to correspond well between the two maps. Class 3 (blue) on the other hand has a larger regional coverage in the machine map compared to the manually produced map. Quantitatively, the similarity between the two maps is 0.6. This might at first indicate a closer similarity to the manually produced map, compared to the independent S and T channels which had a similarities of 0.57 and 0.59 respectively, when considered individually. However, this ‘improvement’ may not be real or relevant. The aim of the work here is not to fit the machine map to the manually produced map but to show that the proposed process and methods produces a plausible classification, substantiated with numerical (probabilistic) support. The specimen, manually produced map is used purely for comparison, it is not a benchmark and the class types and their boundaries are uncertain. It may be the case that the machine map is more representative of the classes and their distributions than the manually produced map. So, if a closer similarity to the manually produced map is achieved, it may in fact indicate that a less representative machine map has been produced, if the manual map is not representative of the real class distributions and their boundaries. Thus combining the S and T channels may have made the class map less representative but equally, it is not an implausible representation. Resolution of this ambiguity is not a tractable problem.



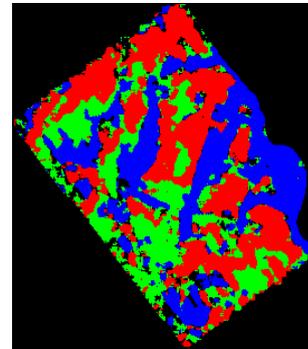
(a) distribution of ground truth points in manually produced class map.



(a) distribution of ground truth points in class map produced from the fusion of unsupervised classifications.



(c) Manually produced specimen class map



(d) Machine class map (fusion of unsupervised classifications)

Figure 8.12. Comparison of manually produced (left) and fusion of unsupervised machine classifications (right.)

8.5.4 Overall discussion

Taking into consideration the qualitative and quantitative outcomes of the different processes considered in the experiments and the external validation in the form of ground truth and a specimen class map, the single channel (filter bank features) in the unsupervised process is the best performer. Although the unsupervised ensemble approach was also efficacious from a qualitative perspective, in the discrimination of the main classes and the quantitative similarity to the specimen class map, it is computationally more expensive than a single channel, single classifier approach and far more challenging to implement.

A manual approach is subjective, time consuming, inconsistent and expensive. The machine approach (irrespective of whether supervised, unsupervised, single or multiple model) in the experiments here is conservatively, at least three orders of magnitude faster than the manual approach, as the process does not take longer than about 5 minutes. However, the quality of the output is also important. An important advantage that the machine approach has over a manual approach is (at least in the methods here), it provides numerical support for the certainty of the classification at each location in the class map. As was demonstrated, this support can be modified to take into consideration the impurity of the classes, as defined by the mixture of different ground truth classes within the classified regions.

Although it is possible to apply a global class support using the conditional probabilities of the ground truth classes within the regions of the classes in a manually produced map, it is not possible to assign that support to specific points. Further, the unsupervised model is also rigorously internally validated, in that the optimal unsupervised model is chosen, based on multiple random samplings and pre-clusterings so that the most appropriate number of provisional classes is identified and the best solution for that number of classes is selected. Thus, machine production of the map has speed, internal validation and outputs quantitative support for the certainty of the classification results. A further advantage is, all other things being equal, consistency. Given the same data, processing flow, feature sets and operational parameters, the machine process will deterministically produce identical classification results – there is no intra-rater variability. This is not the case with human/manual processing. Both manual and machine approaches are capable of producing plausible classifications but the machine has some clear advantages as described above.

The classification objective is of paramount importance. In the test case here, the objective was to use the machine approaches to classify the sonar image according to mean sediment types (defined by grain size distributions.) Using the same data, features and methods but a different classification objective will produce a different outcome, for instance, a classification of sediment types according to the UK SeaMap classes. The choice of features is naturally an important decision. A surprising outcome was that it was not known beforehand if the Gabor filter bank features would be useful for the

classification objective described here. They produced a classification similar to the specimen class map, identifying the three main classes at a various resolutions. The statistical features produced classifications that were predominantly binary in the supervised and unsupervised cases.

In general, the supervised approach did not work as well as the unsupervised approach and this is probably due to fully automating the process using image samples selected around the ground truth points. Label noise and class heterogeneity are possibly responsible for the induction of unrepresentative classification models. In future work, manual, visual selection of training regions will be considered for the supervised process, as in Blondel and Gómez-Sichi (2009.) Manual selection of training regions may be more appropriate for the Sabellaria discrimination problem where the image textures are quite distinct.

Overall, this investigation has been very successful, since the flexibility of the process has been implicitly demonstrated. The good results obtained from the contingent application scenario indicate that the process and methods are versatile and therefore have the potential for application to a variety of seabed target discrimination problems.

8.6 Evaluation, conclusions, recommendations and scope for further work

8.6.1 Evaluation

The aim of this work was to devise a machine process and methods for fully automated multi-class classification of subjective targets in real-world mosaic imagery. The hybrid, unsupervised process has internal validation components for identifying the number of provisional classes and finds the best clustering solution for a given number of provisional classes. This facilitates induction of a probabilistic unsupervised classification model which can be applied consistently to the individual blocks in the block processing scheme. The research in this case study was primarily exploratory and descriptive and used mixed methods to investigate the plausibility of the process, methods and outcomes in the (contingent) application context, a sediment classification task. An objective was to evaluate classification outcomes in a qualitative comparison

between supervised, and unsupervised approaches and the types of features used, compared to a specimen manual classification. The classification objective was to divide the image into regions corresponding to different sediment target types (as defined by their grain size ranges). A curious outcome was that the features derived from the Gabor filter bank channel operating in isolation identified more classes, relatively consistently, over a range of resolutions, in the single-channel unsupervised process. There was no evidence available beforehand to indicate that the Gabor filter bank would be a useful feature generator for this task. In fact, it was expected that the statistical features would be the most salient. The unsupervised ensemble approach using filter bank and statistical features also produced plausible results but there was no strong evidence to indicate that this more complex, expensive process is better than the unsupervised approach using the single, textural channel. Predominantly, the supervised approach led to a binary classification outcome. This is most likely due to label noise and hence the induction of unrepresentative models. The hybrid, unsupervised process, using the Gabor filter bank features showed reasonably good qualitative and quantitative similarities to the specimen manual classification. Gabor filter banks were the most promising approach for the sediment discrimination context considered here. The machine class map was produced in a few minutes, compared to days of work for a human analyst. The main aim and objectives of this work have been successfully achieved. Importantly, this study has demonstrated the versatility of the process and methods, intimating the potential transferability to other target discrimination problems. A paper is being prepared from the work in this case study, to submit to a journal for consideration. The main contribution of this work is to the fully automated processing of subjective target discrimination tasks in real-world imagery.

An experimental harness was designed and implemented in Matlab, comprising about 40 separate m-files. Further scripts were written to carry out additional post-processing of the intermediate results generated in multiple runs of the harness. Several of the components of the experimental harness were later integrated, with modifications, in a prototype sonar mosaic image segmentation/classification process, for the industrial stakeholders.

8.6.2 Conclusions

Novel processes and methods have been proposed and evaluated on the application context of the discrimination of sediment target regions in a real-world sonar mosaic image. The hybrid, unsupervised approach using the Gabor filter bank features is the most promising approach for the discrimination of sediments considered here, according to the prescribed classification objective. This combination of features and processing methods produced the best overall results, over a range of resolutions, in comparison to the other approaches considered. The results are plausible and importantly, the classifications have numerical support. The aim and objectives of the case study have therefore been successfully achieved. The work also intimates the potential of the process and methods for application to a variety seabed target discrimination problems.

8.6.3 Recommendations

The following recommendations are provided, regarding the machine learning process configuration. Gabor filter bank features are recommended as the first choice for a single feature channel set-up but before use, they will need parameter tuning and validation on samples of the image texture classes.

If a supervised approach is to be used, it is recommended that patches of visually identifiable texture (preferably with ground truth validation) are selected for classifier training, rather than using the immediate vicinity of the ground truth locations as seed locations. The texture patches used for feature evaluation and selection may also be used for classifier training.

The number of random locations to be sampled is the only parameter for the unsupervised process. Some experimentation, involving visual inspection of the output class map may be required to find a good number of training points in the unsupervised case, as this will vary with image size, kernel size, features used and number of tentative classes. For the approximately 16 MP image tested, with 3-4 classes, 100 sample locations (up to 900 training patches) worked well.

A single parameter, the block width, in tiles, determines the (square) block size used in the block processing scheme. The tiles are the image neighbourhoods, sized according to multiples of the kernel size used for the feature creation. There are no specific recommendations for the block size as this may depend greatly on the PC processing power available. Using a block width of $100 \times$ the feature kernel size should be a useful starting point.

8.6.4 Scope for further work

A mini-proposal for further work on Sabellaria classification is presented in chapter 10 (10.2.2). Some other avenues for investigation include:

- Evaluation with a variety of different classification objectives, for instance, rock and sediment classification, habitat classification, other sediment classification schemes and complex morphological regimes.
- Integration of other data types (and features) in the classification solution, such as multibeam bathymetry or single beam echosounder.
- Investigation of a semantic labeling process taking into consideration the local variations in spatial uncertainty at the ground truth locations.
- Further investigation of feature channel selection and combination methods for the ensemble process.
- Investigation of the manual selection of training regions (patches of imagery) for the supervised approach, compared against using ground truth seed locations.

It is important to point out that the further work is dependent on the availability of good data sets, with comprehensive ground truth and manually produced specimen class maps to improve robustness of the external validation.

Evaluation

Chapter 9

Contents

9.1 Overall evaluation

9.2 Research project management

9.2.1 Stakeholder influence on research design

9.2.2 Division of project time

9.2.3 Publication output and schedule

9.2.4 Execution of project case studies

9.2.5 Risk revisited

The purpose of this chapter is to provide an overall evaluation, bringing together various aspects of the research in section 9.1. Some of the research project management issues and outcomes that were pertinent to the project are also considered, in section 9.2.

9.1 Overall evaluation

Overall, the research project has been successful and the aims and objectives have been met, as described in each of the individual case studies. The individual studies also demonstrated that context limited contributions to advances in knowledge were achieved.

The research carried out concerned machine learning methods for discriminating natural targets in seabed imagery. The methods were applied at stages 3 and 4 in figure 2.2, to interpretative discrimination/classification problems, usually carried out in a manual process, involving a human analyst. In all of the case studies, it was found that either full- or semi-automation of the process was a tractable problem. The important industrial application tasks, included pockmark discrimination and Sabellaria discrimination.

The machine learning methods aspect of the research focused on how computational and machine learning algorithms could be applied in the processes and how they could be combined, as part of the approach to solving the specific industrial problems. The classification and clustering algorithms *per se* were of secondary importance, being treated as black boxes for performing a classification or partitioning within the framework of the machine methods and processes. In the case of binary discrimination problems such as the discrimination of pockmark or Sabellaria targets from the background, the BVM classifier was found to work well. For multi-class problems such as sediment classification, the Naïve Bayes classifier was suitable and provided probabilistic support for the class predictions of individual pattern instances.

For the novel processes devised for the specific application problems considered in this research, the feature creation methods and the evaluation of the feature configurations and parameterisations is fundamental to the success of the approach. Salient features need to be identified and applied with appropriate configurations and parameterisations, to extract useful information about the targets. Feature creation and evaluation methods are therefore a central and integral aspect of the machine learning methods, hence the emphasis on these methods in three of the four case studies. The bespoke feature based approaches were effective and efficient solutions to the problems, successfully completing the tasks in a fraction of the time compared to a manual approach. However, careful set-up and choice of features, feature configurations and parameterisation is required for each situation using different image data, targets and classification objectives. Evaluation of the features is a significant challenge, and more fundamentally, measuring the robustness of the method used to evaluate the features for the prescribed tasks. In other words, when measuring the saliency of a feature, it is not necessarily known how robust the measurement method is across a variety of tasks and feature configurations. An approach to the issue of determining a robust ranking of parametric features using distance measures and novel committees was considered in the third case study (chapter 7.) Even if a few individual features are chosen to create a classification model though, it can still be difficult to associate the properties of the targets observed in the imagery with the properties of the feature creation methods and the information these methods capture about the targets. There was clear evidence showing the Gabor filters were good discriminators of Sabellaria textures (and also, unexpectedly, sediment types.) Yet, the underlying connection between the properties of

the Sabellaria textures and the reason why the Gabor filters work well on this task is not known (another avenue for further work.)

There are numerous permutations of machine learning methods that can be applied, from the palette of feature creation methods, feature evaluation and selection strategies, classification and clustering algorithms and post-processes. The approach and the methods depend not only on the classification objective and data but also on the size of the input space. Further technical issues arise as the size of the input space grows, as was evident in case study 4 (chapter 8) where a number of earlier ideas and methods were integrated in the context of real-world mosaic imagery. The novel, hybrid unsupervised approach produced plausible classification results on the sediment classification task, in a few minutes of processing time. The process is versatile and the researcher considers that it should be generalisable to the discrimination of other subjective natural targets in seabed imagery, such as Sabellaria and pockmarks (with appropriate features and feature tuning.) Further, the full automation of pockmark and Sabellaria discrimination/classification is feasible within this processing framework.

The mixed methods analyses of the results proved useful in creating a holistic overview of the properties of the machine methods and processes, as applied to the subjective, interpretative problems. It is helpful to consider the outcomes of applying a machine method to an interpretative problem using a variety of metrics, not just classification accuracy. The ‘true’ interpretation and classification is unknown (and generally unknowable) and therefore any solution (human or machine) is uncertain.

It should be apparent, if there were proven, reliable, methods and hardware for wide coverage remote acoustic discrimination of targets such as Sabellaria and pockmarks, that an interpretative approach may not be needed. So, expediting the required laborious, manual discrimination and classification tasks with machine methods is a reasonable proposition. Despite the preference amongst many researchers for the emergent automated, objective approaches to seabed classification problems using calibrated backscatter, there is still a niche for bespoke, feature based approaches to such subjective interpretative problems.

Standardisation (or lack of it) is still a pressing issue in many seabed classification problems though. One means of advancing the study of seabed texture discrimination and classification would be to build a database of ground-truthed sonar images. The database could contain real-world imagery of different sediments, bedforms, habitats, biogenic and geological structures. A selection of classification and discrimination objectives and metrics could be prescribed to facilitate a standardised evaluation of different proposed machine learning methods for automation of the tasks. The database could be made freely available to researchers around the world and would enable the comparison of results from different machine methods on benchmark objectives and data.

9.2 Research project management

The purpose of this section is to outline some of the key successes and difficulties of the project in relationship to the project structure and management. There is some overlap of the project management with research approach and design described in chapter 4.

9.2.1 Stakeholder influence on research design

Much real-world research is agenda driven and multiple parties or stakeholders are often involved. The stakeholders influence the themes and direction of the research. In some cases, research is tendered through a bid process and contract, with aims already prescribed by the stakeholder (tendering organisation). This is in contrast to “blue sky” projects where, say, research aims are proposed relatively independently by the researcher, without necessarily having a clear practical or application driven goal. Funding for these projects can be sought through a competitive grant application. Even so, the researcher is still subject to external influences, since the proposal for a “blue sky” project is likely to be constrained by the availability of funding for specific types of research in relationship to the researchers’ domain and the proposal details. In other words, the work undertaken by a researcher can never be fully independent of external influences and some degree of flexibility is therefore required to adapt the themes and design of the research according to these factors. Unfortunately, external influences may be adverse and are not always positive or constructive. This may lead to a variety of problems with inevitable delays or even failure of a project. The project risk is particularly high in research work where a large proportion of the time and effort is consumed with software development tasks, as these carry their own spectrum of risk factors. See, for instance, Schmidt et al, (2001.)

The research project involved industrial and academic stakeholders who exerted influence on the themes and direction of the research within a general interpretation of the aims. The design originally proposed by the researcher had a central theme concerning the discrimination and classification of pockmarks. This piece of research involved the researchers own ideas and decisions, relatively independently of the stakeholders. As well as containing a “blue sky” element it also had clear practical and commercial goals. A point of departure from the original design and plan was marked

by an agreed change of direction to the problem of Sabellaria discrimination as this was a priority on the industrial stakeholder agenda. Unfortunately, data provision issues encountered by the industrial stakeholder, connected with this line of enquiry later induced a further change of the research themes and design to accommodate even greater flexibility and alternative pathways. Consequently the research project was fragmented into a collection of case-studies. The global case-study structure is therefore evolutionary, being brought about due to external factors, rather than a structure initially planned by the researcher.

9.2.2 Division of project time

The timeline of research activities (in their most general terms) was summarised in chapter 4, figure 4.2. Figure 9.1 shows the estimated division of time spent on the main project tasks. The most time consuming component of the work was experimental harness development. Each change in direction of the project required further substantial development work, data sourcing and preparation. Data preparation and experimental harness development (software design, engineering and coding) are not direct research activities, although clearly, they are essential to support the research and are thus labeled as indirect research in figure 9.1. Although unavoidable due to the external factors driving the changes in direction and because of the lack of any suitable pre-existing software implementations, the repeated experimental harness development for different tasks was a less effective use of the time budget.

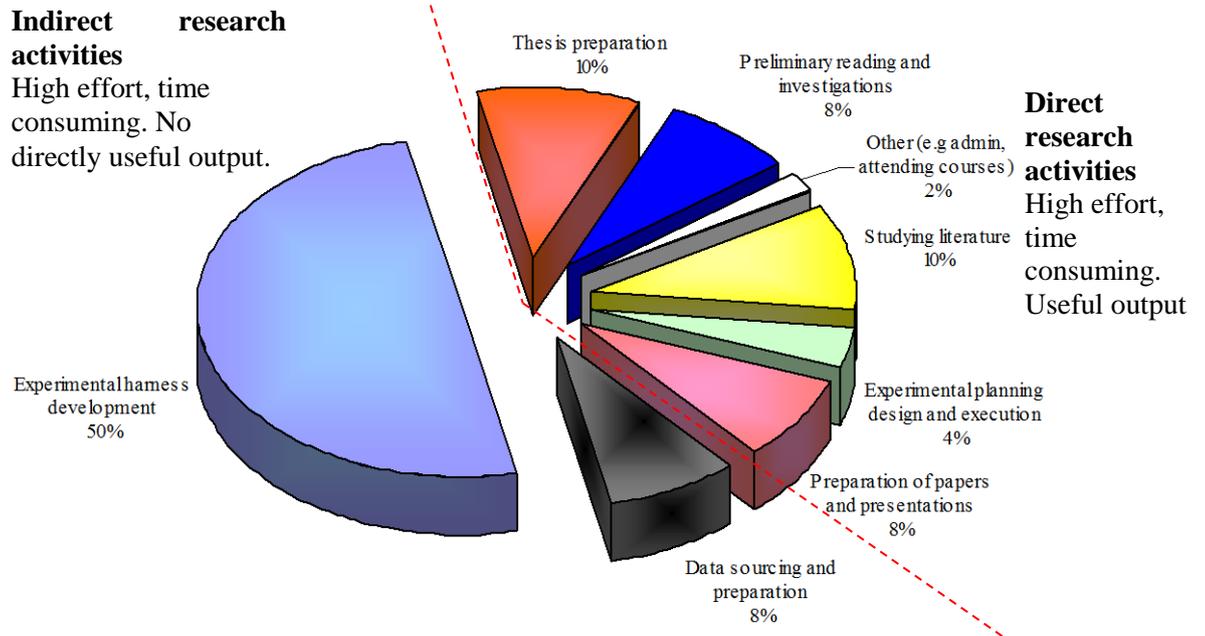


Figure 9.1 Division of time on the main project tasks. Indirect research activities are to the left of the dashed red line.

Extending 1.25 years beyond the 3-year completion benchmark and given the 8 months of preliminary work, the project has obviously encountered severe delays. The researcher considers that project time management was based on sound principles and practice and that overrunning was due to adverse stakeholder influence on the project and issues with the provision/acquisition of suitable data.

9.2.3 Publication output and schedule

The fact that five peer reviewed conference and journal publications have arisen during the course of the project is evidence that the research output can withstand the test of external scrutiny. The researcher considers that three conference publications and one journal publication would be a satisfactory goal to achieve regarding publication output. This element of the project has been successful. The timeline for the papers and presentations is shown in figure 9.2.

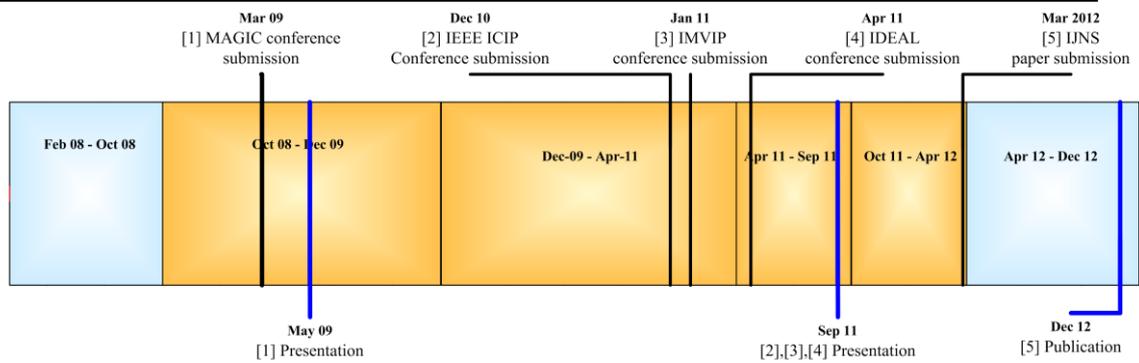


Figure 9.2 Conference and journal paper submission and presentation/publication timeline.

9.2.4 Execution of project case studies

Some of the main project investigation themes are listed in table 9.1. Not all of these themes evolved into case studies in the thesis. The SMART assessment of objectives is applied to the themes as a means of comparing their relative merits. The ‘Data’ column indicates if sufficient data were in the possession of the researcher at the time the decision was made to follow a particular theme.

Description	Data	S	M	A	R	T
1. Pockmark discrimination in MBES data	✓	✓	✓	✗	✓	✓
2. Application of machine learning ensemble	-	✓	✓	✗	✗	✓
3. Sabellaria image texture discrimination	-	✓	✓	✓	✓	✓
4. Distance measure committee for feature evaluation	-	✓	✓	✗	✓	✗
5. Active learning for sabellaria discrimination	-	✓	✓	✓	✓	✗
6. Sediment classification in sidescan sonar imagery	✓	✓	✓	✗	✓	✓

Table 9.1 Comparison of research themes according to the SMART criteria.

Description	P	Alt	C	EF	Pu	CS
					b	
1. Pockmark discrimination in MBES data	✓	-	-	✓	1	5
2. Application of machine learning ensemble	✓	-	-	✓	0	-
3. Sabellaria image texture discrimination	✓	-	✓	✓	1	6
4. Distance measure committee for feature evaluation	-	✓	✓	✓	2	7
5. Active learning for sabellaria discrimination	-	✓	-	✓	0	-
6. Sediment classification in sidescan sonar imagery	-	✓	✓	✓	0	8

Table 9.2 Further comparison of research themes. See text for description.

Table 9.2 shows the research themes, with an indication of whether it was a planned (P) or alternative (Alt) (contingent) pathway, whether the intended investigation was completed (C) and if it could be extended further (EF). It also shows if the work led to any publications (Pub) and whether or not the work is included in the thesis as a separate case study (CS), in which case, the chapter number is given. Some points concerning the research management issues and outcomes in each theme are outlined below:

1. Pockmark discrimination in MBES data

The researcher had originally planned this theme as the main problem solving pathway for the thesis. Whilst appropriate data were in the researchers' possession and a realistic, focused plan for progression had already been devised, continuation of this work to include pockmark classification and the use of a fusion of backscatter and bathymetry data did not meet stakeholder approvals. The work was suspended when the direction of the research was changed to Sabellaria discrimination. In chapter 10 (section 10.2.1), a mini-proposal is given for advancing the pockmark research theme.

2. Application of machine learning ensemble

Whilst there was considerable stakeholder pressure to utilise ensemble methods, it was the opinion of the researcher that problem solving should be carried out as a primary activity and that appropriate technology should be used as necessary to solve the industrial problem. For example, the problem orientated research question is “can a machine learning process be used to discriminate pockmarks?” rather than “can a machine learning ensemble be used to discriminate pockmarks?” It is reasonable to choose more complex methods only when it is clear that simpler methods are unsuitable.

3. Sabellaria image texture discrimination

The researcher agreed to this theme on the premise that suitable data would be provided by the industrial stakeholders. One partially complete data set was made available and this facilitated a comprehensive investigation into the discrimination of Sabellaria textures on a relatively small scale. However, it was not possible to extend this line of work to larger data sets as the data were not available. The methods and process designed for discriminating the Sabellaria textures in large real world mosaics were applied to an alternative problem in case study 4 (point 6 below.)

4. Distance measure committee for feature evaluation

This work was more abstract and less applied than the work in the other case studies. The researcher had initial reservations concerning this line of investigation due to the additional time required for designing the experiments, preparing data, developing the experimental harnesses and writing the papers – which did eventually lead to a journal publication. Clearly though, feature evaluation has been an important central theme in the thesis and the work in this case study is novel and contributed strongly to the ‘machine learning methods’ aspect of the research work.

5. Active learning for Sabellaria discrimination

This work consumed a considerable amount of time, again predominantly in experimental harness development. Good results were obtained from a new query strategy for SVM based active learning, devised by the researcher. Also, it was shown that initial selections of training sets by different human interpreters converged over 10-15 epochs, to very similar classification outcomes. In other words, it could be feasible to use active learning to reduce inter-rater variability, so that different experts could arrive at similar sabellaria texture segmentations. However, taking into consideration the severe delays already incurred, despite the promising results obtained thus far, a decision was made to halt the line of enquiry.

6. Sediment classification in sidescan imagery

The work in this specific case study was carried out independently by the researcher, as an exploratory, proof of concept study using a novel hybrid process and methods, originally intended for supervised/unsupervised Sabellaria discrimination in real-world mosaic imagery. Although the sediment discrimination application instance was not considered to be ideal, given the time and data available, it was a reasonable course of action to take. The outcome of this case study demonstrated the versatility of the proposed process and methods and the potential for transfer to various target discrimination applications. A mini proposal is presented in chapter 10, (section 10.2.1), for extending this work to the task of Sabellaria texture discrimination.

In the researchers' opinion, all of the themes followed were quite successful, since the revised aims and objectives were met. However, it is also considered that the externally imposed, fragmentation of the project has hampered productivity.

9.2.5 Risk revisited

As stated earlier, and in chapter 4, a research project whose outcomes depend on several intensive software design and development stages for the experimental harnesses and on the provision of data has a very high level of intrinsic risk. However, it appeared that supply and provision of sonar image data containing Sabellaria was the main source of

issues with this project. The originally supplied data set turned out to be incomplete and the missing data could not be traced. Other potential data sets were identified but were ultimately not provided or not available for use. A data set from an external organisation was acquired but there were issues with the re-processing of the mosaic imagery and further, the interpretation of regions of Sabellaria in the imagery was not available.

The issues with the supply of data were unusual in this case, since the industry stakeholders provided partial funding for the project and also had a material interest in the outcomes of research utilising these data. However, whilst unusual, it is not entirely unexpected since, as Petre and Rugg (2010, p 111) state, two of the ignoble truths of doing research are “resources won’t be there when you need them” and people will not deliver. Nonetheless, a researcher should have certain expectations regarding data provision otherwise no research requiring data provision would be commenced or concluded.

Conclusions and scope for further work

Chapter 10

Contents

10.1 Conclusions

10.1.1 Case study specific conclusions

10.1.2 General conclusions

10.1.3 General recommendation

10.2 Scope for further work

10.2.1 Mini proposal 1: Pockmark classification and trapped gas hazards

10.2.2 Mini proposal 2: Unsupervised/supervised Sabellaria discrimination

Results from the four case studies were critically analysed and discussed in the respective chapters, 5-8. In this final chapter of the thesis, section 10.1 summarises the conclusions derived from the analyses of each case study (10.1.1). The main contributions are also stated. Some general conclusions from an analysis of the research as a whole are also drawn (10.1.2) and a general recommendation is stated (10.1.3). Two-mini proposals are outlined in section 10.2, designed to advance the achievements of the work, in the context of other novel applications.

10.1 Conclusions

10.1.1. Case study specific conclusions

Case study 1: A new approach to the automated mapping of pockmarks in multi-beam bathymetry

A novel feature based machine learning process was proposed and demonstrated to be effective for rapidly identifying pockmarks and mapping their boundaries. Salient features and appropriate scales of analysis at which to apply the feature kernels were identified in a comprehensive feature evaluation/validation process. These achievements showed the objectives of the study were successfully met. The work contributed to the wider domain of DBM analysis and particularly to the feature based discrimination of specific landform objects in a DBM.

Case study 2: Feature based discrimination of Sabellaria spinulosa textures in sidescan sonar imagery

The study was the first of its kind concerning the novel task of machine discrimination of Sabellaria texture surrogates in sidescan sonar imagery. Features created from the signal processing methods were generally found to be the best discriminators on the waterfall imagery. A Gabor filter bank configuration and range of filter parameters for effectively discriminating Sabellaria textures in mosaic imagery was established.

It was speculated that for visually discernible Sabellaria textures in the mosaic at a ground resolution, R_G pixels per metre, the filter kernel size, $k \approx 3R_G$ and filter envelope $\sigma \approx 1.5R_G$ may provide good discrimination results. These outcomes showed the objectives of the study were successfully met. The main contribution of this work was to the novel task of machine discrimination of Sabellaria textures.

Case study 3: Novel consensus approaches to the reliable ranking of features for seabed imagery classification

A novel framework for empirically evaluating distance measures for feature evaluation and ranking was proposed. The subsequent experimental campaign and results analysis showed that a consensus feature evaluation and ranking approach improved the correlation of measured feature saliency to classification accuracy and the correlation of feature rankings to a baseline rank on sonar imagery, under prescribed conditions. Under the same conditions, the variability of the correlation metrics was reduced, i.e. the feature evaluation and ranking robustness were improved. These achievements showed that the objectives of the study were successfully met. The work contributed to the application domain of sonar imagery feature evaluation and to the machine learning methods domain of feature evaluation and ranking. Specifically, the measurement of robustness of feature evaluation methods.

Case study 4: Unsupervised classification of sonar imagery

A novel, hybrid process and methods were proposed for the rapid unsupervised discrimination of natural targets in qualitative, real-world sidescan sonar mosaic imagery. The novel methods devised included a pre-clustering stage for inducing a sub-optimal probabilistic unsupervised model and a method for combining unsupervised classifications from independent feature channels. The unsupervised classifications using the Gabor filter bank features showed reasonable qualitative and quantitative similarities to the manual classification. The filter bank features led to a more representative classification of the sediments over a range of resolutions – an unexpected result. The supervised process was not as efficacious as the unsupervised, since in most of the experiments two classes were identified on the statistical and filter bank feature channels. The machine methods are versatile and the process can potentially be applied to a variety of seabed target discrimination problems in real-world imagery. The work contributed to the application domain of machine discrimination of qualitative targets in sonar imagery.

10.1.2 General conclusions

Evidence for the tractability of the pockmark and Sabellaria discrimination problems (and also sediment classification) with feature based machine learning methods and processes has been presented in these case studies.

Of the machine methods considered, the feature creation and evaluation methods were of central importance to the success of the approaches in the specific application contexts. The features need to be tuned appropriately to the data, targets and classification objectives.

The novel, hybrid unsupervised approach that was devised, produced plausible classification results on the sediment classification task, in a few minutes of processing time. The process is versatile and the researcher considers that it should be generalisable to the discrimination of other subjective natural targets in real-world seabed imagery, such as Sabellaria and pockmarks (with appropriate features and feature tuning.) Further, the full automation of pockmark and Sabellaria discrimination/classification is feasible within this processing framework.

Using calibrated acoustic backscatter in an objective approach to seabed classification is a method preferred by some researchers. However, there are certain discrimination and classification problems that are (currently) more suited to or depend on a subjective interpretative approach. In cases such as these, bespoke, feature based machine learning approaches can be devised to assist the human interpreter, greatly expediting the discrimination task.

A useful artefact that would help to further advance the study of seabed texture/target discrimination in general would be a database of texture samples and real-world mosaic image exemplars that could be used for benchmarking different methods and processes and a prescribed set of tasks and metrics to ensure a common evaluation framework for comparing different approaches.

10.1.3 General recommendation

Engineering a machine learning system to undertake tasks traditionally carried out by a human is an immensely challenging undertaking. Within this system, the critical aspects of feature evaluation and selection are in many cases based on measurement methods with mathematical foundations, such as distance measures. However, in order to gain a more reliable view of the usefulness of the features (and their parameterisations) for a particular task, it is beneficial to consider multiple measurements using a range of measuring methods. Individual methods have different reliabilities (and computational costs), dependent on the classification objectives, the features and the data. Therefore, the researcher makes a general recommendation that, considering feature evaluation as an art, a holistic approach should be taken to the evaluation process, using multiple measurement methods, experimental approaches and processes, as was practised in case studies 1-3.

10.2 Scope for further work

Two mini-proposals are presented that build on the achievements of the case studies in the thesis.

10.2.1 Mini-proposal 1: Pockmark classification and trapped gas hazards

The general aim of this research is to fuse morphological (DBM) attributes and backscatter attributes to classify pockmarks (see Sections 8 and 9 of the ICIP poster in Appendix 1) and establish if any particular type of pockmark is associated with potential trapped gas hazards.

The reason for doing this is to generate and analyse information that could potentially be used to better understand the connection between pockmark type, activity and other geohazards. Pockmark fields are often associated with trapped gas hazards and these can create serious problems for (very expensive) drilling operations. Is there a distinct type of pockmark or pockmark activity/inactivity that is associated more strongly with potential trapped gas hazards? If so, could it be used as an indicator for the potential risk when identified in the seabed imagery (DBM, backscatter image)?

Different pockmark types can be discriminated by considering clusters of their morphological and backscatter attributes. These clusters could be analysed, with respect to spatial locations of potential trapped gas hazards that have been identified in seismic data from the pockmark field, together with any in-situ measurement data. It may also prove fruitful to consider historical data (if available) where known blowouts have occurred and identify what types of pockmark (if any) were associated with the event.

Resources required are primarily data, including; DBM, multibeam backscatter and seismic sections covering the pockmark field and information on in-situ measurements and samples. An experimental harness would need to be designed and developed to accommodate the additional data layers.

A specific objective involving an inductive element of research is embodied in the null hypothesis: “There is no evidence to indicate that there is a distinct type of pockmark

distributed spatially in such a way that its occurrence is correlated spatially with the occurrence of trapped gas hazards identified in seismic profiles.”

An estimated time scale for carrying out this work is one year. Output from the research is one conference publication/presentation and one journal publication. Results would have a wide domain interest, covering the machine learning and geosciences communities and the oil and gas industry, exploration and production sector.

10.2.2 Mini-proposal 2: Unsupervised/supervised Sabellaria discrimination

The general aim of this research is apply the unsupervised (or supervised) process proposed in case study four, to the discrimination of potential Sabellaria colonies and possibly sub-classifications of the colonies, in large, high-resolution sidescan sonar mosaics.

Experiments would be conducted with different features and parameters to establish if it is possible to distinguish regions colonised by Sabellaria from other areas of the seabed. A further objective is to find out if (numerical) features or attributes can be identified for reliably representing properties of the colonisation so that a sub-classification can be established, similar to say, that of Hendrick and Foster-Smith (2006).

Resources required are mostly data; sidescan sonar mosaics containing known Sabellaria colonies, ground truth descriptions and human interpretation/classification (i.e. manually produced class maps, preferably from more than one rater) delineating the regions of the seabed and showing the different types of Sabellaria colonisation, for instance, sparse, moderate and dense.

An estimated time scale for carrying out this work is one year. Output from the research is one conference publication/presentation and one journal publication. Results would have domain interests covering machine learning and the marine habitat classification/mapping communities.

Appendix 1

IEEE International conference on image Processing (ICIP), poster presentation, Brussels, September 2011.

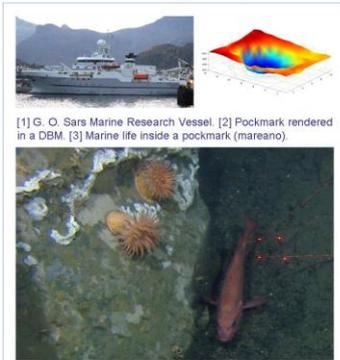
A NEW APPROACH TO THE AUTOMATED MAPPING OF POCKMARKS IN MULTI-BEAM BATHYMETRY

Richard Harrison^{1,2}, Valérie Bellec³, Dave Mann², Wenjia Wang¹

¹School of Computing Sciences, University of East Anglia, Norwich, UK, ²Gardline Geosurvey Ltd., Great Yarmouth, UK, ³Geological Survey of Norway (NGU), Trondheim, Norway.

1 Pockmarks

Pockmarks are natural seabed depressions caused by fluid venting. They can occur in swarms containing thousands of individuals. Escaping gases may form Methane Derived Authigenic Carbonates (MDAC). MDAC structures inside pockmarks are important habitats, protected by the EC. They are also indicators of possible hydrocarbon reserves. Seabed instability caused by pockmarks has implications for engineering projects too. So, identifying and mapping the pockmarks is clearly an important task.



2 Problem

Pockmarks are rendered in Digital Bathymetry Models (DBM), generated from data captured by a Multi-Beam Echo-Sounder (MBES). Mapping pockmarks manually is tedious and subjective. It would therefore be useful to devise a process for automatically identifying and mapping the pockmarks.

3 Aims

Assuming a pockmark is a collection of DBM grid nodes, we investigate features for node representation and create supervised machine learning models for their classification.

4 Feature creation

Features are created at multiple resolutions and include, amongst others,

•Laplacian of a Gaussian

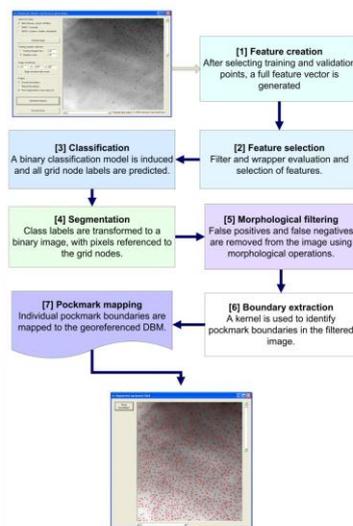
$$z(i, j) = \frac{1}{N} \nabla^2 G(x, y) * Z(i, j)$$

x, y are local coordinates with origin i, j and N is a scale dependent coefficient.

•Co-occurrence textural features

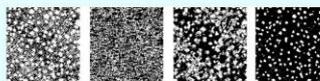
Contrast, entropy and correlation were used for investigating the class discriminatory potential of DBM textures.

Process overview



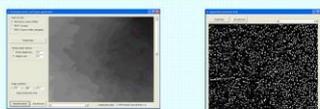
5 Feature selection

Pockmarks have wide ranging shapes and physical dimensions. Resolution of the objects also varies in different DBM's. A salient set of features is required to ensure accurate and efficient classification. Feature responses in filter and wrapper evaluations on validation data are used to select suitable candidates. The qualitative impact of different feature sets is shown below.



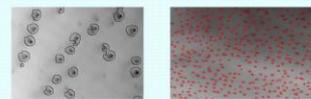
6 Classification

A Ball Vector Machine (BVM) was deployed in the test cases for hard labelling of grid nodes. Segmentation is accomplished by mapping classified nodes to a binary image space. Morphological filtering is applied to remove false-positives and false-negatives.



7 Mapping

After filtering, a kernel is applied to extract and map object boundaries. Individual objects are identified with up to 96 % accuracy.

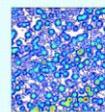


8 Pockmark types

Individual pockmarks have diverse characters. They can be grouped into different types, according to their shapes and other properties such as the ratio of surface area to the projected surface area, R .

$$R(i, j) = \frac{1}{((k - 1)d)^2} \left\{ \sum_{m=1}^{|i|-1} \sum_{n=1}^{|j|-1} \Delta(m, n) - ((k - 1)d)^2 \right\}$$

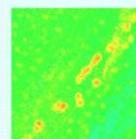
k is the computational kernel dimension, d of the DBM resolution and Δ_t the elemental triangular area, from Heron's formula.



Pockmarks with red centres have relatively high values of R . This could be related to specific physical or geological properties, peculiar to these individuals.

9 Data fusion

Combining data from multiple sensors broadens the scope of the process and facilitates refinements in the discrimination of pockmark types, based on their natural properties and the attributes used to represent them. Using features derived from MBES bathymetry together with acoustic backscatter responses is an example.



Pockmarks with red centres have strong backscatter responses. Higher acoustic reflectivity indicates these pockmarks are distinct from the others in some way.

10 Summary

A new, feature based machine learning approach to the automated mapping of pockmarks has been proposed. Prototype software has been designed to efficiently perform this task with proven accuracy.



Acknowledgements: Ph.D research supported by the Grant (R17336) jointly funded by EPSRC, UEA and Gardline Geosurvey Ltd., of Great Yarmouth, UK. Thanks to Margaret Dolan of the NGU who worked on the identification of pockmarks in these data using a GIS package. Margaret produced shape files and other useful information which was helpful during our research.

References

A

Aksoy, S. and Haralick, R. (2000) Probabilistic vs. geometric similarity measures for image retrieval, proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2, 357–362.

Allain, C. and Cloitre, M. (1991) Characterising the lacunarity of random and deterministic fractal sets, *Physical review A*44 (6).

Amayeh, G., Tavakkoli, A. and Bebis, G. (2009) Accurate and efficient computation of gabor features in real-time applications, *Advances in Visual Computing*, 243-252, Springer.

Anderson, J., Holliday, D., Kloser, R., Reid, D. and Simard, Y. (2008) Acoustic seabed classification: current practice and future directions, *ICES Journal of Marine Science*, 65, 1004 –1011, doi:10.1093/icesjms/fsn061.

Androutsos, D., Plataniotiss, K. and Venetsanopoulos, A. (1998) Distance measures for color image retrieval, proceedings of the IEEE International Conference on Image Processing, 2, 770–774.

Apel, J. (1987) *Principles of Ocean Physics*, International Geophysics series, Volume 38, Academic Press, London.

Atallah, L., (2004) *Learning from Sonar Data for the Classification of Underwater Seabeds*, Ph.D thesis, Robotics Research Group, Department of Engineering Science, University of Oxford, <http://www.doc.ic.ac.uk/~latallah/thesis.pdf>.

Augustin, J. and Lurton, X. (2005) Image amplitude calibration and processing for seafloor mapping sonars, proceedings of IEEE Oceans 2005 – Europe, 1, 698 – 701, doi:10.1109/OCEANSE.2005.1511799.

B

Baruque, B., Corchado, E. and Yin, H. (2011) The S2 - ensemble fusion algorithm, *International Journal of Neural Systems*, 21(6), 505–525.

Basseville, M. (1989) Distance measures for signal processing and pattern recognition, *Signal Processing*, 18 (4), 349–369.

Bazeley, P. (1999) The bricoleur with a computer: Piecing together qualitative and quantitative data, *Qualitative Health Research*, 9 (2), 279-287.

Beaudoin, J., Hughes Clarke, J., Van Den Aemele, E. and Gardner J. (2002) Geometric and radiometric correction of multibeam backscatter derived from Reson 8101 systems, *Proceedings of the Canadian Hydrographic Conference 2002*, Canadian Hydrographic Association, Ottawa, Ontario,
http://www.omg.unb.ca/omg/papers/Beaudoin_Multibeam_Backscatter_Reson_8101_Systems.pdf.

Begon, M., Harper, J. and Townsend, C. (1990) *Ecology: Individuals, Populations and Communities*. Blackwell Scientific Publications.

Bell, J. (1995) *A model for the synthesis of sidescan sonar data*, PhD thesis, Heriot-Watt University.

Bell, J. Chantler, M. and Wittig, T. (1999) Sidescan sonar: a directional filter of seabed texture? *IEE Proceedings on Radar, Sonar and Navigation*, 146(1), 65-72.

Bell, J. and Linnett, L. (1997) Simulation and analysis of synthetic sidescan sonar images, *IEE Proceedings on Radar, Sonar and Navigation*, 144 (4), 219-226.

Bellec, V., Van Lancker, V., Degrendele, K., Roche, M. and Le Bot, S. (2010) Geo-environmental characterization of the Kwinte Bank, *Journal of coastal Research*, 51, 63-76, doi:10.2112/SI51-006.1.

- Bhattacharyya, A. (1943) On a measure of divergence between two statistical populations defined by their probability distributions, *Bulletin of the Calcutta Mathematical Society*, 35, (99-109).
- Bianconi, F. and Fernández, A. (2007) Evaluation of the effects of Gabor filter parameters on texture classification, *Pattern Recognition*, 40, 3325–3335.
- Bianconi, F., Fernández, A. and Mancini, A. (2008) Assessment of rotation-invariant texture classification through Gabor filters and discrete Fourier transform, *Proceedings of the 20th International conference on Graphical Engineering*, Valencia, Spain, The Spanish Association for Graphical engineering (INGEGRAF.)
- Billi, A., Minelli, L., Orecchio, B. and Presti, D. (2010) Constraints to the Cause of Three Historical Tsunamis (1908, 1783, and 1693) in the Messina Straits Region, Sicily, Southern Italy *Seismological Research Letters*, 81, 907-915.
- Birchall, R. (2007) Determining the extent of Sabellaria reefs during seabed surveys, *Proceeding of the 6th international offshore site investigation and geotechnics conference, Confronting new challenges and sharing knowledge*, London, UK.
- Blondel, P. (1996) Segmentation of the Mid-Atlantic Ridge south of the Azores, based on acoustic classification of TOBI data. In: MacLeod, C., Tyler, P., Walker, C. (Eds.), *Tectonic, Magmatic, Hydrothermal and Biological Segmentation of Mid-Ocean Ridges*. Geological Society Special Publication No. 118, Boulder, CO, 17–28.
- Blondel, P. (2007). *The handbook of sidescan sonar*. Springer.
- Blondel, P. and Gómez Sichi, O. (2009) Textural analyses of multibeam sonar imagery from Stanton Banks, Northern Ireland continental shelf, *Applied Acoustics*, 70 (10), 1288–1297.
- Blondel, P., Parson, L. and Robigou, V. (1998) TexAn: textural analysis of sidescan sonar imagery and generic seafloor characterisation, *Proceedings of IEEE Oceans*, 419-423.

Blondel, P., Sempéré, J., Robigou, V. (1993) Textural analysis and structure tracking for geological mapping: Applications to sonar data from Endeavour Segment, Juan de Fuca Ridge, proceedings of IEEE Oceans, 209-213.

Bogaart, P. and Troch, P. (2006) Curvature distribution within hillslopes and catchments and its effect on the hydrological response, *Hydrology and Earth System Sciences Discussions* 3 (3),1071-1104.

Bovik, A., Clark, M. and Geisler, W. (1990) Multichannel texture analysis using localized spatial filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (1), 55-73.

Breiman, L. (2001) Random forests, *Machine Learning*, 45(1), 5–32.

Brown, C. and Collier, J. (2008) Mapping benthic habitat in regions of gradational substrata: An automated approach utilising geophysical, geological, and biological relationships, *Estuarine, Coastal and Shelf Science*, 78, 203-214, doi:10.1016/j.ecss.2007.11.026.

Bruzzone, L, Chi, M and Marconcini, M (2006) A novel transductive SVM for semisupervised classification of remote-sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 44 (11), 3363.

Bugatti, P., Traina, A. and Traina, C. (2008) Assessing the best integration between distance-function and image-feature to answer similarity queries, in *Proceedings of the ACM Symposium on Applied Computing*, 1225–1230.

Buhl-Mortensen, P., Dolan, M. and Buhl-Mortensen, L. (2009) Prediction of benthic biotopes on a Norwegian offshore bank using a combination of multivariate analysis and GIS classification, *ICES Journal of Marine Science*.

Butson, C. and King, D. (2006) Lacunarity analysis to determine optimum extents for sample-based spatial information extraction from high-resolution forest imagery, *International Journal of Remote Sensing*, 27 (1), 105-120.

C

- Camapum Wanderley, J. and Fisher, M. (2001) Multiscale color and texture invariants for image recognition, proceedings of the IEEE International Conference on Image Processing, 1, 862-865.
- Campbell, F. and Kulikowski, J. (1966) Orientational selectivity of the human visual system, *The Journal of physiology*, 187 (2), 437.
- Capus, C., Banks, A., Coiras, E., Ruiz, I, Smith, C. and Petillot, Y. (2008) Data correction for visualisation and classification of sidescan SONAR imagery, *Radar, Sonar & Navigation, IET*, 2 (3), 155-169, doi:10.1049/iet-rsn:20070032.
- Carmichael, D., Linnett, L., Clarke, S. and Calder, B. (1996) Seabed classification through multifractal analysis of sidescan sonar imagery, *IEE Proceedings on Radar, Sonar and Navigation*, 143(3).
- Celik, T. and Tjahjadi, T. (2011) A novel method for sidescan sonar image segmentation, *IEEE Journal of Oceanic Engineering*, 36, 186–194.
- Cervenka, P. and de Moustier, C. (1993) Sidescan sonar image processing techniques, *IEEE Journal of Oceanic Engineering*, 188 (2), 108-122, doi: 10.1109/48.219531.
- Cexus, J. and Boudraa, A. (2003) Multi-channel model for sonar image segmentation, in the Seventh International Symposium on Signal Processing and Its Applications, *Proceedings*, 2.
- Cha, S. (2007) Comprehensive survey on distance/similarity measures between probability density functions, *International Journal of Mathematical Models and Methods in Applied Sciences*, 4 (1), 300–307.
- Chailloux, C., Le Caillec, J., Gueriot, D. and Zerr, B. (2011) Intensity-Based Block Matching Algorithm for Mosaicing Sonar Images, *IEEE Journal of Oceanic Engineering*, 36 (4), 627-645, doi:10.1109/JOE.2011.2141850.

-
- Chan, J., Huang, C. and DeFries, R. (2001) Enhanced Algorithm Performance for Land Cover Classification from Remotely Sensed Data Using Bagging and Boosting, *IEEE Transactions on Geoscience and Remote Sensing*, 39 (3), 693-695.
- Chan, J. and Paelinckx, D. (2008) Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery, *Remote Sensing of Environment*, 112 (6), 2999-3011.
- Chan, T. and Vese, L. (2001) Active contours without edges, *IEEE Transactions on Image Processing*, 10 (2), 266-277.
- Chand, S., Rise, L., Ottesen, D., Dolan, M., Bellec, V., and Bøe, R. (2009) Pockmark-like depressions near the Goliat hydrocarbon field, Barents Sea: Morphology and genesis, *Marine and Petroleum Geology*, 26 (7), 1035–1042.
- Chantler, M. and McGunnigle, G. (1995) Compensation of illuminant tilt variation for texture classification, *IEE Fifth International Conference on Image Processing and its Applications*, 767-771.
- Chantler, M., Russell, G. and Linnett, L. (1994) Illumination: A directional filter of texture? *British Machine Vision Conference*, 2, 449-458.
- Chavez, P., Isbrechta, J., Galanis, P., Gabel, G., Sides, S., Soltesz, D., Ross, S. and Velasco, M. (2002) Processing, mosaicking and management of the Monterey Bay digital sidescan-sonar images, *Marine Geology*, 181 (1–3), 305–315.
- Cherkauer, K. (1996) Human expert-level performance on a scientific image analysis task by a system using combined artificial neural networks, *Working Notes of the AAAI Workshop on Integrating Multiple Learned Models*, 15-21.
- Chotiros, N. (2006) Seafloor acoustic backscattering strength and properties from published data, *proceedings of IEEE Oceans - Asia Pacific*, doi:10.1109/OCEANSAP.2006.4393905.

- Christensen, O. (2007) SUSHIMAP Survey strategy and methodology for marine habitat mapping, PhD thesis, NTNU, Dept. of Electronics and Telecomm., Trondheim, Norway.
- Clausi, D. (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, 28 (1), 45–62.
- Clausi, D. and Jernigan, M. (2000) Designing Gabor filters for optimal texture separability, *Pattern Recognition*, 33 (11), 1835–1849.
- Cobra, D., Oppenheim, A. and Jaffe, J. (1992) Geometric distortions in side-scan sonar images: a procedure for their estimation and correction, *IEEE Journal of Oceanic Engineering*, 17 (3), 252-268, doi:10.1109/48.153442.
- Coggins, J. and Jain, A. (1985) A spatial filtering approach to texture analysis, *Pattern Recognition Letters*, 3 (3), 195-203.
- Cohen, W., Ravikumar, P. and Fienberg, S. (2003) A comparison of string distance metrics for name-matching tasks, in *Proc. IJCAI-2003 Workshop on Information Integration on the Web*, 73–78.
- Coiras, E., Ruiz, I., Petillot, Y., Lane, D. (2004) Fusion of multiple side-scan sonar views, *proceedings of IEEE Oceans*, 2036 - 2054, doi: 10.1109/OCEANS.2004.1406456.
- Collier, J. and Brown, C. (2005) Correlation of sidescan backscatter with grain size distribution of surficial seabed sediments, *Marine Geology*, 214, 431 – 449, doi:10.1016/j.margeo.2004.11.011.
- Cutter, G., Rzhhanov, Y. and Mayer, L. (2003) Automated segmentation of seafloor bathymetry from multibeam echosounder data using local Fourier histogram texture features, *Journal of Experimental Marine Biology and Ecology*, 285,355-370.
- Czarnecki, M. (1979). An application of pattern recognition techniques to side scan sonar data, *proceedings of IEEE Oceans*,112-119.

D

- Dartnell, P. and Gardner, J. (2004) Predicting Seafloor Facies from Multibeam Bathymetry and Backscatter Data, *Photogrammetric Engineering and Remote Sensing*, 70, 1081 - 1091.
- Dartnell, P. and Gardner, J. (2009) Seafloor terrain analysis and geomorphology of the greater Los Angeles Margin and San Pedro Basin, Southern California. *Earth science in the urban ocean: the Southern California continental borderland*, 454.
- Dasarathy, B. (1994) *Decision fusion*, IEEE computer Society Press, Los Alamitos, CA.
- Daugman, J. (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Journal of the Optical Society of America*, 2 (7), 1160-1169.
- Degraer, S., Moerkerke, G., Rabaut, M., Vanhoey, G., Dufour, I., Vincx, M., Henriët, J. and Vanlancker, V. (2008) Very-high resolution side-scan sonar mapping of biogenic reefs of the tube-worm *Lanice conchilega*, *Remote Sensing of Environment*, 112 (8), 3323-3328, doi:10.1016/j.rse.2007.12.012.
- De Moustier, C. (2009) Oblique incidence characterization methods, Volume 2, Section 22, In: Hughes Clarke, J., de Moustier, C., Mayer, L. and Wells, D. (2009) 50th UNB-OMB / UNH-CCOM Multibeam Sonar Training Course Stavanger Norway.
- Diesing, M., Coggan, R. and Vanstaen, K. (2009) Widespread rocky reef occurrence in the central English Channel and the implications for predictive habitat mapping, *Estuarine Coastal and Shelf Science*, 83 (4), 647-658.
- Dietterich, T. (2000). Ensemble methods in machine learning. Multiple classifier systems, 1-15.
- Dietterich, T. and Michalski, R. (1983) A comparative review of selected methods for learning by example. In, Michalski, R., Carbonell, J. and Mitchell, T. (Eds.), *Machine Learning: An artificial intelligence approach*, 41-81, Tioga Publishing Co., Palo Alto.

Dolan, M., Grehan, A., Guinan, J., Brown, C. (2008) Modelling the local distribution of cold-water corals in relation to bathymetric variables: Adding spatial context to deep-sea video data. *Deep Sea Research Part I: Oceanographic Research Papers*, 55 (11), 1564-1579.

Domingos, P. and Pazzani, M. (1997) On the optimality of the simple Bayesian classifier under zero-one loss, *Machine Learning*, 29 (2-3), 103-130.

Du Buf, J. and Heitk, P. (1991) Texture features based on Gabor phase, *Signal Processing*, 23 (3), 227–244.

Duda, R. and Hart, P. (1972) Use of the Hough transformation to detect lines and curves in pictures, *Communications of the ACM*, 15 (1), 11-15.

Duda, R., Hart, P., Stork, D. and others (2001) *Pattern Classification*, 2nd edn. John Wiley and Sons Ltd., New York.

Dudani, A. (1976) The distance-weighted k-nearest-neighbor rule, *IEEE Transactions on Systems, Man and Cybernetics*, 4, 325-327.

Dunlop, J. (1997) Statistical modelling of sidescan sonar images, *proceedings of IEEE Oceans*, 1, 33-38.

Dunn, D., Higgins, W. and Wakeley, J., (1994) Texture segmentation using 2-D Gabor elementary functions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 (2), 130-149.

E

Ehsani, A. and Quiel, F. (2008) Geomorphometric feature analysis using morphometric parameterization and artificial neural networks, *Geomorphology* 99 (1-4), 1-12.

Emran, S. and Ye, N. (2001) Robustness of canberra metric in computer intrusion detection, proceedings of the IEEE Workshop on Information Assurance and Security, West Point, NY, USA.

F

Fan, G. and Xia, X. (2003) Wavelet-based texture analysis and synthesis using hidden markov models, *IEEE Transactions on Circuits and Systems*, 50, 106-120.

Fernández, A., Ghita, O., González, E., Bianconi, F. and Whelan, P. (2010) Evaluation of robustness against rotation of LBP, CCR and ILBP features in granite texture classification, *Machine Vision and Applications*, 22 (6), 913-926.

Fernández, A., X. Álvarez, M. and Bianconi, F., (2011) Image classification with binary gradient contours, *Optics and Lasers in Engineering*, 49, 1177–1184.

Fish, J. and Carr, H. (2001) *Sound reflections: advanced applications of side scan sonar*, Lower Cape Publishing.

Fisher, P., Wood, J. and Cheng, T. (2004) Where is Helvellyn? Fuzziness of multi-scale landscape morphometry, *Transactions of the Institute of British Geographers*, 29 (1), 106–128.

Fix, E. and Hodges, J., Jr. (1951) Discriminatory analysis, nonparametric discrimination: Consistency Properties, USAF School of Aviation Medicine, Randolph Field, TX, Project No. 21-49-004, Rep. No. 4, Contract No. AF41(128)-31.

Folk, R. and Ward, W. (1957) Brazos river bar: A study in the significance of grain size parameters, *Journal of Sedimentary Petrology*, 27 (1), 3-26.

Fonseca, L., Mayer, L., Orange, D., Driscoll, N. (2002) The high frequency backscattering angular response of gassy sediments: model/data comparison from the Eel River margin, California, *Journal of the Acoustical Society of America*, 111 (6), 2621–2631.

Fonseca, L. and Mayer, L. (2007) Remote estimation of surficial seafloor properties through the application of Angular Range analysis to multibeam sonar data, *Marine Geophysical Research*, 28, 119-126, doi:10.1007/s11001-007-9019-4.

Fonseca, L., Brown, C., Calder, B., Mayer, L. and Rzhanov, Y. (2009) Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures, *Applied Acoustics*, 70 (10), 1298-1304.

Foote, J. (1997) Content-based retrieval of music and audio, *Multimedia Storage and Archiving Systems II, Proc. of SPIE*, 3229, 138–147.

Fowler, J., Devillers, R., Simms, A., Bolduc, A. (2008) Identification and characterization of seabed pockmarks in the St. Lawrence Estuary, Quebec 2008: 400 Years of Discoveries. Joint Meeting of the Geological Association of Canada, Mineralogical Association of Canada, Society of Economic Geologists and the Society for Geology Applied to Mineral Deposits. May 26-28, 2008. Quebec City Convention Centre, Quebec.

Franklin, S. (1987) Geomorphometric processing of digital elevation models, *Computers and Geosciences*, 13, (6), 609.

Franklin, S. and Peddle, D. (1987) Texture analysis of digital image data using spatial cooccurrence, *Computers and Geosciences*, 13 (3), 293-311.

Fraschetti, S., Terlizzi, A. and Boero, F. (2008) How many habitats are there in the sea (and where)? *Journal of Experimental Marine Biology and Ecology*, 366 (1), 109-115.

Fred, A. and Jain, A. (2002) Data clustering using evidence accumulation, In proceeding of the 16th IEEE international conference on Pattern Recognition, 4, 276-280.

Friedman, G. and Sanders, F. (1978) *Principles of sedimentology*, Wiley, New York.

G

Gabor, D. (1946) Theory of Communication, Journal of the Institute of Electrical Engineers (London), 93, 429-457.

Gafeira, J., Long, D. and Diaz-Doce, D. (2012) Semi-automated characterisation of seabed pockmarks in the central North Sea, Near Surface Geophysics, 10, 301-312
doi:10.3997/1873-0604.2012018

Galloway, J. (2008) Systematic acoustic seafloor habitat mapping of the British Columbia coast, Habitat Mapping Technology for Alaska, J.R. Reynolds and H.G. Greene (eds.), Alaska Sea Grant College Program, University of Alaska Fairbanks.
doi:10.4027/mhmta.2008.14.

Galparsoro, I., Borja, A., Bald, J., Liria, P. and Chust, G. (2009) Predicting suitable habitat for the European lobster (*Homarus gammarus*), on the Basque continental shelf (Bay of Biscay), using Ecological-Niche Factor Analysis. Ecological Modelling, 220 (4), 556-567.

Gefen, Y., Meir, Y., Mandelbrot, B. and Aharony, A. (1983) Geometric implementation of hypercubic lattices with non-integer dimensionality by use of low lacunarity fractal lattices, Physical Review Letters, 50, 145–148.

Geman, D., Geman, S., Graffigne, C. and Dong, P (1990) Boundary detection by constrained optimization, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (7), 609–628.

Gensane, M. (1989) A statistical study of acoustic signals backscattered from the sea bottom, IEEE Journal of Oceanic Engineering, 14 (1), 84-93.

Gislason, P., Benediktsson, J. and Sveinsson, J. (2006) Random Forests for land cover classification, Pattern Recognition Letters, 27 (4), 294-300.

Goff, J., Olson, H. and Duncan, C. (2000) Correlation of side-scan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin. *Geo-Marine Letters*, 20 (1), 43-49.

Gonzalez, R. and Woods, R. (2008) *Digital image processing*, Third Edition, Prentice Hall.

Gorini, M. (2009) Physiographic Classification of the Ocean Floor: A Multi-Scale Geomorphometric Approach, *Proceedings of Geomorphometry*, Zurich, Switzerland, 31 August - 2 September.

Grigorescu, S., Petkov, N. and Kruizinga, P. (2002) Comparison of texture features based on Gabor filters, *IEEE Transactions on Image processing*, 11 (10), 1160–1167.

Guinan, J., Grehan, A., Dolan, M. and Brown, C. (2008) Quantifying relationships between video observations of cold-water coral cover and seafloor features in Rockall Trough, west of Ireland, *Marine Ecology Progress Series*, 375, 125-138.

Guyon, I. and Elisseeff, A. (2003) An introduction to variable and feature selection, *The Journal of Machine Learning Research*, 3, 1157–1182.

H

Hadjitodorov, S., Kuncheva, L. and Todorova, L. (2006) Moderate diversity for better cluster ensembles, *Information Fusion*, 7(3), 264-275.

Hamilton, L. (2005) A bibliography of acoustic seabed classification, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management, Technical Report No. 27. http://www.ozcoasts.gov.au/pdf/CRC/27-Hamilton_Acoustic_Biblio.pdf [Accessed February 2012.]

Hammerstad, E. (2000) Backscattering and Seabed Image Reflectivity, Kongsberg hydroacoustic products, Multibeam echosounders, EM series, EM Technical Note, [http://www.km.kongsberg.com/ks/web/nokbg0397.nsf/AllWeb/226C1AFA658B1343C1256D4E002EC764/\\$file/EM_technical_note_web_BackscatteringSeabedImageReflectivity.pdf?OpenElement](http://www.km.kongsberg.com/ks/web/nokbg0397.nsf/AllWeb/226C1AFA658B1343C1256D4E002EC764/$file/EM_technical_note_web_BackscatteringSeabedImageReflectivity.pdf?OpenElement).

Hanbury, A., Kandaswamy, U. and Adjeroh, D. (2005) Illumination-invariant morphological texture classification, in Proceedings of the 7th International Symposium on Mathematical Morphology, Paris (France).

Haralick, R., Shanmugam, K. and Dinstein, I. (1973) Textural features for image classification, IEEE Transactions on systems, man and cybernetics, 3 (6), 610–621.

Harris, P. and Baker, E. (2012) Seafloor Geomorphology as benthic habitat: Geohab Atlas of seafloor geomorphic features and benthic habitats, Harris, P. and Baker, E., Eds., Elsevier, London.

Harrison, R. (2009), A spatial model of a towed cable system and USBL acoustic beacon, in Società Geologica Italiana, proceedings of the International conference on seafloor mapping for geohazard assessment, Eds., Chiocci, F., Ridente, D., Casalbore, D. and Bosman, A., 7, 59-62.

Harrison, R., Birchall, R., Mann, D. and Wang, W. (2011) A novel ensemble of distance measures for feature evaluation: Application to sonar imagery, IDEAL 12th International Conference Proceedings, Lecture Notes in Computer Science, 6936, 327–336.

Hellequin, L, Boucher, J. and Lurton, X. (2003) Processing of High-Frequency Multibeam Echo Sounder Data for Seafloor Characterization, IEEE Journal of Oceanic Engineering, 28 (1), 78 – 89, doi:10.1109/JOE.2002.808205.

Hendrick, V. and Foster-Smith, R. (2006) Sabellaria spinulosa reef: a scoring system for evaluating reefiness in the context of the Habitats Directive, Journal of the Marine Biological Association of the United Kingdom, 86, 665–677.

- Henry, L., Davies, A. and Murray Roberts, J. (2010) Beta diversity of cold-water coral reef communities off western Scotland, *Coral Reefs*, 29 (2), 427-436.
- Holliday, D. (2007) Theory of sound-scattering from the seabed, in ICES cooperative research report, acoustic seabed classification of marine physical and biological landscapes, John T Anderson, Editor, <http://www.ices.dk/pubs/crr/crr286/CRR286.pdf>.
- Holmes, K., Van Niel, K., Radford, B., Kendrick, G. and Grove, S. (2008) Modelling distribution of marine benthos from hydroacoustics and underwater video, *Continental Shelf Research*, 28 (14), 1800-1810.
- Hou, T. and Hough, L. (2004) Seabed characterization using normalized backscatter data by best estimated grazing angles, *International symposium on underwater technology*, 153-160, doi: 10.1109/UT.2004.1405514.
- Hovland, M., Gardner, J. and Judd, A. (2002) The significance of pockmarks to understanding fluid flow processes and geohazards, *Geofluids*, 2 (2), 127–136.
- Hubert, L. and Arabie, P. (1985) Comparing partitions, *Journal of Classification*, 2 (1), 193-218.
- Hughes Clarke, J. (2004) Seafloor Characterization Using Keel-Mounted Sidescan: Proper Compensation for Radiometric and Geometric Distortion, *Proceedings of the Canadian Hydrographic Conference*, Ottawa.
- Hughes Clarke, J., Danforth, B. and Valentine, P. (1997) Areal seabed classification using backscatter angular response at 95 kHz. In *SACLANTCEN Conference on High Frequency Acoustics in Shallow Water*, 243-250.
- Hughes Clarke, J., de Moustier, C., Mayer, L. and Wells, D. (2009) 50th UNB-OMB / UNH-CCOM Multibeam Sonar Training Course, Stavanger Norway.

I

Ierodiaconou, D., Laurenson, L., Burqc, S. and Reston, M. (2007) Marine benthic habitat mapping using multibeam data, georeferenced video and image classification techniques in Victoria, Australia, *Journal of Spatial Science*, 52 (1), doi:10.1080/14498596.2007.9635105.

IHO (2011), *Manual of Hydrography*, IHO publication C-13, http://www.iho.int/iho_pubs/CB/C-13/english/C-13_Chapter_4.pdf

Iwahashi, J. and Pike, R. (2007) Automated classifications of topography from DEM's by an unsupervised nested-means algorithm and a three-part geometric signature, *Geomorphology*, 86 (3-4), 409-440.

J

Jackson, D. (1994) Models for scattering from the seabed, *Proceedings of the Institute of Acoustics*, 16 (6).

Jackson, D. and Richardson, M. (2007) *High-frequency seafloor acoustics*, The underwater acoustics series, Springer, New York.

Jackson, D., Winebrenner, D. and Ishimaru, A (1986) Application of the composite roughness model to high-frequency bottom backscattering, *Journal of the Acoustical Society of America*, 79 (5), 1410-1422.

Jain, A. and Farrokhnia, F. (1991) Unsupervised texture segmentation using Gabor filters, *Pattern recognition*, 24 (12), 1167–1186.

Jakeman, E. and Pusey, P. (1984) A model for non-Rayleigh sea echo, *IEEE Transactions on Antennas and Propagation*, 24, 804-814.

Jin, H., Liu, Q., and Tong, X. (2004) Face detection using improved LBP under Bayesian framework,” in *Proceedings of the 3rd International Conference on Image and Graphics*, 306–309.

- Johnson, H. and Helferty, M. (1990) The geological interpretation of side-scan sonar, *Reviews of Geophysics*, 28 (4), 357-380, doi:10.1029/RG028i004p00357.
- Jonsson, P., Sillitoe, I., Dushaw, B., Nystuen, J., and Heltne, J. (2009) Observing using sound and light—a short review of underwater acoustic and video-based methods, *Ocean Science Discussions*, 6, 819–870.
- Judd, A. (2001) Pockmarks in the UK sector of the North Sea, Technical Report TR 002 produced for the Strategic Environmental Assessment – SEA2.
- Judd, A. and Hovland, M. (2007) *Seabed Fluid Flow*, Cambridge University Press, New York.

K

- Kalcic, M. and Bibee, D. (2004) Textural segmentation of high-resolution sidescan sonar images, Sixth International Symposium on Technology and the Mine Problem, Monterey, CA, 9-13 May 2004.
- Kalousis, A., Prados, J. and Hilario, M. (2007) Stability of feature selection algorithms: A study on high dimensional spaces, *Knowledge and Information Systems* 12(1), 95–116.
- Karoui, I., Fablet, R., Boucher, J. and Augustin, J. (2005) Statistical discrimination of seabed textures in sonar images using co-occurrence statistics, *proceedings of IEEE Oceans*, 605–610.
- Karoui, I., Fablet, R., Boucher, J., Pieczynski, W. and Augustin, J. (2008) Fusion of textural statistics using a similarity measure: application to texture recognition and segmentation, *Pattern Analysis and Applications*, 11 (3), 425-434.
- Kaye, B. (1989) *A random walk through fractal dimensions*, VCH Publisher, New York.

Kenny, A., Todd, B. and Cooke, R. (2001) Procedural Guidelines No. 1-4–The Application of Sidescan Sonar for Seabed Habitat Mapping, Marine Monitoring Handbook, The Joint Nature Conservation Committee, Peterborough, United Kingdom.
http://www.jncc.gov.uk/marine/mmh/MMH_0601.pdf

Kieser, R., Preston, J., Orłowski, A. and Chapman, R. (2007) Acquiring and preparing acoustic data in ICES cooperative research report, acoustic seabed classification of marine physical and biological landscapes, John T Anderson, Editor,
<http://www.ices.dk/pubs/crr/crr286/CRR286.pdf>.

Kim, J., Muller, J., van Gasselt, S., Morley, J. and Neukum G. (2005) Automated crater detection, a new tool for Mars cartography and chronology, *Photogrammetric engineering and remote sensing*, 71 (10).

King, L. and MacLean, B. (1970) Pockmarks on the Scotian shelf, *Geological Society of America Bulletin*, 81 (10).

Kingsbury, N. (1998) The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters, *Proceeding of the 8th IEEE DSP Workshop*, 8.

Kloser, R (2007) Seabed backscatter, data collection and quality overview, in ICES cooperative research report, acoustic seabed classification of marine physical and biological landscapes, John T Anderson, Editor,
<http://www.ices.dk/pubs/crr/crr286/CRR286.pdf>.

Kocak, D., Dalglish, F., Caimi, F. and Schechner, Y. (2008) A focus on recent developments and trends in underwater imaging, *Marine Technology Society Journal*, 42 (1), 52–67.

Kohavi, R. and John, G. (1997) Wrapper for feature subset selection, *Artificial Intelligence*, 97(1), 273–324.

Kohavi, R. and Sommerfield, D. (1995) Feature subset selection using the wrapper model: Overfitting and dynamic search space topology, in *Proceeding of the 1st International Conference on Knowledge Discovery and Data Mining*, 192–197.

Kohonen, T. (1990) The self-organizing map, proceedings of the IEEE, 78 (9), 1464–1480.

Kononenko, I. (1994) Estimating Attributes: Analysis and extensions of RELIEF, in Proceedings of the 7th European Conference on Machine Learning, 171–182.

Křížek, P., Kittler, J. and Hlaváč, V. (2007) Improving stability of feature selection methods, in Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns, 929–936.

Kullback, S. and Leibler, R. (1951) On information and sufficiency, Annals of Mathematical Statistics, 22(1), 79–86.

Kuncheva, L. and Hadjitodorov, S. (2004) Using diversity in cluster ensembles, proceedings of the IEEE conference on Systems, man and cybernetics, 2, 1214-1219.

Kurmyshev, E., Sánchez-Yáñez, R. and Cuevas, F. (2003) A framework for texture classification using the Coordinated Clusters Representation, Pattern Recognition Letters, 24, 21–31.

L

Lalgudi, K., Papaefthymiou, M. and Potkonjak, M. (2000) Optimizing computations for effective block-processing, ACM Transactions on Design Automation of Electronic Systems (TODAES),5(3), 604-630.

Lanaaya, H., Martin, A., Aboutajdine, D. and Khenchaf, A. (2005a) A new dimensionality reduction method for seabed characterization: supervised curvilinear component analysis, proceedings of IEEE Oceans, 1,339-344.

Lanaaya, H., Martin, A., Khenchaf, A. and Aboutajdine, D. (2005b) Feature selection using genetic algorithm for sonar images classification with support vector, European Conference on Propagation and Systems, Brest, France.

-
- Lastras, G., Canals, M., Urgeles, R., Hughes-Clarke, J. and Acosta, J. (2004) Shallow slides and pockmark swarms in the Eivissa Channel, western Mediterranean Sea, *Sedimentology*, 51 (4), 837–850.
- Lawrence, N. and Scholkopf, B. (2001) Estimating a kernel Fisher discriminant in the presence of label noise, In *machine learning international workshop*, 306-313.
- Lawver, L., Davis, M., Wilson, T. and Party, S. (2007) Neotectonic and other features of the Victoria Land Basin, Antarctica, interpreted from multibeam bathymetry data, 10th International Symposium on Antarctic Earth Sciences.
- Le Bas, T. and Huvenne, V. (2009) Acquisition and processing of backscatter data for habitat mapping – Comparison of multibeam and sidescan systems, *Applied Acoustics*, 70 (10), 1248-1257.
- Lewin, K. (1946) Action research and minority problems, *Journal of Social Issues*, 2, 34-46.
- Li, M. and Sethi, I. (2006) Confidence-based active learning, *IEEE transactions on pattern analysis and machine intelligence*, 1251-1261.
- Li, T., Ogihara, M. and Ma, S. (2004) On combining multiple clusterings, *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, 294-303.
- Limpenny, D., Foster-Smith, R., Edwards, T., Hendrick, V., Diesing, M., Meadows, W., Crutchfield, Z., Pfeifer, S. and Reach, I. (2010) Best methods for identifying and evaluating *Sabellaria spinulosa* and cobble reef. Aggregate Levy Sustainability Fund Project MAL0008. Joint Nature Conservation Committee, Peterborough, ISBN - 978 0 907545 33 0.
- Liu, H. and Motoda, H. (2008) *Computational Methods of Feature Selection*, Chapman and Hall.

-
- Liu, Z., Sang, E. and Liao, Z. (2005) Sonar image segmentation using snake models based on cellular neural network, proceedings of the IEEE International Conference on Information Acquisition, 448-452, doi:10.1109/ICIA.2005.1635130.
- Long, D (2006) BGS detailed explanation of seabed sediment modified Folk classification, <http://www.searchmesh.net/PDF/BGS%20detailed%20explanation%20of%20seabed%20sediment%20modified%20folk%20classification.pdf>
- Lucieer, V. (2007a) Morphometric characterisation of rocky reef using multibeam acoustic bathymetric data, IEEE International Geoscience and Remote Sensing Symposium , 905-909.
- Lucieer, V. (2007b) Spatial uncertainty estimation techniques for shallow coastal seabed mapping, Ph.D thesis, University of Tasmania.
- Lucieer, V. (2008) Object-oriented classification of sidescan sonar data for mapping benthic marine habitats, International Journal of Remote Sensing, 29, 905 - 921.
- Lundblad, E., Wright, D., Naar, D., Donahue, B., Miller, J., Larkin, E. and Rinehart, R. (2004) Classifying deep water benthic habitats around Tutuila, American Samoa, In: Proceedings of the 24th Annual ESRI User Conference, San Diego, California.
- Lyons, A. and Abraham, D. (1999) Statistical characterization of high-frequency shallow-water seafloor backscatter, The Journal of the Acoustical Society of America, 106 (3), 1307-1315.

M

- Malhi, Y. and Román-Cuesta, R. (2008) Analysis of lacunarity and scales of spatial homogeneity in IKONOS images of Amazonian tropical forest canopies, Remote Sensing of Environment, 112 (5), 2074-2087.
- Malik, J. and Perona, P. (1990) Preattentive texture discrimination with early vision mechanisms, Journal of the Optical Society of America A, 7 (5), 923–932.

- Manjunath, B. and Ma, W. (1996) Texture features for browsing and retrieval of image data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18 (8), 837-842.
- Marr, D. and Hildreth, E. (1980) Theory of edge detection, *Proceedings of the Royal Society of London., Series B. Biological Sciences*, 207 (1167),187-217.
- Marsh, I. and Brown, C. (2009) Neural network classification of multibeam backscatter and bathymetry data from Stanton Bank (Area IV), *Applied Acoustics*, 70 (10), 1269-1276, doi:10.1016/j.apacoust.2008.07.012.
- Martin, A. (2005) Comparative study of information fusion methods for sonar images classification, in: *The Eighth International Conference on Information Fusion*, Philadelphia, USA, 25-29 July, 2005.
- Martin, A., Laanaya, H. and Arnold-Bos, A. (2006) Evaluation for uncertain image classification and segmentation, *Pattern Recognition*, 39, 1987 – 1995, doi:10.1016/j.patcog.2006.05.015.
- Martin, A. and Osswald, C. (2008) Experts Fusion and Multilayer Perceptron Based on Belief Learning for Sonar Image Classification, *proceedings of the IEEE 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, 1-6, doi:10.1109/ICTTA.2008.4530035.
- Martin, A., Sévellec, G. and Leblond, I. (2004) Characteristics vs. decision fusion for sea-bottom characterization, *In Situ Sea Bed Characterization Conference*, Brest.
- Martins, R., Pina, P., Marques, J. and Silveira, M. (2009) Crater detection by a boosting approach, *IEEE Geoscience and Remote Sensing Letters*, 6 (1), 127–131.
- McGonigle, C., Brown, C., Quinn, R. and Grabowski, J. (2009) Evaluation of image-based multibeam sonar backscatter classification for benthic habitat discrimination and mapping at Stanton Banks, UK, *Estuarine, Coastal and Shelf Science*, 81, 423-437.
- McIntyre, D. and Leftheriou, A. (2005) *Methods for the study of marine benthos*, 3 ed., Blackwell Science Ltd.

- McIntyre, N. and Wiens, J. (2000) A novel use of the lacunarity index to discern landscape function, *Landscape Ecology*, 15 (4), 313-321.
- Melgani, F. and Bruzzone, L. (2004) Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42 (8), 1778-1790.
- MESH (2009) Mapping European Seabed Habitats: Development of a Framework for Mapping European Seabed Habitats, 2004–2008, www.searchmesh.net. [Accessed February 2012].
- Micallef, A., Berndt, C., Masson, D. and Stow, D. (2007) A technique for the morphological characterization of submarine landscapes as exemplified by debris flows of the Storegga Slide, *Journal of Geophysical Research*, 112, (F2).
- Mignotte, M. and Collet, C. (2000) Markov Random Field and Fuzzy Logic Modeling in Sonar Imagery: Application to the Classification of Underwater Floor, *Computer Vision and Image Understanding*, 79, 4–24, doi:10.1006/cviu.2000.0844.
- Mitchell, H. (2007) *Multi-sensor data fusion: an introduction*, Springer.
- Mitchell, N (1996) Processing and analysis of Simrad multibeam sonar data, *Marine Geophysical Researches*, 18 (6), 729-739, doi:10.1007/BF00313883.
- Movellan, R. (2002). Tutorial on Gabor filters. Open Source Document.
- Müller, R, Eagles, S, Hogarth, P and Hughes, M (2007) Automated textural image analysis of seabed backscatter mosaics: A comparison of four methodologies, http://www.geosci.usyd.edu.au/users/dietmar/Pdf/Muller_etal_Geohab_2007.pdf.

N

Naudts, L., Greinert, J., Artemov, Y., Beaubien, S., Borowski, C. and Batist, M. (2008) Anomalous sea-floor backscatter patterns in methane venting areas, Dnepr paleo-delta, NW Black Sea, *Marine Geology*, 251 (3-4), 253–267.

Nelson, H., Thor, D., Sandstrom, M. and Kvenvolden, K. (1979) Modern biogenic gas generated craters (sea-floor pockmarks) on the Bering Shelf, Alaska, *Bulletin of the Geological Society of America*, 90 (12).

Novarini, J. and Caruther, J. (1998) A simplified approach to backscattering from a rough seafloor with sediment inhomogeneities, *IEEE Journal of Oceanic Engineering*, 23 (3), 157-166, doi:10.1109/48.701188.

O

Ojala, T., Pietikäinen, M. and Mäenpää, T. (2002) Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 971–987.

Oza, N. and Tumer, K. (2008) Classifier ensembles: Select real-world applications, *Information Fusion*, 9 (1), 4-20.

P

Pal, M. (2007) Ensemble learning with decision tree for remote sensing classification, *Proceedings of World Academy of Science, Engineering and Technology*, 26, 735-737.

Panish, R. and Taylor, M. (2011) Achieving high navigation accuracy using inertial navigation systems in autonomous underwater vehicles, *proceedings of IEEE Oceans, Spain 2011*, 1-7, doi:10.1109/Oceans-Spain.2011.6003517.

- Pardoe, D., Ryoo, M. and Miikkulainen, R. (2005) Evolving neural network ensembles for control problems, Proceedings of the conference on Genetic and evolutionary computation, ACM, 1384.
- Parnum, I. (2008) Benthic habitat mapping using multibeam sonar systems. PhD Thesis, Curtin University of Technology: Australia.
- Parnum I., Siwabessy, P. and Gavrilov, A. (2004) Identification of seafloor habitats in coastal shelf waters using a multibeam echosounder, Proceedings of ACOUSTICS 2004, <http://www.cmst.curtin.edu.au/publicat/2004-41.pdf>.
- Patrinou, P., Alexandridis, A., Ninos, K. and Sarimveis, H. (2010) Variable selection in nonlinear modeling based on RBF networks and evolutionary computation, International journal of neural systems, 20 (05), 365-379.
- Penney, G., Weese, J., Little, J., Desmedt, P. and Hill, D. (1988) A comparison of similarity measures for use in 2D-3D medical image registration, IEEE Transactions on Medical Imaging, 17(4), 586–595.
- Penrose, J., Siwabessy, P., Gavrilov, A., Parnum, I., Hamilton, L., Bickers, A., Brooke, B., Ryan, D. and Kennedy, P. (2005) Acoustic Techniques for Seabed Classification, Cooperative Research Centre for Coastal Zone Estuary and Waterway Management Technical Report 32.
- Petre, M. and Rugg, G. (2010) The Unwritten Rules of PhD Research, Open University Press.
- Petrou, M. and Sevilla, P. (2006) Image Processing: Dealing with Texture, John Wiley and Sons Inc.
- Philip, D. (2003) An evaluation of USBL and SBL acoustic systems and the optimisation of methods of calibration, Part 1, The Hydrographic Journal, 108, 18-25.
- Phillips, E. and Pugh, D. (2010). How to get a PhD. Open University Press.

- Pican, N., Trucco, E., Ross, M., Lane, D., Petillot, Y. and Tena Ruiz, I. (1998) Texture analysis for seabed classification: co-occurrence matrices vs. self-organizing maps, proceedings of IEEE Oceans, 1, 424-428, doi:10.1109/OCEANS.1998.725781.
- Pinet, N., Duchesne, M., Lavoie, D., Bolduc, A. and Long, B. (2008) Surface and subsurface signatures of gas seepage in the St. Lawrence Estuary (Canada): Significance to hydrocarbon exploration, *Marine and Petroleum Geology*, 25 (3), 271-288.
- Plotnick, R., Gardner, R. and O'Neill, R. (1993) Lacunarity indices as measures of landscape texture, *Landscape Ecology*, 8 (3), 201-211.
- Pollen, D. and Ronner, S. (1983) Visual cortical neurons as localized spatial frequency filters, *IEEE Transactions on Systems, Man, and Cybernetics*, 13 (5), 907-916.
- Preston, J., Christney, A., Collins, W. and Bloomer, S. (2004) Automated acoustic classification of sidescan images, proceedings of IEEE Oceans, 4, 2060-2065.
- Preston, J. (2009) Automated acoustic seabed classification of multibeam images of Stanton Banks, *Applied Acoustics*, 70 (10), 1277-1287, doi:10.1016/j.apacoust.2008.07.011.
- Puzicha, J., Hofmann, T. and Buhmann, J. (1997) Nonparametric similarity measures for unsupervised texture segmentation and image retrieval, proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 267-272.
- R**
- Ramesh, V., Glass, R. and Vessey, I. (2004) Research in computer science: an empirical study, *The Journal of systems and Software*, 70, 165-176.
- Rao, A. and Lohse, G. (1993) Identifying high level features of texture perception, *CVGIP: Graphical Models and Image Processing*, 55 (3), 218-233.
- Reed, T. (1987) Digital image processing and analysis techniques for SeaMARC II sidescan sonar imagery, Ph.D dissertation, University of Hawaii, Honolulu.

- Reed, T. and Hussong, D. (1989) Digital Image Processing Techniques for Enhancement and Classification of SeaMARC II Side Scan Sonar Imagery, *Journal of Geophysical research*, 94 (B6), 7469-7490, doi:10.1029/JB094iB06p07469.
- Rish, I. (2001) An empirical study of the naïve Bayes classifier, *IJCAI workshop on empirical methods in artificial intelligence*, 3 (22), 41-46.
- Roff J., Taylor, M. and Lauhgren, J. (2003) Geophysical approaches to the habitat classification, delineation and monitoring of marine habitats and their communities, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13 (1), 77–90, doi:10.1002/aqc.525.
- Rogers, J., Kelley, J., Belknap, D., Gontz, A. and Barnhardt, W. (2006) Shallow-water pockmark formation in temperate estuaries: A consideration of origins in the western gulf of Maine with special focus on Belfast Bay, *Marine Geology*, 225, (1-4), 45–62.
- Rubner, T., Puzicha, J., Tomasi, C. and Buhmann, J. (2001) Empirical evaluation of dissimilarity measures for color and texture, *Computer Vision and Image Understanding*, 84 (1), 25–43.
- Ruiz, L., Fdez-Sarría, A. and Recio, J. (2004) Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study, 20th ISPRS Congress.
- S**
- Saeyns, Y., Abeel, T. and Van de Peer, Y. (2008) Robust feature selection using ensemble feature selection techniques, *Machine Learning and Knowledge Discovery in Databases*, 313–325.
- Samiee, K. and Rad, G. (2008) Textural Segmentation of Sidescan Sonar Images Based on Gabor Filters Bank and Active Contours without Edges, *proceedings of the IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, 3–8.

- Scherer, R. (2010) Designing boosting ensemble of relational fuzzy systems, *International Journal of Neural Systems*, 20(5), 381–399.
- Schmidt, R., Lyytinen, K., Keil, M. and Cule, P. (2001) Identifying software project risks: An international Delphi study, *Journal of management information systems*, 17 (4), 5-36.
- Schumann, G., Trigg, M., Sefton-Nash, E., Bates, P., Wilsher, W., Seiderer, L. and Pearce, B (2010) Automated mapping of marine habitats from marine sonar and bathymetric data, Published by the Marine Aggregate Levy Sustainability Fund (MALSF), Crown Copyright, Courtesy of Cefas, ISBN-978 0 907545 53 8.
- Seabeam (2000) Multibeam sonar theory of operation, L-3 Communications SeaBeam Instruments, 141 Washington Street, East Walpole, MA,
<http://www.mbari.org/data/mbsystem/sonarfunction/SeaBeamMultibeamTheoryOperation.pdf>
- Shang, C. and Brown, K. (1992) Feature-based texture classification of sidescan sonar images using a neural network approach, *Electronics Letters*, 28 (23), 2165-2167.
- Shang, C. and Brown, K. (1993) *Texture Classification of Sidecan Sonar Images with Neural Networks*, the Institution of Electrical Engineers. Printed and published by the IEEE. Savoy Place, London WCPR OBL. UK.
- Sharifi, M., Fathy, M. and Mahmoudi, M. (2002) A classified and comparative study of edge detection algorithms, proceedings of the IEEE International Conference on Information Technology: Coding and Computing, 117-120.
- Simard Y. and Stepnowski A. (2007) Classification methods and criteria, ICES Cooperative Research Report No. 286, 1 – 6,
<http://www.ices.dk/pubs/crr/crr286/CRR286.pdf>.
- Sinha, A., Chen, H., Danu, D., Kirubarajan, T. and Farooq, M. (2008) Estimation and decision fusion: A survey, *Neurocomputing*, 71, 2650-2656.

Sobel, I., (1990) An Isotropic 3×3 Gradient Operator, *Machine Vision for Three Dimensional Scenes*, Freeman, H., Academic Pres, NY, 376-379.

Sonka, M., Hlavac, V. and Boyle, R. (2008) *Image processing, analysis and machine vision* (Third edition), Thomson.

Steele, B. and Patterson, D. (2001) Land cover mapping using combination and ensemble classifiers, *Proceedings of the 33rd Symposium on the Interface*. Interface Foundation of North America.

Stewart, W., Jiang, M. and Marra, M. (1994) A neural network approach to classification of sidescan sonar imagery from a midocean ridge area, *IEEE Journal of Oceanic Engineering*, 19 (2), 214-223.

Stewart, W. Marra, M. and Jiang, M. (1992) A Hierarchical Approach To Seafloor Classification Using Neural Networks, *proceedings of IEEE Oceans, Mastering the Oceans Through Technology*, 1, 109-113.

Strehl, A. and Ghosh, J. (2002) Cluster ensembles - a knowledge reuse framework for combining multiple partitions, *The Journal of Machine Learning Research*, 3, 583-617.

Sun, N. and Shim, T. (2008) *Sonar Images Classification of Seabed Physiognomy Based on the Information Fusion Methods*, 2008 Congress on Image and Signal Processing, 2.

Susman, G. and Evered, R. (1978) An assessment of the scientific merits of action research, *Administrative science quarterly*, 582-603.

T

Tan, P., Steinbach, M. and Kumar, V. (2006) *Introduction to data mining*, Pearson Addison Wesley, Boston.

Thacker, N., Aherne, F. and Rockett, P. (1997) The Bhattacharyya metric as an absolute similarity measure for frequency coded data, *Kybernetika* 34 (4), 1–7.

Thoros, E. and Jackson, D. (1989) The validity of the perturbation approximation for rough surface scattering using a Gaussian roughness spectrum, *Journal of the Acoustical society of America*, 86, 261-277.

Tichy, W. (1998) Should computer scientists experiment more? *IEEE Computer*, 31 (5), 32-40.

Torre, V. and Poggio, T. (1986) On edge detection, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 2, 147-163.

Tsang, I., Kocsor, A. and Kwok, J. (2007) Simpler core vector machines with enclosing balls, in *Proceedings of the 24th international conference on Machine learning*, ACM New York, NY, USA, 911–918.

Tsang, I., Kwok, T. and Cheung, P. (2005) Core vector machines: Fast SVM training on very large data sets, *Journal of Machine Learning Research*, 6 (1), MIT Press.

Tumer, K. and Agogino, A. (2008) Ensemble clustering with voting active clusters, *Pattern Recognition Letters*, 29 (14), 1947-1953.

Tweed, T. and Miguet, S. (2002) Automatic detection of regions of interest in mammographies based on a combined analysis of texture and histogram, *Proceedings of the 16 th International Conference on Pattern Recognition*, 2, 448-452.

U

Urbach, E. and Stepinski, T. (2009) Automatic detection of sub-km craters in high resolution planetary images, *Planetary and Space Science*, 57 (7), 880–887.

Urick, R. (1967) *Principles of underwater sound for Engineers*, McGraw-Hill Book Co.

V

Van der Sanden, J. and Hoekman, D. (2005) Review of relationships between grey-tone co-occurrence, semivariance and autocorrelation based image texture analysis approaches, *Canadian Journal of Remote Sensing*, 31 (3), 207-213.

Vapnik, V. (1982) *Estimation of Dependences Based on Empirical Data* [in Russian]. Nauka, Moscow, 1979. (English translation: Springer Verlag, New York, 1982).

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer, New York.

Vapnik, V. (1998) *Statistical Learning Theory*. John Wiley and Sons, Inc., New York.

Varma, M. and Zisserman, A. (2005) A statistical approach to texture classification from single images, *International Journal of Computer Vision*, 62 (1), 61-81.

Vasconcelos, N. and Lipmann, A. (2000) A unifying view of image similarity, *proceedings of the IEEE International Conference on Pattern Recognition*, 1, 38–41.

Vinh, N., Epps, J. and Bailey, J. (2010) Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance, *The Journal of Machine Learning Research*, 11, 2837-2854.

W

Wackernagel, H. (2003) *Multivariate Geostatistics*, Springer.

Wang, W., Richards, G. and Rea, S. (2005) Hybrid Data Mining Ensemble for Predicting Osteoporosis Risk, *IEEE-EMBS 2005, 27th Annual International Conference of the Engineering in Medicine and Biology Society*, 886-889.

Wang, Y., Z. Liu, Z., Sang, E. and Ma, H. (2007) Sonar image classification based on directional wavelet and fuzzy fractal dimension, in *2nd IEEE Conference on Industrial Electronics and Applications*, 118–120.

-
- Waske, B. and Benediktsson, J. (2007) Fusion of support vector machines for classification of multisensor data, *IEEE transactions on geoscience and remote sensing*, 45 (12), 3858.
- Webb, A. (1999) *Statistical pattern recognition*, John Wiley and Sons Ltd., Chichester, England.
- Webb, K. (2009) *Ecology and geology of pockmarks*, PhD thesis, University of Oslo.
- Wedding, L. and Friedlander, A. (2008) Determining the influence of seascape structure on coral reef fishes in Hawaii using a geospatial approach, *Marine Geodesy*, 31 (4), 246-266.
- Weiss, A. (2001) Topographic positions and landform analysis (poster), ESRI International User Conference, July 2001, San Diego, CA.
- Welch, R., Gupta, S. and Kuo, K. (1988) Marine stratocumulus cloud fields off the coast of southern California observed using LANDSAT imagery. Part II: Textural analysis. *Journal of Applied Meteorology*, 27, 363–378.
- Williams, K. (2001) An effective density fluid model for acoustic propagation in sediments derived from Biot theory, *Journal of the Acoustical Society of America*, 110 (5), 2276–2281.
- Wilson, M., O'Connell, B., Brown, C., Guinan, J. and Grehan, A. (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope, *Marine Geodesy*, 30 (1), 3-36.
- Wood, J. (1996) *The Geomorphological Characterisation of Digital Terrain Models*. Ph.D. thesis, University of Leicester.
- X**
- Xie, X. and Mirmehdi, M. (2008) A galaxy of texture features, in *Handbook of texture analysis*, Mirmehdi, M., Xie, X. and Suri, J., Eds., pp. 375–406. Imperial College Press.

Z

Zhou, L. Lai, K. and Yu, L. (2010) Least squares support vector machines ensemble models for credit scoring, *Expert Systems with Applications*, 37, 127-133.

Zhou, X. and Chen, Y. (2005) Seafloor Classification of Multibeam Sonar Data Using Neural Network Approach, *Marine Geodesy*, 28, 201 – 206, doi: 10.1080/01490410590953785.

Zhou, Z. and Tang, W. (2006) Clusterer ensemble, *Knowledge-Based Systems*, 19 (1), 77-83.

Zieger, S., Stieglitz, T. and Kininmonth, S. (2009) Mapping reef features from multibeam sonar data using multiscale morphometric analysis, *Marine Geology*, 264 (3-4), 209-217.

Zortea, M., De Martino, M. and Serpico, S. (2007) A SVM ensemble approach for spectral-contextual classification of optical high spatial resolution imagery, *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007*, 1489-1492.