

Perceptual Dimensions for Surface Texture Retrieval

Khemraj Emrith

Submitted for the degree of Doctor of Philosophy



Heriot-Watt University

School of Mathematical and Computer Sciences

May 2008

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that the copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author or of the University (as may be appropriate).

Abstract

This thesis presents a methodology for developing perceptually relevant surface texture retrieval systems. Generally such systems have been researched using image texture which has been captured under unknown or uncontrolled conditions (e.g. Brodatz). However, it is known that changes in illumination affect both the visual appearance of surfaces and the computational features extracted from their images. In contrast this thesis either uses surface information directly, or computes features obtained from images captured under controlled lighting conditions.

Psychophysical experiments were conducted in which observers were asked to place texture samples into groups. Multidimensional Scaling was applied to the resulting similarity matrices to obtain a more manageable reduced perceptual space. A four-dimensional representation was found to capture the majority of the variability. A corresponding feature space was created by linearly combining selected trace transform features. Retrieval was performed simply by determining the n closest neighbours to the query's feature vector. An average retrieval precision of 60% was obtained in blind tests.

Acknowledgments

This thesis will be incomplete without me expressing my gratitude to all those who have helped and supported me throughout my PhD years.

My thanks go mainly to Prof. Mike J. Chantler. I have been extremely fortunate to have you as my research supervisor. Your professional attitude and dedication as a supervisor is exemplary. You have always ensured that I stay focused and maintain a steady work ethic. Your ideas and clarity of thoughts have always been sources of encouragement.

I would like to thank the members of the Texture Lab for their invaluable suggestions and inputs to this thesis. I have always been made to feel at ease by you guys and was always provided by valuable advices when needed. Special thanks to Fraser Halley who went through the painstaking task of reading my thesis.

I would like to thank the Heriot-Watt University for offering me the “James Watt Scholarship” to pursue this thesis.

Finally, my biggest thanks goes to my wife Bintee without whom I wouldn’t have coped with the stressful days and nights that is part and parcel of a PhD. I thank you for your love and moral support and for always being by my side in difficult moments.

ACADEMIC REGISTRY
Research Thesis Submission



Name:	Khemraj Emrith		
School/PGI:	School of Mathematical and Computer Sciences		
Version: <i>(i.e. First, Resubmission, Final)</i>	Final	Degree Sought:	PhD in Computer Science

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

- 1) the thesis embodies the results of my own work and has been composed by myself
- 2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
- 3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
- 4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
- 5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.

Signature of Candidate:		Date:	21 st of May 2008
-------------------------	--	-------	------------------------------

Submission

Submitted By <i>(name in capitals)</i> :	KHEMRAJ EMRITH
Signature of Individual Submitting:	
Date Submitted:	21 st of May 2008

For Completion in Academic Registry

Received in the Academic Registry by <i>(name in capitals)</i> :			
Method of Submission <i>(Handed in to Academic Registry; posted through internal/external mail):</i>			
E-thesis Submitted			
Signature:		Date:	

Contents

Abstract	ii
Acknowledgments	iii
Research Thesis Submission	iv
Contents	v
List of Figures.....	ix
List of Tables	xiii
Table of Terms	xiv
Table of Notations	xv
Chapter 1 Introduction	1
1.1 Motivation and driving issues.....	1
1.2 Goals	3
1.3 Scope of this thesis.....	4
1.4 Novelties and Contributions	5
1.5 Thesis organisation	6
Chapter 2 Texture Retrieval: Challenges, Approaches and Techniques	7
2.1 Introduction.....	7
2.2 Challenges in Texture Retrieval.....	8
2.2.1 Challenges from the Computer Vision community	9
2.2.2 Challenges from a Vision Science perspective	12
2.3 Brief overview of Computational Approaches	14
2.4 Perceptual Approaches to Texture Retrieval	18
2.5 Tools and Techniques for Perceptual Texture Retrieval.....	21
2.5.1 Techniques to capture human judgments	21
2.5.2 Techniques for analysing psychophysical data	24
2.5.3 Mapping computational features to a perceptual space	27
2.6 Assessing requirements to develop a Perceptual Texture Retrieval System	29
2.6.1 Reliability of Rao and Lohse’s psychophysical results.....	29
2.6.2 Dataset for a new psychophysical experiment	32
2.6.3 Selecting tools for the capture and analysis of psychophysical data.....	34
2.7 Conclusion	36

Chapter 3	Design and Implement the Psychophysical Experiment.....	37
3.1	Introduction.....	37
3.2	Basic Approach.....	38
3.3	Specification of the Stimuli	39
3.3.1	Dataset consisting mainly of believable surface textures.....	40
3.3.2	Granularity and Roughness	40
3.3.3	Controlled Illumination and viewpoint conditions	41
3.3.4	Variety of texture samples.....	41
3.3.5	Matte surfaces	41
3.3.6	Constant albedo surfaces	42
3.3.7	Size of the datasets	42
3.3.8	Size and resolution of images.....	42
3.4	Acquisition of Stimuli.....	43
3.4.1	Sourcing the Datasets	44
3.4.2	Tex1 – Capturing height-maps of real surfaces	45
3.4.3	Tex1 – Synthetic texture generation.....	47
3.4.4	Preparation of samples for experiment.....	57
3.5	Experiment Design.....	58
3.5.1	Grouping task	58
3.6	Implementation of experiment.....	58
3.6.1	Comments collected on the experiment	59
3.7	Aggregate Data (generate similarity matrices)	61
3.8	Conclusions.....	62
Chapter 4	Psychophysical Data Analysis	63
4.1	Introduction.....	63
4.2	Data representation	64
4.3	Cluster Analysis	65
4.3.1	Dendrograms	65
4.3.2	Analysing the Tex1 Dataset	67
4.3.3	Analysing the MoMA Dataset	72
4.4	Dimensional Analysis	75
4.4.1	Alignment error	75
4.4.2	Stress	77
4.5	Traits in major dimensions.....	79

4.5.1	The Tex1 dataset	79
4.5.2	The MoMA dataset.....	82
4.6	Conclusions.....	85
Chapter 5	Identifying features for texture retrieval	86
5.1	Introduction.....	86
5.2	Feature Selection Criteria	87
5.2.1	Phase sensitive features	87
5.2.2	Power Spectrum sensitive features.....	88
5.2.3	Position independent features	88
5.2.4	Generating large pool of features	88
5.2.5	Avoiding Redundant Features or Feature Sets	89
5.2.6	Features that are inexpensive and simple to compute	89
5.3	Investigating Feature Extraction Methods	90
5.3.1	Local Binary Patterns	90
5.3.2	Gabor features (Phase and power spectrum features)	93
5.3.3	Simoncelli's features	95
5.3.4	Trace Transform features	97
5.4	Feature Set Selection.....	101
5.4.1	Feature Normalisation	102
5.5	Conclusion	102
Chapter 6	Surface Texture Retrieval.....	103
6.1	Introduction.....	103
PART I: MFS Based Texture Retrieval.....	104	
6.2	Overview of the development and retrieval processes	105
6.3	Stage I – Determining the number of perceptual dimensions to be used for retrieval	107
6.3.1	Results of the analysis of Reduced Perceptual Space (RPS)	108
6.4	Stage II – Producing an RPS to MFS mapping using regression analysis.....	110
6.4.1	The Prediction Model.....	110
6.4.2	Feature Selection	112
6.4.3	Cross-Validation.....	113
6.4.4	Model Selection to estimating MFS for Tex1 and MoMA data sets	118
6.4.5	Determine the number of features for MFS mapping	120

6.4.6	Mapping results	121
6.5	Stages III & IV – Query Feature Vector Calculation and Texture Retrieval..	125
6.5.1	Retrieval results.....	129
6.6	Blind Testing.....	136
6.6.1	Test Sample Selection	136
6.6.2	Effect of varying the number of features (per dimension).....	137
6.6.3	Results – further details.....	139
6.7	Summary for Part I.....	141
PART II: Full Perceptual Space Based Texture Retrieval.....		143
6.8	Introduction.....	144
6.9	Overview of modified retrieval processes	144
6.10	Optimisation Model	145
6.10.1	Statement of the problem	145
6.10.2	Levenberg-Marquardt (LM) Algorithm	146
6.11	Texture retrieval using the optimised model	148
6.11.1	Retrieval rates v/s number of features.....	148
6.11.2	Blind test results using the OFS	149
6.11.3	Computation times	152
6.12	Summary for Part II	152
6.13	Performance Evaluation and Discussion	153
6.14	Conclusion	156
Chapter 7 Summary, Conclusion and Future Works.....		157
7.1	Summary of Research.....	157
7.2	Conclusion	158
7.3	Future works	160
Appendix A: Texture Datasets – Tex1 and MoMA		163
Appendix B: Grouping Experiment Instructions		169
Appendix C: Similarity Matrix.....		171
Appendix D: Grouping Results.....		173
Appendix D.1:	Tex1 Groups	174
Appendix D.2:	MoMA Groups	178
Appendix E: Trace Transform Functionals		181
Appendix F: Retrieval Results		185
References		190

List of Figures

<i>Figure 1.1 – Processes involved in retrieval of perceptually similar surface textures.....</i>	<i>4</i>
<i>Figure 2.1– Effect of changing illumination direction on appearance of surface.....</i>	<i>30</i>
<i>Figure 2.2 – Sinusoidal behaviour of texture features when applied to four different surfaces that have been illuminated under varying illumination conditions</i>	<i>30</i>
<i>Figure 3.1 - Steps required for psychophysical experiment setup</i>	<i>38</i>
<i>Figure 3.2– Lambertian reflection (left) and specular reflection (right) on somewhat matte and mirror-like surfaces respectively.....</i>	<i>41</i>
<i>Figure 3.3 –Capture geometry</i>	<i>45</i>
<i>Figure 3.4 – Photometric images for the anaglypta surface taken at a fixed slant of 45° and tilt angles 0°, 90° and 180°</i>	<i>46</i>
<i>Figure 3.5 - Row 1 shows the height maps of some real world surfaces and row 2 shows their corresponding images.....</i>	<i>47</i>
<i>Figure 3.6 – (a) Mulvaney surface (b) Ogilvy surface</i>	<i>49</i>
<i>Figure 3.7– Row 1 shows different distribution of frequencies and row 2 shows resultant patterns</i>	<i>50</i>
<i>Figure 3.8 – Phase rich textures</i>	<i>51</i>
<i>Figure 3.9– (a) 1D signal folding (b) corresponding folding within a 2-D plane representing the same signal.....</i>	<i>52</i>
<i>Figure 3.10 – Dark regions identify the placements on surfaces with varied high frequency information</i>	<i>53</i>
<i>Figure 3.11 – Texture primitives.....</i>	<i>53</i>
<i>Figure 3.12 – Stages (I) to (III) illustrate the complete process of applying folding, thresholding and primitives mapping to generate a synthetic surface. Stage (IV) represents the illuminated surface.</i>	<i>54</i>
<i>Figure 3.13 - (column1) natural surfaces recovered using photometric stereo, (column2) visually similar synthetic surfaces.</i>	<i>55</i>
<i>Figure 3.14 -(column1) natural texture images and (column2) visually similar synthetic texture images</i>	<i>56</i>
<i>Figure 3.15 - Tex1 sample photograph provided to observers for psychophysical experiment</i>	<i>57</i>
<i>Figure 3.16 - Occurrence Matrix.....</i>	<i>61</i>
<i>Figure 4.1- Ideal dendrogram.....</i>	<i>66</i>

<i>Figure 4.2 - Complete dendrogram with leaves being texture samples used in the psychophysical experiment. Dissimilarity level units correspond to number of observers.</i>	68
<i>Figure 4.3- Dendrogram showing the dissimilarity level when 6 ‘major’ clusters are considered</i>	68
<i>Figure 4.4 - Groups 1 to 6 from dendrogram represented by columns 1-6</i>	71
<i>Figure 4.5 – Dendrogram representing linkage between MoMA textures to form groups</i>	72
<i>Figure 4.6– Dendrogram partition showing 2 different partitioning</i>	73
<i>Figure 4.7– Groups 1 to 6 for MoMA database</i>	74
<i>Figure 4.8– Graph showing how alignment error decreases as dimensionality increases</i>	76
<i>Figure 4.9 - Scree plot showing “elbow” effect</i>	78
<i>Figure 4.10- Normalised stress for the first 20 dimensionalities used to represent the MoMA and Tex1 datasets</i>	79
<i>Figure 4.11 – Spatial representation of dataset after 120*120 Similarity Matrix is reduced to a 2D space using MDS</i>	80
<i>Figure 4.12 – Span of different categories of textures when the first two dimensions are considered</i>	81
<i>Figure 4.13– Spatial arrangement of MoMA textures in a 2D plane</i>	82
<i>Figure 4.14-Variation of different texture categories along trend line TL1</i>	83
<i>Figure 4.15- Variation of different texture categories along trend line TL2</i>	84
<i>Figure 5.1 - Magnitude only, (b), and phase only, (c), representations of a checkerboard, (a)</i>	87
<i>Figure 5.2 - Original LBP operator with associated weights</i>	90
<i>Figure 5.3 - Circular symmetric neighbour sets used generated at three different scales</i>	91
<i>Figure 5.4 – Variation of histogram values for when operator $LBP_{8,1}$ is applied to texture T89</i>	92
<i>Figure 5.5 - (top row) spatial representation of Gabor wavelet pair, and (bottom row), corresponding frequency domain representation</i>	93
<i>Figure 5.6 - Dyadic Gabor filter bank with 4 orientations and 3 frequencies</i>	94
<i>Figure 5.7 - Steerable filters at 2 different orientations and scales</i>	95
<i>Figure 5.8- (a) Parameters associated with a tracing line, (b) Converting 2D surface to a 1D function</i>	97

<i>Figure 5.9 - Triple feature construction from original texture (a), to2D function (b), and transformed to a 1D function (c). The final result is a scalar obtained from (c)</i>	<i>99</i>
<i>Figure 6.1–The four main stages of the proposed MDS based Retrieval Model.....</i>	<i>105</i>
<i>Figure 6.2- Graph of precision versus dimensionality, (a) precision for Tex1 (b) for MoMA dataset</i>	<i>109</i>
<i>Figure 6.3 – Diagram showing selection of test (shaded cell) and training data for 10-fold CV</i>	<i>114</i>
<i>Figure 6.4- Average error for LOOCV and 10-Fold CV methods when training the Tex1 dataset (top) and MoMA dataset (bottom).</i>	<i>118</i>
<i>Figure 6.5- Variation of precision with increasing number of features used to approximate a 4D RPS for (a)the Tex1 dataset and (b)the MoMA dataset.....</i>	<i>120</i>
<i>Figure 6.6 – RPS (vertical axes) vs. MFS (horizontal axes) values for Tex1 texture samples (rows correspond to the first four dimensions, columns 1,2,&3 correspond to increasing number of features - 1, 25 and 75 respectively.</i>	<i>123</i>
<i>Figure 6.7 – RPS (vertical axes) vs. MFS (horizontal axes) values for MoMA texture samples (rows correspond to the first four dimensions, columns 1,2,&3 correspond to increasing number of features - 1, 25 and 65 respectively.</i>	<i>124</i>
<i>Figure 6.8 – 10-bins precision histograms for 10, 20 and 30 retrieved samples for Tex1 textures.</i>	<i>126</i>
<i>Figure 6.9 – 10-bins precision histograms for 10, 20 and 30 retrieved samples for MoMA textures.</i>	<i>127</i>
<i>Figure 6.10- First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture T55(lower quartile) as query</i>	<i>132</i>
<i>Figure 6.11-First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture T34(median) as query</i>	<i>133</i>
<i>Figure 6.12- First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture T116(upper quartile) as query</i>	<i>134</i>
<i>Figure 6.13- First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture M1954(upper quartile) as query.....</i>	<i>135</i>
<i>Figure 6.14 – (a) Tex1 and (b) MoMA test textures randomly selected from 10 and 8groups obtained by applying the cluster analysis to the two datasets</i>	<i>137</i>
<i>Figure 6.15 – Variation of precision with increasing number of features for (top) Tex1 test textures and (bottom) MoMA textures (blind test).....</i>	<i>138</i>
<i>Figure 6.16 – Test texture T62</i>	<i>139</i>

<i>Figure 6.17- First 20 retrievals from (a) a 110D FPS and (b) the MFS feature space using texture T62 as query</i>	140
<i>Figure 6.18 - The three main stages of the proposed Optimisation model. Shaded part show the corresponding stages from the model proposed in Section 6.2</i>	144
<i>Figure 6.19 – Average precision rates for the Tex1(top) and MoMA(bottom) datasets with increasing number of features used to map the FPS to the OFS</i>	148
<i>Figure 6.20 - Texture T83</i>	150
<i>Figure 6.21- first 20 retrievals from (a) a 110D FPS and (b) the OFS feature space using texture T83 as query</i>	151
<i>Figure 6.22- Processing time v/s number of features selected to create the OFS</i>	152
<i>Figure 6.23- Effect of increasing texture features on retrieval success</i>	154
<i>Figure A.1- Texture images for Tex1 dataset, Part I (T1 to T48)</i>	164
<i>Figure A.2- Texture images for Tex1 dataset, Part II (T49 to T96)</i>	165
<i>Figure A.3- Texture images for Tex1 dataset, Part III (T97 to T120)</i>	166
<i>Figure A.4- Texture images for MoMA dataset, Part I (M12 to M1765)</i>	167
<i>Figure A.5- Texture images for MoMA dataset, Part II (M1792 to M2607)</i>	168
<i>Figure B.1– Instruction sheet presented to subjects participating in the psychophysical experiment</i>	170
<i>Figure C.1 Partial Similarity matrix showing pairwise occurrence of Tex1 images R1 to R40 (generated from surfaces T1 to T40). Matrix constructed from data coming from 8 subjects</i>	172
<i>Figure D.1- Tex1: Group1 (regular textures), group size =33</i>	174
<i>Figure D.2- Tex1: Group2 (irregular textures), group size=18</i>	175
<i>Figure D.3- Tex1: Group3 (patchy textures), group size=10</i>	175
<i>Figure D.4- Tex1: Group4 (vertical textures), group size=25</i>	176
<i>Figure D.5- Tex1: Group5 (Circular textures), group size =15</i>	177
<i>Figure D.6- Tex1: Group6 (Horizontal textures), group size= 19</i>	177
<i>Figure D.7- MoMA: Group1, group size =13</i>	178
<i>Figure D.8- MoMA: Group2, group size =17</i>	178
<i>Figure D.9- MoMA: Group3, group size =13</i>	179
<i>Figure D.10- MoMA: Group4, group size =12</i>	179
<i>Figure D.11- MoMA: Group5, group size =14</i>	180
<i>Figure D.12- MoMA: Group6, group size =12</i>	180

List of Tables

<i>Table 3.1– Summary of criteria for dataset selection</i>	43
<i>Table 4.1- Linkage functions and associated parameter values</i>	67
<i>Table 4.2– Stress values with corresponding goodness of fit interpretation</i>	77
<i>Table 5.1 – Eligibility of selected features with respect to chosen criteria. ✓ means eligible and ✗ means ineligible.</i>	101
<i>Table 6.1 – quantitative analysis of the precision value distributions for the Tex1 dataset</i>	128
<i>Table 6.2 – quantitative analysis of the precision value distributions for the MoMA dataset</i>	128
<i>Table 6.3 – Average precision for Tex1 groups.....</i>	129
<i>Table 6.4 – Comparative figures for average precision rates of Tex1 test textures in the 3 different perceptual spaces.</i>	139
<i>Table 6.5 - Comparative results for average precision rates of Tex1 test textures for retrievals in MFS(B) and OFS(B).....</i>	149
<i>Table 6.6 – Precision values for different retrieval methods obtained when applied to (a) the Tex1 dataset and (b) the MoMA dataset.....</i>	155
<i>Table 7.1– Average performance (blind testing) for Tex1 and MoMA datasets for retrievals in the MFS and OFS. The performance by chance for 110 and 73 target textures from Tex1 and MoMA is provided for comparison.</i>	159
<i>Table E.1 – Trace functionals, T. N represents the number of points along trace and x_i is the i^{th} sample.</i>	183
<i>Table E.2 – Diametric functionals, P.....</i>	183
<i>Table E.3 – Circus functionals, Φ.....</i>	184
<i>Table F.1- Retrieval results for query textures within a 4-d MFS(75 features)for the Tex1 dataset</i>	186
<i>Table F.2 - Retrieval of the 30 most similar textures to query textures (1st cell-bold) from the FPS of the Tex1 Dataset</i>	187
<i>Table F.3 - Retrieval results for query textures within a 4-d MFS(65 features)</i>	188
<i>Table F.4 - Retrieval of the 30 most similar textures to query textures (1st cell-bold) from the FPS of the MoMA Dataset</i>	189

Table of Terms

General Terms

Terminologies	Acronyms
Bidirectional Reflectance Distribution Function	BRDF
Bidirectional Texture Function	BTF
Classification and Regression Trees	CART
Content Based Image Retrieval	CBIR
Cross Validation	CV
Direct Magnitude Estimation	DME
Filter-Rectify-Filter	FRF
Hierarchical Cluster Analysis	HCA
Leave One Out Cross Validation	LOOCV
Levenberg Marquardt	LM
Multidimensional Scaling	MDS
Museum Of Modern Art	MoMA
Perceptual Texture Space	PTS
Principal Component Analysis	PCA
Trace Transform	TT

Thesis Related terms

Terminologies	Acronyms	Short definition	Page
Full Perceptual Space	FPS	Denotes the $N \times N$ Perceptual space representing the similarity Matrix	104
Reduced Perceptual Space	RPS	Refers to the low-dimensional space obtained after applying MDS to the FPS	104
MDS derived Feature Space	MFS	The feature space approximation of RPS	104
Optimised Feature Space	OFS	Feature space derived directly from the FPS	142

Table of Notations

Symbols	Meaning
N	Number of textures in a dataset
f	Total number of features extracted to represent each textures
d	Dimensionality of RPS and MFS
O_i	Represents observation or texture sample i
Ω	Set of N objects or textures for a specific dataset
h_{ij}	Height of the internal node specifying the smallest class to which objects O_i and O_j belong to
α and γ	Parameters associated with clustering strategies
R_d^i	Set of textures, sorted in ascending order of distance when compared to a texture i in the RPS of dimensionality d
R_P^i	Set of textures, sorted in descending order of similarities when compared to a texture i in the FPS
\mathbf{X}	Feature matrix of size $N \times f$
\mathbf{x}_i	Feature vector of size f associated with texture i
x_i^f	f^{th} feature component from feature vector \mathbf{x}_i
\mathbf{Y}	Vector of size N position of textures within a single dimension of the RPS
y_i	Coordinate of texture i in for a particular axis of MFS
ε	random noise
$\boldsymbol{\beta}$	Vector of regression coefficients of size $t+1$ for t selected features or also used as vector of optimisation parameters
$R(\boldsymbol{\beta})$	Empirical error from prediction model when computed using set of coefficients $\boldsymbol{\beta}$
ρ_{ij}	Pearson correlation coefficient between any feature i and another feature j
K	Number of partitions for Cross Validation
k	k^{th} partition
M_k	Size of k^{th} partition
λ	Dampening factor
\mathbf{H}	Hessian matrix
τ	Tilt angle
σ	Slant angle

$i(x,y)$	Reflectance intensity at point (x,y)
I	Illumination vector
L	Illumination matrix
n	Surface normal unit vector
m	Scaled surface normal vector
S	Similarity matrix
$S(i,j)$	Similarity of texture samples i and j
D	Dissimilarity matrix
d_{ij}	Dissimilarity of texture samples i and j
T^k	Trace functional k
P^j	Diametric functional j
Φ^i	Circus functional i
f^{ijk}	Scalar features generated from triple function using T^k , P^j and Φ^i

Chapter 1

Introduction

Efficient and perceptually relevant texture retrieval has been and remains a challenging and very important area of research. Due to the omnipresence of texture information within our natural environments, image features relating to different texture properties have been extensively exploited by Content Based Image Retrieval (CBIR) systems. Additionally, with a number of specialised applications such as medical diagnosis dealing with large databases of texture images, retrieval systems that provide robust search and retrieval facilities are constantly being sought.

1.1 Motivation and driving issues

Research in the field of CBIR has focused mostly in finding perceptual features to represent textures in order to bridge the semantic gap between low-level image content and high-level concepts used by humans in discriminating textures [Smeulders00]. All approaches considered so far in building retrieval models have used texture image datasets where the conditions under which the images have been captured have been completely neglected. The Brodatz dataset of digitised texture images has been used as the *de facto* source of images to train and test retrieval models for textures. However, the image capture conditions used have not been described in the literature. Thus it is not known that these surfaces were imaged under consistent lighting conditions. This is unfortunate, because changes in illumination can cause significant variations in both observers' perceptions and the values of computed texture features [Chantler94]. Thus the current work addresses the texture retrieval problem by considering either surface textures or texture images that have been generated under controlled and known illumination and viewpoint conditions.

Another issue of this research is the unavailability of a specific set of features that can be universally related to different perceptual traits of textures. Ever since work performed by Tamura *et al.* [Tamura78] in finding texture features relating to human perception, numerous attempts have been made in finding structural information that could represent different categories of textures. Although the texture features described in the literature have proved to be successful in some areas of texture processing,

namely texture segmentation and synthesis; content-based retrieval on the other hand has failed to meet user expectations [Rui99, Smeulders00 & Datta08]. Despite being equipped with more powerful feature extraction techniques and indexing mechanisms, retrieval systems have constantly been outsmarted by the efficiency and precision with which the Human Visual System discriminates between different categories of textures. Both frequency domain and spatial domain features have been investigated, however, the challenge provided by the semantic gap still remains. It has recently been argued by Petrou *et al.* [Petrou07] that the failure of having specific features to represent textures can be related to the fact that preconceived perceptual attributes are kept in mind while designing and extracting the feature sets. This largely biases and restricts the way the textures can be represented, especially in the case of retrieval. Thus the current research does not assume any perceptual attributes of textures within the feature extraction phase, and addresses the problem of texture representation by using a large set of features.

In order to develop retrieval systems that are consistent with the way humans perceive textures, recent research has investigated how texture features can be mapped to perceptual dimensions such that the latter could be exploited for retrieval. Following the excellent work undertaken by Rao *et al* [Rao93b] in investigating perceptual dimensions to create a taxonomy for textures, other researchers such as Long *et al.* [Long01] and Payne *et al.* [Payne05] have utilised results from psychophysical experiments on texture similarity so as to construct perceptually robust retrieval systems. However, all this research has been carried using texture imagery captured under unknown or varying conditions.

As in the case of Rao's experiments, the current research uses psychophysical data to identify perceptual dimensions after which computational features are mapped to those dimensions leading to so called perceptual texture spaces. However, these data are obtained using images of textures that have been obtained under controlled conditions.

1.2 Goals

The focus of this thesis is to research technologies and methodologies for perceptually relevant retrieval of surface textures. By perceptually relevant we imply retrievals that users would perceive as similar to a given query texture. To achieve this primary objective, it is important to investigate how humans perceive different categories of textures and use the captured judgments in training a retrieval model that can “mimic” human perception. Thus the main goals of the research are:

1. To select or create a database of surface height maps so that controlled illumination and viewpoint conditions can be applied to generate texture images for psychophysical studies.
2. To capture similarity judgments of different texture pairs taken from a set of texture images illuminated under controlled conditions.
3. To derive methods for developing appropriate retrieval systems.

1.3 Scope of this thesis

The research focuses in developing an automated retrieval model that takes as input an “unknown” texture query and uses a selected set of relevant features to retrieve perceptually similar textures. A broad overview of the processes involved in meeting the goals set-out in Section 1.2 is given in Figure 1.1.

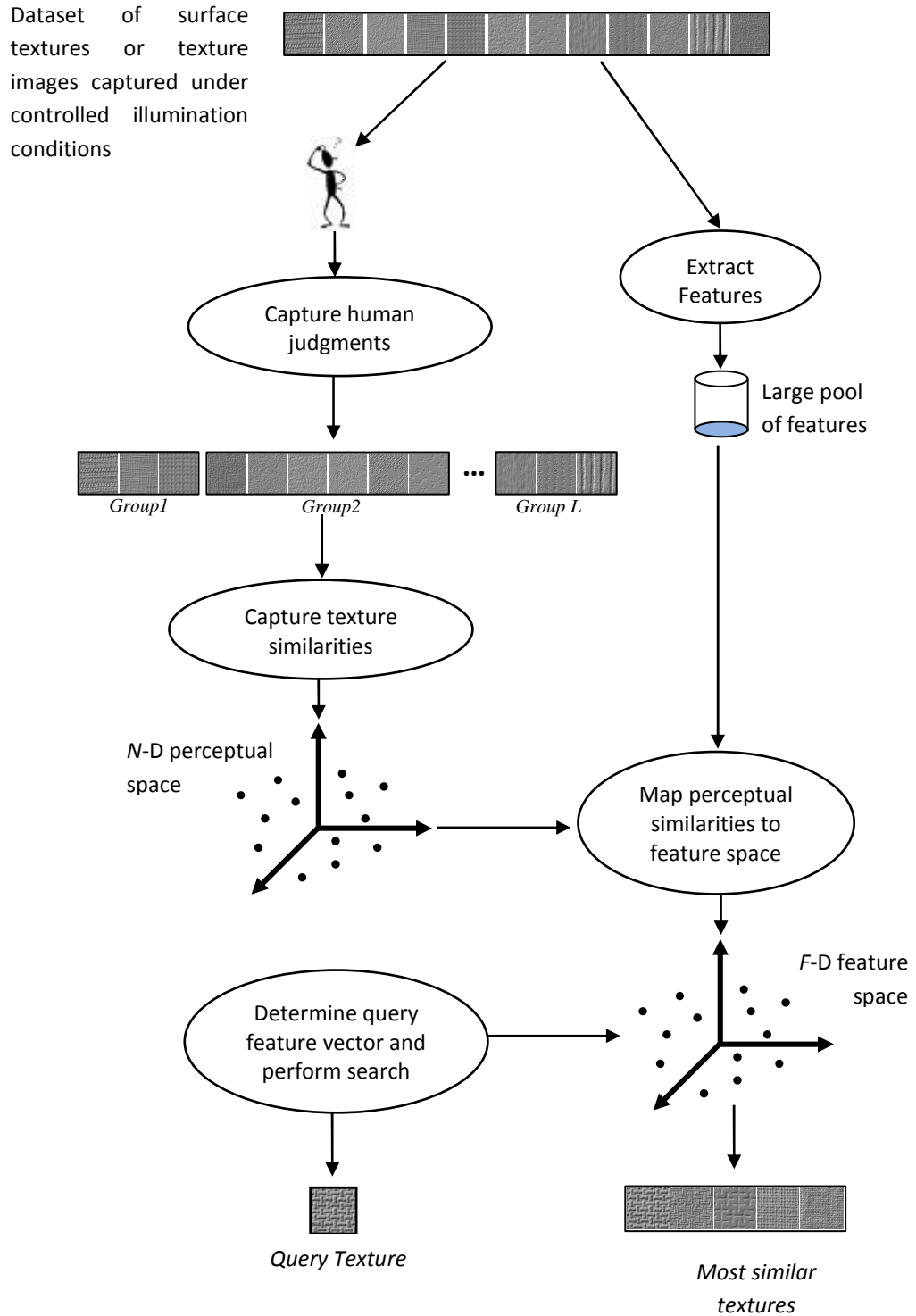


Figure 1.1 – Processes involved in retrieval of perceptually similar surface textures

The current work deals mainly with surface textures captured using a stills camera and point light sources. Lack of sufficiently homogeneous natural textures means that synthetic textures have also been generated to provide for a texture dataset with a good spread of different categories. A dataset of specialised texture images captured in a controlled environment was also made available and has been used to test the retrieval models proposed.

The way in which humans judge similar textures is explored and the results of the different observers are aggregated to form an N dimensional perceptual space, where N represents the number of textures. Human judgments are captured through properly designed and implemented psychophysical experiments.

Within the scope of this thesis we employ dimensionality reduction to reduce the high-dimensional perceptual space to a more manageable F -D feature space that is exploited for automatic retrieval of textures.

Relevant features used in mapping the perceptual space are chosen from a large pool of features. Retrieval of textures similar to a query texture results from (1) mapping the query feature into the F -D feature space and (2) locating the n nearest textures within that space.

1.4 Novelties and Contributions

The main novelty of this thesis lies in the use of surface textures, rather than images, in developing a perceptually consistent texture retrieval system. To the author's knowledge no work has attempted to characterise surface textures both in terms of computational features and in terms of human perception.

1.5 Thesis organisation

This thesis is structured in the following way.

Chapter 2 investigates how the computer vision and vision science community have been dealing with the problem of finding perceptually relevant features for texture retrieval. Based on the investigation provided, a list of tools and techniques used in perceptual retrieval of texture is identified.

Chapter 3 covers the steps required in performing a psychophysical experiment for texture similarity. The chapter provides design issues in creating the datasets to be used in the experiment and also implementation procedures through which human observations are captured. The outcome of this chapter is high dimensional perceptual data representing similarity information between texture pairs.

Chapter 4 provides an analysis of the high dimensional data obtained from the psychophysical experiments. The first part of the chapter investigates whether structural information exists within the texture groups created by observers. Cluster analysis is applied to high dimensional perceptual data to create a random number of groups and visual inspection allows us to investigate for consistency within the groups.

The second part of the chapter demonstrates how dimensionality reduction is applied to reduce the full perceptual space to lower dimensional space. Consequently the reduced perceptual space is examined to identify major perceptual texture attributes that could be exploited for retrieval purposes.

Chapter 5 investigates some popular texture description approaches in order to select a potential feature set to be used in mapping the reduced perceptual space to a corresponding feature space.

Part I of **Chapter 6** describes how the actual mapping of the feature space to the Reduced Perceptual Spaces is performed. Retrievals from the corresponding feature space are presented and analysed.

An alternative approach using the full perceptual space is presented in Part II of this chapter. The objective is to see whether a more direct approach would provide for better retrieval performance. The results from the two approaches are evaluated and discussed.

Chapter 7 summarises the work undertaken within the context of this thesis and relates how the objectives set out in Chapter 1 have been met. The results from the previous chapters are discussed and the contributions outlined.

Chapter 2

Texture Retrieval: Challenges, Approaches and Techniques

2.1 Introduction

Search, retrieval or navigation of large databases of multimedia information have for long been very active areas of research. However, with major advances in the fields of data capture and data storage, the amount of research work undertaken in the field of Content Based Image Retrieval (CBIR) has moved leaps and bounds within the last five years. CBIR is a vast and wide area of research and discussing its progress is beyond the scope of this thesis. However it is worth mentioning, at this stage, some excellent surveys made by Rui *et al.* [Rui99], Smeulders *et al.* [Smeulders00], and others such as Dai *et al.* [Dai04] and Liu *et al.* [Liu07] who have helped us to be up-to-date with the evolution of CBIR research.

The main objective of this thesis is to come up with a surface texture retrieval system that can represent human judgements of textures as closely as possible. In order to achieve this objective, the goals set out for Chapter two are as follows:

1. To investigate the challenges that researchers have faced and are facing to provide texture retrieval systems that generate perceptually consistent results,
2. To investigate how psychophysics has influenced the field of content-based retrieval and determine whether a new psychophysical experiment needs to be performed to capture how humans categorise textures, and,
3. To determine the requirements, in terms of tools and techniques, to build a perceptual texture retrieval system.

Chapter two is organised in the following way: Section 2.2 addresses the challenges of developing retrieval systems for texture. It summarises the challenges from two different perspectives: (1) from the computer vision aspect and (2) from the vision science community point of view. Section 2.3 provides a brief summary of the computational approaches to texture retrieval, whereas Section 2.4 looks into the perceptual approaches. Tools and techniques used so far by computer vision researchers and cognitive scientists in capturing, analysing and integrating human judgments within a retrieval framework are presented in Section 2.5. Section 2.6

identifies the requirements in building a surface texture retrieval system using psychophysical data. Finally, Section 2.7 summarises how the goals for this chapter were met.

2.2 Challenges in Texture Retrieval

Even if no universally accepted definition exists for texture, it has always been considered to be a very important aspect of visual information that humans constantly use to analyse different scenes. Given the abundant presence of textured surfaces in natural environments, humans generally use knowledge about those surfaces to discriminate between scenes coming from their environments. As reported by Gurnsey *et al.* [Gurnsey01], studies for texture properties have either been motivated for an ecological cause or from a signal processing perspective. Ecologically because of the omnipresence of texture information in the real world and from a signal processing perspective in order to examine how the human visual system encodes texture information.

Texture information was initially exploited within a very narrow area of machine vision, mainly in the early phases of remote sensing for radar or satellite image interpretation [Haralick73]. Since then texture-based research has rapidly and widely spread to areas of computer vision, image processing and computer graphics, be it for the analysis of texture information for image classification, the extraction of texture features for segmentation purposes, or simply for the use of texture data for visualisation. Analysis of texture information for synthesis purposes has also been used by image compression applications. Within that span of time, different texture models have been proposed to suit different applications. In their survey, Tuceryan and Jain [Tuceryan98] classified these models as statistical approaches, geometrical or structural approaches, model-based approaches and signal processing approaches. The survey published by Smeulders *et al.* [Smeulders00] provides more insight into how much work has been done till the year 2000 within the field of texture information representation and processing. However, the area is growing at such a frightening pace that seven years later the amount of work done on texture based research has almost doubled compared to research done pre year 2000.

This increase can be explained by the fact that textures, due to their aesthetical properties, are nowadays very much involved in consumer-oriented design, marketing and selling of different products. Moreover extensive application of texture in medical

diagnosis and industrial inspection has put a lot of emphasis on content-based retrieval of texture images. With cognitive scientists realising that texture, as a visual cue, plays a significant role in a variety of cognitive tasks, coupled with the fact that current texture retrieval systems are still perceptually inconsistent, strong interests have arisen from the vision science community in the field of texture retrieval. The vision science community is mostly interested in how the human visual system discriminates among different texture categories compared to the computer vision community which is constantly trying come up with a computational model that allows the representation of textures through a relevant feature space. In the rest of this section are presented the challenges faced by both the computer vision community and the vision science researchers in developing perceptually consistent texture retrieval systems.

2.2.1 Challenges from the Computer Vision community

In computational vision, we try to model and implement the vision processes at a conscious level rather than a subconscious level as in the case of human vision. Early interests in texture from the computer vision community relate mostly to the derivation of computational measures in analysing and synthesizing textures. Research in the field of CBIR imposed more challenges to the texture-oriented researchers. Those challenges are discussed below:

- *More features to represent larger categories of texture images*

The failure by CBIR systems to meet users' expectations has led researchers to believe that the main cause of this failure is the insufficiency or incompleteness of the feature set available to represent the textures (applicable to image retrieval in general). Throughout the years, various representations of texture information have been proposed, namely power spectral features, Gabor features, wavelets, moments, fractals, higher order statistics and so on. Even if these representations have performed sufficiently well in certain texture processing areas such as segmentation and synthesis, their application to retrieval has been largely unsuccessful. Most of the feature sets presented in literature have been generated with regard to specific perceptual attributes of textures such as directionality, contrast, regularity and others. However, no universal set of features has been identified so far that are used by humans to distinguish between different categories of textures. In a recent attempt to do so, Petrou *et al.* [Petrou07] have used the Trace Transform to generate very large sets of features to represent

textures. They argue that since the perceptual attributes used by humans to categorise textures are still an open problem, extracting features based on specific perceptual attributes biases and thus limits the texture representations.

- *Dealing with high dimensionality data*

With more and more features used to represent textures, the obstacle that researchers undoubtedly had to face is the curse of dimensionality. Processing thousands of features drastically decreases the performance of retrieval systems. Thus, researchers have been working on different ways to reduce a large set of features to a smaller subset that can more accurately and efficiently represent the dataset being investigated or searched. Feature selection targets mainly the problem of high dimensionality; however, it also allows the identification and removal of irrelevant and redundant features that result in more accurate learning models. A good review of the evolution of feature selection is provided by Liu *et al.* [Liu05].

- *Find ways to compensate for varying data capture conditions*

Existing works on texture retrieval have so far used sets of texture images (mostly the Brodatz dataset) where the viewpoint condition under which those images were acquired is unknown. To solve the general viewpoint invariance problem, several translation, rotation and scale invariant feature extraction techniques have been proposed in literature. Alignment techniques and structural descriptions have also been employed to bypass the viewpoint problem. Moreover, recent studies by Chantler *et al.* [Chantler94 & Chantler05] have also demonstrated the influence of texture appearance under changing illumination. Thus, illumination invariant texture representation is also an active area of research.

- *Multidimensional indexing techniques*

Even after applying feature selection/reduction techniques, the dimensions of the feature space used to represent a dataset are quite large. Traditional data structures are proving to be inefficient in storing and indexing the current crop of features or feature sets. Sophisticated multidimensional indexing techniques are constantly being explored and utilised to meet the computational demands of high-dimensional feature sets and also to reduce the response time of retrieval systems. Moreover, due to the interactive nature of current retrieval systems, we can no longer assume that features are extracted and stored in advance. A major challenge for researchers in this area is to cater for scalable image

and feature sets. Thus, much work is being done on dynamic indexing strategies. Rui *et al.* [Rui99], Smeulders *et al.* [Smeulders00] and Datta *et al.* [Datta08] provide very good resumés on multidimensional indexing techniques as used by CBIR systems.

- *Advance query modelling facilities and interactive systems*

A common limitation of the early texture retrieval systems has been the rigid interface provided for users to formulate their query. To remedy the situation, researchers have been working on several ways to expose the premise of the retrieval system in a more intuitive and natural way. Recent researches have put a lot of emphasis on ‘interactive retrieval systems’ and techniques like relevance feedback have been explored to capture users’ needs through an iterative feedback and query refinement process [Rui98, Rui99, & Lew06].

Besides providing more facilities for human interaction with retrieval systems, researchers have also considered multimodal queries in order to seek the best description of users’ needs. Thus, novel user interfaces, querying models and result visualisation techniques are constantly being explored [Lew06].

- *Semantic information extraction and learning-based approaches*

Learning based approaches are being investigated and implemented within retrieval systems in order to dynamically modify feature sets or similarity measures used in the retrieval process. With the assumption that a unique feature set cannot represent diverse categories of texture images, learning methods allow the fine-tuning of image signatures [Rui99 & Lew06].

2.2.2 Challenges from a Vision Science perspective

Interest from psychophysicists and cognitive scientists to study the visual perception of textures is not new, however, as compared to the work done on colour, the state of understanding of perceptual texture properties used in texture discrimination and categorisation is still very poor. With the inability of computer scientists to bridge the semantic gap and to provide for similarity measurements that are perceptually consistent, there has been fast growing interest from the vision science community to meet the existing challenges in texture categorisation.

- *Understand and model low-level human vision*

The main interest by vision scientists lies in what representations and rules are utilised by the human visual system to process textures. Most the work at this level relates to identification of cognitive mechanisms in the process of texture segregation. Early pioneers such as Julesz [Julesz62, Julesz75 & Julesz81] and Beck [Beck87] made use of synthetic texture stimuli to explain the discriminability of textures. Julesz proposed the “theory of textons” to explain the preattentive discriminability of texture. However, most, if not to say all, of the earlier texture perception models were based on synthetic textures and have proved difficult to formalise for real world textures. Thus, a large chunk of research undertaken by vision scientists is still dedicated in finding models that could explain how the human visual system discriminates between different categories of textures.

- *Support findings from neurophysiology*

Researches performed by neurophysiologists have suggested that the cortical cells of the human brain have receptive fields that are sensitive to spatial frequency and orientation. Inspired by those findings recent psychophysical studies have proposed different processing mechanisms that could relate to the way the brain decomposes an input texture image [Kourtzi06]. In order to mimic the operations of the visual cortex psychophysicists have applied linear filters that are selective for spatial frequency and orientation. The common framework of the mechanisms employed consists of two layers of filtering separated by a non-linearity, with the first stage of filtering more sensitive to higher spatial frequencies [Landy04 & Johnson04]. However, those mechanisms mostly focus on mapping statistical properties of texture to the processing of the visual cortex.

- *Identify salient features that capture human attention*

Psychophysicists have tried, through different studies, to understand and thus measure perceptual similarity of texture. After initial work allowed the discovery that a limited set of visual properties are used for the pre-attentive discrimination of textures, vision scientists showed lots of consideration in identifying the salient features used by the low-level visual system in analysing textures [Malik90, Heaps99 & Iqbal99]. However, it is important to point out that most work on saliency accounts for the speed and ease with which the salient features are identified. The challenge remains of how salient features could account for dominant perceptual dimensions.

- *Identify perceptual texture dimensions*

The identification of primary colours to represent the whole colour spectrum has led vision scientists to think that there might be some basic texture properties or terms that can be used to represent all the visual properties of texture. Few have attempted to solve this puzzle. To date, the work performed by Rao and Lohse [Rao93a & Rao93b] remain the most referenced and valued research work in identifying perceptual dimensions to represent textures.

- *Establish a standardised taxonomy to represent texture categories*

Along with efforts in finding dominant perceptual dimensions for texture representation, psychophysical studies have also been performed to understand how humans classify textures into meaningful and structured hierarchical categories. Again, the work done by Rao *et al.* [Rao93b] seems to be the only noticeable research that could be accounted for.

2.3 Brief overview of Computational Approaches

Computational approaches towards texture characterisation and texture processing relate mostly to the application of mathematical models that can identify and explain the perceptual qualities of textures in images. The fact that no precise definition of texture has been accepted so far by the research community means that the models used to describe texture have targeted different aspects of texture based on its perception and its application. The properties of texture considered by computer vision researchers that have aided them to formulate descriptive approaches are:

1. Texture is an organised area phenomenon and cannot be defined at a single point,
2. Texture is described by the type, the density and also the spatial distribution of its primitives, and,
3. Texture is normally perceived at different scales and resolutions.

Studies performed by Tamura *et al.* [Tamura78] and Laws [Laws80] identified a number of perceptual properties that humans use to discriminate between different categories of textures. The properties that they investigated into were uniformity, density, coarseness, roughness, regularity, linearity, directionality, frequency, and phase. Based on these conceptual properties, different computational approaches to texture representation and retrieval have been proposed in literature. In an early review, Haralick [Haralick79] summed up those approaches in two main categories: structural and statistical/stochastic. Later surveys on texture analysis extended these two categories with a third one: spectral approaches. Another similar taxonomy provided by Tuceryan and Jain [Tuceryan98] outlined the approaches as statistical, geometrical, model-based and signal processing approaches.

Statistical approaches consist mainly of fitting mathematical functions to the spatial distribution of gray level values representing the texture images. Haralick [Haralick79] suggested the use of autocorrelation functions, spatial gray level co-occurrence probabilities and autoregressive models to compute statistical texture features. Variants on these models, proposed later in literature, provided either better texture representation or used less memory and computational speed to generate the same results. The sum and difference histogram methods proposed by Unser [Unser86] were similar to Haralick's co-occurrence matrices, however, they used memory and processing power in a more efficient manner. Davis *et al.* [Davis79] suggested the use

of Generalised Co-Occurrence matrices to describe the spatial distributions of local features, such as edges and lines, rather than the spatial distributions of intensity.

Another statistical method that has been commonly used for texture analysis is the Gray-Level Run Length method. Introduced by Galloway [Galloway75], this method identifies sets of consecutive, collinear image points that have the same gray level and computes the length of each run or set. Stochastic or probabilistic measures have been proposed in order to model the interdependencies of pixels together with their neighbourhood. Haralick [Haralick79] did exploit this property through his autoregressive model. Other random field models have also been given considerable attention within the field of texture analysis. Markov Random Fields (MRFs) for instance have been used due to their capability to capture the local contextual information in an image [Tuceryan98]. Derin and Elliott [Derin87] used Gibbs Random Fields to model and segment textured images. Fractals, due to their capability to model properties such as roughness and self-similarity at different scales, have been used mostly in the generation of synthetic surfaces that have very near resemblance to natural surfaces such as plaster or rock. Due to these statistical properties fractals have also been used a lot in analysing image textures.

Structural approaches model and describe textures by assuming that textures are made up of primitives or texture elements. It is imperative, within structural methods, to be able to identify the primitives that make up the texture. The extracted primitives are then used in two different approaches for analysis. The “bottom-up” approach computes the statistical properties of the primitives and defines the mutual spatial relationship between them. The “top-down” approach extracts the placement rule that describes the texture, mainly using the Fourier transform [Matsuyama83]. Structural methods differ by their interpretation, extraction and representation elements.

Ever since the “theory of textons” was put forward by Julesz [Julesz81], much research has been undertaken in describing natural textures by extracting primitives which appear in near-regular repetitive spatial arrangements. Commonly referred to as textons, texels, tokens or blobs, these primitives are basically homogeneous regions of pixels with some invariant properties that may be defined by their distribution of intensity values or shapes. Voorhees and Poggio [Voorhees87] used a bank of Laplacian of Gaussian masks, applied at different scales and orientations, to extract blobs for texture discrimination. Similarly Tuceryan and Jain [Tuceryan90] used Difference of Gaussian filters to extract primitives from a texture image. They then used the extracted primitives to generate a Voronoi tessellation for the texture image

considered and thereafter extracted features from Voronoi polygons within the tessellation. While these methods are mainly “bottom-up”, some “top-down” have also been proposed. Zucker [Zucker76] and Connors [Connors79] used the gray-level co-occurrence matrices of texture images to determine their periodicity. Mathematical morphology has also been used for structural representation of textures. Haralick [Haralick79] used structural elements of different shapes to erode a texture image and extracted textural properties of the image as a result of the erosion process. Zucker [Zucker80] used semi-regular or regular tessellations of ideal textures which are then morphed to represent a real world texture.

In more recent versions of structural approaches, various new types of textons have been investigated to represent texture surfaces and images. With the fact that changes in illumination and viewpoint directions influence the appearance of surface textures heavily, textons that incorporate this element of visual texture have attracted strong interests. In order to model texture surface in terms of both reflectance and geometric information, Wang and Dana [Wang04] have presented a method that defines the local geometry of a surface texture in terms of a finite number of *geometric textons*. Lately, Zhu *et al.* [Zhu05] provided a study of the geometric, dynamic and photometric structures of textons in order to account for motion and illumination variations.

The use of spectral approaches, or channel-based approaches, was motivated by studies of human perception revealing that the human visual system decomposes the retinal information into a number of channels with varying frequencies and intensities [Beck87]. Several filter based approaches, proposed in literature, have tried to mimic the way the visual cortex functions by decomposing an input visual stimuli through the use of filter banks. Filter banks are designed in such a way that they capture localised information by targeting specific range of spatial frequencies at different orientations. Researchers have exploited both the spatial domain and the frequency domain (Fourier domain in particular) for texture analysis via filter banks. Laws [Laws80] was one of the first to apply filtering approaches for texture identification. Laws proposed a set of twenty five separable masks that were derived from three simple one-dimensional non-recursive filters and used the outputs from these masks as signatures for different textures.

Based on the assumption that the energy distribution in the frequency domain uniquely identifies a texture, a number of filtering methods applied to the power spectral domain were proposed. Coggins and Jain [Coggins85] applied a set of seven dyadically spaced ring filters and four wedge shaped filters to extract features for texture analysis. Banks

of Gabor filters have been extensively used in all areas of texture processing. Jain and Farrokhnia [Jain91] presented multi-filtering approach to texture analysis that uses a Gaussian shaped band-pass filters dyadically tuned to exploit differences in dominant sizes and orientations of different textures.

FRF (Filter-Rectify-Filter) models were used by a number of researchers to investigate the effects of texture element shape, size and spacing on visual perception of textures [Bergen88, Landy91, and Graham92]. FRF models consist of three different stages: (stage 1) a set of linear spatial filters, (stage 2) a point-wise nonlinearity, and (stage 3) further linear spatial filtering. Malik and Perona [Malik90] also used the model for the preattentive analysis of textures.

The wavelet transform [Mallat89] and its variants, such as the discrete wavelet transform and the wavelet packet transform, have also received considerable attention within the field of texture analysis. These are critically sampled filter banks that allowed the decomposition of a texture image into orientation and scale sensitive subbands[Kingsbury99].

Those models have mostly been utilised in the context of texture segmentation and discrimination, however we will only point out the models that have been used within a texture retrieval perspective. In the context of CBIR, texture was initially used as an image feature in combination with colour and shape in order to provide for more robust retrieval systems. QBIC, ImageRover, PhotoBook, Virage and MARS are some well-known CBIR systems that have used texture features in addition to colour and shape features, in order to provide for better retrieval performances [Velkamp02]. However with increasing availability of texture data and its growing application in different areas such as medical diagnosis or remote sensing, retrieval based on texture features only is being actively researched. This necessity has brought forward advances in textured region descriptors such as affine and photometric transformation invariant features.

2.4 Perceptual Approaches to Texture Retrieval

Interests in the field of texture perception date back as far as the early 1950s through work done by J. J. Gibson [Gibson50], however the real major step that brought the scientific research community to draw more attention to it can be attributed to B. Julesz more than a decade later. Most of his work was concentrated basically on finding spatial statistics of texture images that would help in preattentively discriminating between a pair of texture samples with the same visual properties, mainly brightness, contrast and colour [Julesz62 & Julesz75]. To further enhance his findings on texture discrimination Julesz proposed the “textons theory” to represent local textural features whose first and second order characteristics have perceptual significance for preattentive discrimination of textures [Julesz81].

The early investigations about texture perception focused mainly on the segmentation of textures into homogeneous regions that contribute in the discriminability of those textures. Departing from Julesz’s theory of textons, investigators such as Beck *et al.* [Beck87] used synthetic textures constructed by placing micropatterns onto predetermined regular or random placement maps to investigate segregation of textures. The micropatterns employed consisted of small visual stimuli in the form of dots, line segments, *Ls*, *Ts* and *Xs*. Recent investigations on texture segregation can be summarised in the work covered by Landy *et al.* [Landy04].

The use of artificial textures proved to be quite useful in the statistical modelling of textures, however, given that these textures are not representative of the set of natural textures encountered in real world, it did not motivate research about how humans perceive, analyse and categorise different texture categories. The availability of the Brodatz dataset [Brodatz66] allowed for more in-depth and realistic analysis of texture information and led to a number of investigators being interested in finding perceptual cues that humans use to discriminate between textures. Early psychophysical experiments performed by Tamura *et al.* [Tamura78] enabled the latter to determine some textural properties that humans commonly use to discriminate different texture categories. The experiments were performed using Brodatz textures and Tamura *et al.* identified six textural properties to categorise different texture groups, namely: coarseness, directionality, regularity, roughness, contrast and line- or blob-like. In a similar experiment to Tamura’s, Amadasun *et al.* [Amadasun89] asked human subjects to rank a set of 10 texture images chosen from the Brodatz album based on five different perceptual properties of texture: busyness, contrast, coarseness, complexity and

strength. The findings from their experiment indicated a strong correlation between computational features representing coarseness and texture strength, and also between those representing contrast and complexity.

Psychophysical studies undertaken by Rao and Lohse [Rao93a] were a huge boost for cognitive scientists in their attempts to provide a taxonomy for different categories of texture. The experiments performed by Rao and Lohse enabled the latter to identify high level features used to differentiate texture groups. They presented 30 texture images chosen yet again from the Brodatz album and asked human subjects to group similar ones together. Using a combination of hierarchical cluster analysis and multidimensional scaling, they identified that only three high-level perceptual features could account for most of the variability in the texture samples they considered in their experiment. These features were orientation, complexity and repetition.

Rao *et al.* then investigated the taxonomic relationships between texture categories as a follow-up to their previous work. In their new experiment they used 56 Brodatz textures and asked human subjects to rank the textures in a scale of 1 to 9, based on twelve predefined perceptual properties. The subjects were then asked to perform a sorting task to create texture groups [Rao93b]. After analysing the psychophysical data, they found out that only three orthogonal perceptual dimensions were sufficient to represent the 56 textures. They named the different perceptual dimensions with the following high level terms: non-repetitive vs. repetitive, non-granular vs. granular and non-directional vs. directional.

Still as a continuity to Rao and Lohse's work, Bhushan *et al.* [Bhushan97] performed further studies in order to establish a correspondence between texture words and texture images. To do so, they first performed a grouping experiment on 98 texture words from the English Language to determine any underlying common structure. Using hierarchical cluster analysis, they identified eleven major clusters and they termed those groups ranging from 'random' to 'repetitive'. In a second experiment they used the categories of texture images obtained from Rao and Lohse's earlier studies [Rao93a & Rao93b] to determine any systematic correspondence between the different categories of texture words (verbal space) and texture images (visual space). They deduced that the categories in the visual space and the verbal space were strongly associated.

More experiments were conducted by Heaps and Handel [Heaps99] that contributed to the attentive analysis of textures. The authors performed experiments to investigate the model that would best conceptualise the attentive similarity of natural textures. The experiments performed by Heaps and Handel used natural textures from two specific

datasets: VisTex dataset and the Brodatz dataset. The VisTex dataset consisted of 24 textures chosen from the Media Laboratory’s Vision Texture dataset [VisTex95]. The textures were chosen so that some of them resembled the Brodatz textures used by Rao and Lohse [Rao93a]. The sets of Brodatz textures used by Heaps and Handel were exactly the same as the ones used by Rao *et al.* Through their experiments, Heaps and Handel reached the conclusion that perceived similarity is context dependent and that the perceptual dimensions provided by Rao and Lohse in their respective study were somewhat meaningless.

In recent years, psychophysical studies aimed at improving the performance of texture retrieval systems have been investigated. Long and Leow [Long01], for example, identified the low performance of retrieval systems as being related to the perceptual inconsistency of computational features used for texture similarity measurements. Thus, they used psychophysical experiments in order to build a perceptual space to represent human judgments and presented a novel model to map computational features onto the perceptual space.

Payne *et al.* [Payne05] have performed a human study in order to produce a perceptually-derived ranking of similar Brodatz images that could be used as a benchmark to evaluate retrieval performance. In addition they proposed a “mental map” derived from human judgments to provide a scale for psychophysical distance and aid visual comparison of image similarity.

In a more recent study, Petrou *et al.* [Petrou07] question the idea of using preconceived texture properties, such as coarseness, directionality, regularity and others, in order to capture human judgments and then using computational features representing these properties to classify textures. The authors argue that such preconception severely biases the way human subjects would judge different textures when requested to do so. To avoid such problem, they propose the use of thousands of computational features that are not directly related to any high-level texture property. They then perform feature selection in order to identify the features that correlate with the human rankings obtained through 11 individuals asked to create similar groups from 56 Brodatz textures.

2.5 Tools and Techniques for Perceptual Texture Retrieval

In order to reduce the Human Perception Subjectivity, researchers in the field of CBIR have considered including the humans in the retrieval process. Cognitive scientists have made use of psychophysical experiment to understand how humans categorise textures whereas computer vision researchers have included the user in the query formulation process and integrated the human judgments as weights in optimised computational models. We present the different ways which have been employed to capture human judgments of texture. This section also presents a review of how psychophysical data is analysed and mapped to computational features.

2.5.1 Techniques to capture human judgments

Pairwise comparison

Pairwise comparison involves the presentation of two images or objects to a user who is asked to compare the two images according to some preset ranking criteria. This technique has mostly been used to identify perceptual features within images. In most cases, human subjects are presented with a user-friendly interface where the images to be compared are displayed together with a scale that allows the subjects to judge the level of similarity between them. Generally used in face recognition and retrieval systems, pairwise comparison has also been applied within the context of CBIR. Rogowitz *et al.* [Rogowitz98] used paired comparison to investigate the perceptual similarity between each pair of a set of 97 images. Since pairwise comparison requires $n(n - 1)/2$ comparisons for a set of n images, Rogowitz *et al.* [Rogowitz98] used a modified version in order to reduce the complexity. In fact they compared a chosen texture image with eight other texture images presented to human subjects at once. Payne *et al.* [Payne05] performed a similar comparison of Brodatz textures by placing the test texture image at the centre of the screen, surrounded by other textures. Volunteers were then asked to select four textures in decreasing order of similarity with respect to the test texture.

Perceptual grouping

Perceptual grouping is a term that was initially coined by Gestalt psychologists in order to represent the ways in which humans group similar structural elements within an image. The Gestalt theory related grouping to properties such as similarity, proximity, continuation, closure and symmetry [Lowe85]. This theory was used by Julesz where

he simulated textures made up of a combination of textons in order to investigate how humans segregate homogeneous texture regions within an image. As such perceptual grouping was initially used on a preattentive basis and for segregation purposes only. Attentive studies of texture perception led to perceptual grouping being applied as a method to capture human judgments.

Perceptual grouping, as explained by psychologists [Lowe85], refers to the human's visual ability to derive relevant groupings or structures from images without any prior knowledge of the image content. Perceptual grouping has been applied by a number of computer vision researchers in order to deal with human perception subjectivity. Even if most of them focused on finding regions of interests within images, some researchers have also referred to perceptual grouping as a technique to group together texture images that are visually similar. In this context, Rao and Lohse [Rao93a] used perceptual grouping to identify the relevant high order features humans use to group similar textures. Rogowitz *et al.* employed this technique to measure the similarities between any pair of images from a set of 97 photographic images representing a range of semantic categories, of viewing distances and colours.

Perceptual Ordering

Perceptual ordering also forms part of the “laws” of perceptual organisation as established by Gestalt psychologists. In this case psychologists were mainly concerned with how the human mind unifies and orders the perceptual environment when presented with a visual stimulus [Lowe85 & Wenger97]. Perceptual ordering in the context of CBIR basically refers to a process through which human subjects order a set of images when presented with a query image. This aspect of perceptual organisation relies heavily on prior knowledge that the subjects would have on the query image. They tend to ask questions like: *Where did I see something like this before?* Some CBIR systems have exploited this perceptual capability of humans in order to improve the performance of the retrieval engines. These systems are specifically known as ‘Query-By-Example’ (QBE) CBIRs. QBE systems normally capture the user's needs by presenting them with images that are representative of all the categories within the database being searched and allow the user to select one or more images as query image. IBM's *QBIC* for instance allows the formulation of queries based on objects within an image (e.g. find images containing cars), or based on specific image characteristics (e.g. retrieve images with a certain percentage of red or blue), or even based on specific shots within a video segment (e.g. shots with a high percentage of

movement). Other well-known systems that have employed this strategy are MIT's *PhotoBook*, the *Virage* technology which is used by Altavista's PhotoFinder, *CANDID* developed at the Los Alamos National laboratory and many others. Veltkamp *et al.* present a list of thirty nine CBIRs in [Veltkamp02], however their list is by no means exhaustive with more work being done in this area within the past five years.

Using Perceptual/Graphical cues

Graphical or visual cues have been employed by CBIR systems developers in order to give more flexibility to users to formulate their queries. The interfaces are designed in such a way that provides users with tools that represent perceptual features that the developers want to use in the retrieval process. Examples of tools provided are colour palettes, sketch pads, list of natural language keywords, shape representations, directional indications for textures mostly, or even a grid image in order to indicate position [Sclaroff99]. This type of query formulation demands more attention and expertise from the user; however the latter is not required to have any prior knowledge about the perceptual features being sought by the developers. Such kind of query formulation facilities have been applied in some well-known CBIR systems such as QBIC, VisualSEEK, Virage and WISE [Veltkamp02].

Relevance Feedback

In order to further reduce the human perception subjectivity and to allow better integration of the user within the retrieval process, researchers in the area of CBIR have focused on an interactive mechanism that allows a better understanding of the users need to be obtained. Known as Relevance Feedback, this technique can be closely related to QBE techniques whereby users are presented examples to formulate their queries. However, the main difference is the interactive part where the interaction of the user is not a crisp one-off process, but mainly an iterative one whereby the user is allowed to repeat the process of query selection until a satisfactory result is obtained. Users thus provide feedback on the results returned by the retrieval system and this feedback is used to enhance retrieval performance [Rui98]. Recent advancements on the use of relevance feedback show the use of *positive* and *negative* query examples [Kherfi03 & Franco04]. When query images are selected as positive examples, the common features among these images are given stronger weights so that all the target images having the same features are ranked highly in the retrieval phase. Negative examples were used in order to resolve the so called *page zero problem* which is the

situation where the initial query images presented to users are all irrelevant. As a result negative or very low weights would be allocated to selected negative examples so that they don't appear in the retrieval results [Sclaroff99].

2.5.2 Techniques for analysing psychophysical data

Analysis of psychophysical data is performed either to determine structure information within a dataset or otherwise to identify any dominant perceptual dimensions that retain the maximum variability within a set of textures and consequently that can be associated to some high level attributes used by humans to categorise texture. The similarity space which represents the psychophysical results is generally a sparse, high-dimensional space that is very difficult to visualise and assess. Thus, a reduced and more compact perceptual space is required.

Cluster analysis has been commonly used to identify any structural information whereas several dimensionality reduction techniques have been employed to investigate perceptual dimensions within low-dimensional spaces. The two approaches are presented in the remaining part of this subsection.

Cluster analysis

Cluster analysis is a tool that allows the partitioning of data into meaningful subgroups despite the lack of information concerning the number of subgroups or the other information about their composition [Fraley98]. In the context of psychophysical studies, cluster analysis has mostly been used to verify the meaningfulness of acquired perceptual judgments.

The main goal in using the clustering process is to reveal whether “sensible or believable” groupings exist within the dataset that can provide insight about any structural information in the dataset. In an extensive survey done by Jain *et al.* [Jain99], the latter identify two main categories of clustering techniques: partitional and hierarchical clustering. Following strong interests from psychophysicists to come up with a taxonomy of texture categories, the main form of cluster analysis that has been used to analyse psychophysical data is hierarchical cluster analysis.

Hierarchical approaches proceed by creating different sequences of data partitions, with each sequence corresponding to a different number of clusters. They either proceed by merging smaller number of clusters into larger ones, called agglomerative approaches or by splitting larger clusters into smaller ones, called divisive approaches [Gordon87]. Divisive approaches are generally impractical because it is impossible to restrict the

number of splittings. Agglomerative approaches are bounded by the number of groups in the first clustering stage. Additionally, agglomerative approaches are more intuitive to the way that humans create groups.

Dimensionality analysis

Psychometric method

Early work on finding perceptual features for texture perception has used psychometric methods to find the correspondence between human and computational rankings. Popular researches undertaken by Tamura *et al.* [Tamura78] and Amadasun *et al.* [Amadasun89] have mainly considered this method. The latter mainly consist of the computation of a representative ranking for the texture features being analysed by using the rankings performed by humans. The representative rankings are then used to determine the correspondence between computational rankings and human rankings. Thus an indication of which texture feature corresponds better to human judgment is obtained. This technique has recently been used by Abbadeni *et al.* [Abbadeni05] to test how well their autocovariance-based features perform with respect to human perception of texture images.

Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique that extracts the principal components of a feature space by performing a variance optimising rotation of that space. For the purpose of analysing psychophysical data, PCA was initially applied by Rao *et al.* (Rao93b) in order to investigate how much of the total variance of the physical texture space did each of the 12 perceptual properties considered (coarseness, granularity etc...) account for. More recently, Payne *et al.* [Payne05] applied PCA to the ranking scales allocated by human subjects to compare the similarity of regular textures. The aim was two-fold, first to extract principal components so as to have a view of the overall similarity of textures and second to investigate any structure in the similarity of the ranking scales.

Classification and Regression Trees (CART)

CART is a nonparametric regression technique utilised to select variables, and their interactions, from a large set of variables based on how well the variables can explain an expected outcome. This technique has been used in psychophysical based studies in which ratings or scales are provided to users to make judgments. Rao and Lohse

[Rao93b] employed CART to determine if the ratings of twelve scales provided to human subjects to categorise textures could predict the membership of textures within clusters generated using hierarchical clustering.

Multidimensional Scaling

Multidimensional Scaling (MDS) has been extensively used in literature to identify major perceptual dimensions through which the perceived similarity between textures can be represented and also as a visualisation tool that allows for visual inspection of the perceptual space in order to investigate existence of structural information within the dataset. A major assumption associated with MDS is that the latter can transform the original perceived similarity space into some kind of “psychological space” where the distances between textures approximate their perceived similarities. The perceptual space resulting from the application of MDS is bounded by orthogonal dimensions that can be represented by independent perceptual features. Identification of these perceptual features thus allows the creation of a different spatial domain that has been termed as the Perceptual Texture Space (PTS). Within Perceptual Spaces or Perceptual Texture Spaces, smaller distances between texture samples imply larger similarity values.

The use of MDS for texture perception was originally motivated by the fact that it was successfully applied for colour perception. Shephard [Shepard62] demonstrated that applying MDS on similarity judgments of colour patches could reveal the internal organisation of the colour space within only a 2D perceptual space. The latter 2D representation became commonly known as the Colour Wheel that has been associated with colour opponent mechanisms. In connection with texture perception, most research works available in literature relate to MDS as an exploratory technique used to characterise the process of mental representations [Gurnsey99].

In the context of psychophysical studies, the assumption that perceived similarity values behave like distances is too restrictive, especially when human judgments are involved. As a result a non-metric version of MDS has been used within studies investigating perceptual dimensions. Harvey and Gervais [Harvey81] applied MDS to similarities obtained by performing triadic and pairwise comparison of 30 artificial textures to investigate the relationship between the appearances of those visual textures and their fourier spectra. Rao and Lohse [Rao93a & Rao93b] used MDS in an effort to obtain a taxonomical arrangement of texture categories and also to identify perceptual dimensions that would account for most of the variability in 56 of the Brodatz textures.

MDS was also used by Gurnsey *et al.* [Gurnsey99] in an attempt to examine the representational system that determines the appearance of isolated patches of visual texture.

Direct Magnitude Estimation (DME)

DME is a standard psychophysical rating procedure that assumes that the human mind processes information as magnitudes and that cognitive categorisation is a means of delimiting magnitude information [Dewangan05]. In the field of texture perception, DME has been used as a standard rating procedure through which human subjects are asked to assign a number or value to a texture sample when compared to a reference texture based on some texture property such as regularity or coarseness. A rating value is pre-assigned to the reference texture and the human subjects can allocate a value greater or smaller in order to quantify the texture property being tested.

2.5.3 Mapping computational features to a perceptual space

A very important stage in building a perceptual retrieval system is being able to integrate human judgments in the retrieval process. This stage can also be viewed as a learning stage where the retrieval system ‘learns’ how capable the computational features representing the texture dataset are in predicting a retrieval outcome when presented with a query image.

Staying in the context of a “perceptual space” to represent perceptual similarity of texture images, we lay down different approaches employed by researchers to map computational features to a perceptual space. Not much has been done in this aspect and the purpose for constructing a perceptual texture space has been split between classification and retrieval. Payne *et al.* [Payne99], for instance, have used Kendall’s tau to correlate human rankings of Brodatz textures performed by 24 subjects with retrievals of the same textures via the use of a number of different features. In this case no mapping was done and the psychophysical data was used only for evaluation purpose.

Long and Leow [Long01] used a neural network of “invariant and perceptually consistent mapping” to create a perceptual texture space to represent a dataset of target texture images for retrieval. The first layer of the network takes as input the computational features which consist of Gabor features extracted at different spatial frequencies and orientations. The features are then passed to a layer consisting of

several translation invariant maps in order to detect some patterns in the Gabor inputs. The results from those maps are then projected into a scale and orientation invariant (but not illumination invariant) feature space. Long and Leow then used a set of nonlinear regressions, implemented using Support Vector Machines, to map the invariant features to the perceptual similarities of textures.

Petrou *et al.* [Petrou07] used the perceptual groupings performed by the human subjects in order to compute a stability measure for the computational features considered. The stability measure accounts for the variability in each class while applying different features. This method allows the authors to assign a set of weights to each feature representing how well it can represent each perceptual class. Petrou *et al.* then used a weighted distance function as a similarity measure.

2.6 Assessing requirements to develop a Perceptual Texture Retrieval System

The previous sections have helped us to reach a point of understanding of the research in the field of Perceptual Texture Retrieval. The investigation of perceptual approaches provided in Section 2.4 shows that even if there is growing interest to develop retrieval systems that satisfy the end-users' needs, the amount of psychophysical studies performed to understand and learn human's perception is still very limited. Indeed, literature shows that the psychophysical studies performed by Rao and Lohse, undertaken more than a decade ago, are still being referenced by researchers, for example Long and Leow. Recently Payne *et al.* [Payne05] and Petrou *et al.* [Petrou07] have also performed human studies of texture perception, however they used the same texture samples as Rao *et al.* and their studies were for comparative purpose with respect to Rao's results.

2.6.1 Reliability of Rao and Lohse's psychophysical results

The fact that very few psychophysical experiments have been performed to investigate perceptual dimensions for texture retrieval was in itself a big motivation in undertaking this research. However, an important issue to consider was whether Rao and Lohse psychophysical results could be employed within the scope of this thesis.

The main factor that drove us in questioning the reliability of Rao and Lohse's results was the dataset used by the latter to perform their human studies. Since its inception in 1966, the Brodatz album has always been a standard benchmark for texture processing. Rao and Lohse used 56 Brodatz texture photographs to perform their psychophysical study. The intensive use of the Brodatz textures cannot hide the fact that the conditions under which those photographs were taken are still unknown and the effects of illumination variation have largely been ignored.

However, recent studies in the field of Photometric Stereo performed by Chantler *et al.* [Chantler94 & Chantler05] have clearly demonstrated that changes in illumination conditions can drastically change the appearance of texture surfaces and significantly affect the output of the majority of texture features. Figure 2.1 and Figure 2.2 are used to illustrate this statement.

Visual inspection of the two images in Figure 2.1 introduces a bias whereby the image to the left appears to contain vertically oriented structures whereas the one the right does not appear to contain any apparent primitive that repeats itself, though some global

directional information can still be perceived. This contrasting description of the two images would heavily influence the decision of human subjects in putting the two images in the separate groups even if they originate from the same surface.

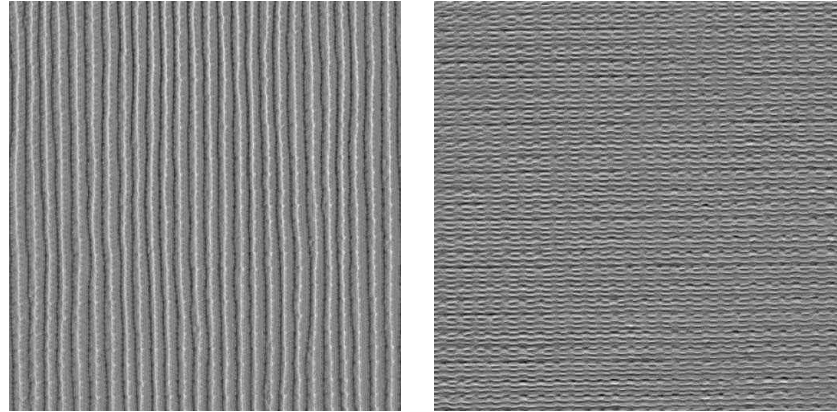


Figure 2.1– Effect of changing illumination direction on appearance of surface

It can be argued that the Human Visual System performs complex processing that allows humans to reconstruct missing or distorted information resulting from illumination variation and thus predict the type of texture that they are viewing. However, when computational measures to extract texture features are involved, the influence of illumination is very drastic. This argument is strengthened by the work done by Chantler *et al.* [Chantler05] on how the output of linear texture features behave with changing illumination conditions.

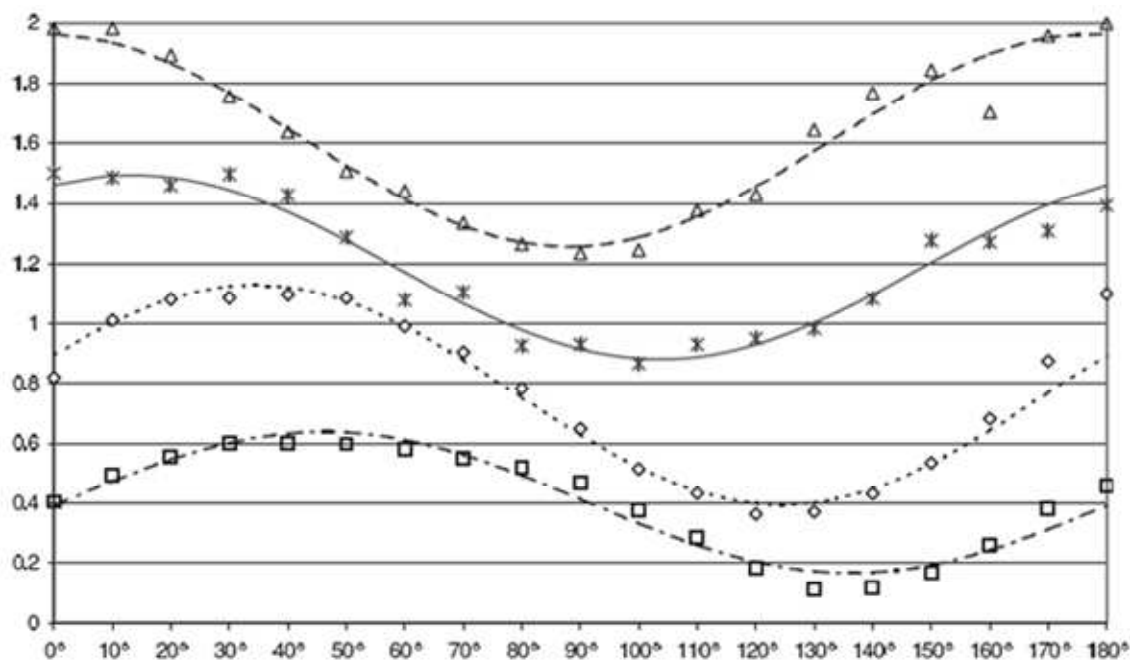


Figure 2.2 – Sinusoidal behaviour of texture features when applied to four different surfaces that have been illuminated under varying illumination conditions

Chantler *et al.* developed a sinusoidal model that explains the dependency of texture features on lighting direction. The outcome of their experiments is illustrated in Figure 2.2. The plots in Figure 2.2 demonstrate how the output of texture features vary when they are repeatedly applied to the same physical texture sample but under varying illuminant tilt angles. The curves show the best fit sinusoids to the measured outputs.

The plots are clearly indicative of the fact that when the illuminant tilt angles change, the outputs of the texture features follow a significant change. Hence, unless the texture features are extracted from the surface texture themselves or from texture images generated under controlled illumination, any kind of processing done on the textures would be heavily biased and are not reliable, especially for retrieval.

As a matter of fact, Rao and Lohse psychophysical results cannot be considered as a reliable source of human judgements to investigate perceptual dimensions for texture retrieval. Hence, this leads to the strong conviction that psychophysical experiment performed under controlled conditions is imperative and has, more than ever, strengthened our motivation in performing this research.

Additionally, researchers [Long01, Payne05 & Petrou07] have also used ‘identical’ textures from the Brodatz album to test for perceptually consistent retrieval systems. While all subimages from a Brodatz original would have been formed under almost identical imaging conditions, the same is not true between Brodatz originals. Thus, the retrieval systems that are tested using the ‘identical’¹ texture approach can exploit the differences in imaging conditions in addition to the difference between textures. As we have seen from Figure 2.1 and Figure 2.2, variation in imaging conditions (particularly those concerning changes in illumination tilt angle) can radically affect the power spectra and associated features of image texture. Hence texture features should be computed (where possible) from the height information of surface textures, or otherwise from texture images that have been under consistent illumination conditions.

In the remaining subsection, we identify the requirements for a new psychophysical study. We also identify tools required in analysing and applying psychophysical data to obtain perceptual dimensions and for texture retrieval.

¹For the purpose of this thesis we have define an ‘identical’ set of image textures to be those cropped from the same parent image of a homogeneous texture. Thus we can produce 9 ‘identical’ textures by dividing an image into 9 non-overlapping (or overlapping) subimages.

2.6.2 Dataset for a new psychophysical experiment

Eliminating the use of the Brodatz dataset and the associated psychophysical studies brings up a major concern: which dataset to use such that illumination and viewpoint conditions can be taken into consideration. Apart from the Brodatz dataset, several other texture datasets have been made public and utilised constantly in other areas of texture processing. A very brief review follows so as to determine whether these databases could provide texture samples to be used for a new human study.

CuRET (or CuRRET) -Columbia-Utrecht Reflectance and Texture Database

The CuRET database consists of three specific texture datasets that have been used to investigate the appearance of real world textures: (1) BRDF (*bidirectional reflectance distribution function*) dataset, (2) BRDF parameter database and (3) BTF (*bidirectional texture function*) database. Besides visual appearance, the CuRET textures have also been used extensively for texture analysis and synthesis. However, the texture surfaces have been captured with both changing illumination and viewing directions, which is not very practical for retrieval (we would like all surfaces to be viewed from the same position). Moreover, the datasets combine both specular and diffuse surfaces which imply different reflectance models to generate the images. In order to avoid any bias in human judgements only one reflectance model is preferred. And most importantly, the surface texture height maps are not available to generate texture samples for the human study. However, texture images generated under the same condition can be selected, thus making the CuRET dataset a potential candidate for psychophysics.

VisTex – Vision Texture Lab database at MIT

The VisTex database was conceived with the intention to provide large set of high quality textures that would be used for, and by most computer vision algorithms in texture processing. However, this set was captured with varied studio lighting conditions including daylight, artificial-fluorescent and artificial-incandescent. Additionally, it does not conform to any rigid frontal perspective. Hence, this makes its use for a human study, where controlled illumination and viewpoint are prime concerns, quite irrelevant.

OuTex - University of Oulu Texture database

This database was generated to test texture segmentation and classification algorithms. The textures captured reflect changes in illumination, surface rotation and resolution. Texture images captured at three different illumination positions are available; however these three illumination positions are coplanar and cannot be used to recover the surface height map through photometric stereo. However as in the case of the CuRET dataset, if texture images are chosen such that they all have the same capture condition, then the OuTex dataset is also a potential candidate to be use for human study.

MeasTex

MeasTex is a texture image database that is accompanied with quantitative measurement framework for image texture analysis and synthesis. MeasTex is solely a collection of 2D texture images with the illumination and viewpoint condition being unknown.

PhoTex – Photometric Texture database at Texture Lab

The PhoTex database consists of a set of rough surfaces that have been captured at different illumination directions and viewpoints. This database satisfies our requirements of acquiring height maps that can be used to generate controlled texture images, however, since it contains only one category of textures (rough surfaces such as plaster or rock), it does not contain sufficient variability for our purposes.

The above investigation in the available texture databases shows that very few texture surfaces can be ‘borrowed’ to create a dataset for a new human study. Additionally, this small set comes from the PhoTex database.

The OuTex and the CuRET datasets can provide images generated under the same condition, which makes them likely candidates for human study, however additional selection criteria need to be investigated to determine whether new samples need to be captured. The other selection criteria are considered in Chapter 3.

2.6.3 Selecting tools for the capture and analysis of psychophysical data

A review of the different techniques to capture human judgements has already been presented in section 2.5.1. Pairwise comparison has been used by several researchers; however the fact that $n(n - 1)/2$ comparisons are required to obtain perceived similarity, this method is not very practical when large datasets are considered. Even if the modified versions applied by Payne *et al.* and Rogowitz *et al.* do reduce the number of comparisons, they, however, only help in increasing the complexity of the psychophysical setup. The modifications imply that (1) a suitable interface needs to be designed to accommodate a test texture with several targets, and (2) there is a main concern about which target textures should be presented with the test sample.

Perceptual ordering, relevance feedback and perceptual cues have been thoroughly exploited by CBIR developers. However, they all need to be employed in a computational context and cannot be considered for the analysis of psychophysical data. Perceptual grouping, on the other hand, has already been used successfully in the field of texture perception to determine perceptual dimensions (cf. Rao and Lohse). It also does not require any complex setup and can be applied to match or compare any size of texture images. Additionally, by providing the human subjects with a view of all the images in the database, perceptual grouping allows the subjects to relate to the context in which the experiment is performed. This technique also has a couple of limitations. Firstly it works well when the number of samples is small. For large datasets (>300) it becomes difficult to present observers with all the samples at once. Secondly grouping large number of samples may also result in boredom and fatigue in observers, hence contributing to biased results. However, the advantages of this technique overshadow its limitations thus making it a strong candidate for capturing judgments.

Further motivation to use this technique comes from the fact that in comparison to other popular techniques such as pairwise comparison, a subject does not need to remember previous judgments as all the images remain in his/her field of vision, whether grouped or ungrouped. Grouping also eases redundancy reduction. As mentioned before natural textures contain a lot of redundant information. However, while comparing the texture samples, subjects need to ignore the redundant information and identify only common features. This process is more difficult when pairwise comparison is performed, whereas with the grouping, task comparison is easier as the subject can use information from as many samples as required to perform the task.

The main objectives in analysing psychophysical data are mainly:

- 1) To identify relevant structure in the comparisons or groupings performed by human subjects,
- 2) To represent the texture samples (judged by the subjects) in a low-dimensional ‘psychophysical’ space such that the separation of the textures within that space represents as closely as possible the perceived similarity of textures.

Hierarchical analysis has been used by several researchers to satisfy the first objective, mainly because, it provides a simple and inexpensive means to create cluster that could be verified “visually” and additionally it allows the investigation of how clusters of similar samples are related. The idea of moving up and down a hierarchical tree causing the merging and splitting of texture categories perfectly with the aim of deriving homogeneous and sensible enough texture groups which could be part of an interpretable taxonomy [Gordon87].

In the case of dimensionality analysis, Multidimensional Scaling, especially its nonmetric version, has proved to be very successful in transforming high-dimensional sparsely sampled spaces to lower dimensional spaces whereby the similarity information being assessed, is still preserved [Rao93a & Rao93b]. Moreover, the values generated by MDS do not have any specific meaning with only the spatial configuration represented by those values being of interest. Besides, MDS also allows the graphical representation of the structure of a complex dataset.

2.7 Conclusion

In this chapter, we identified the challenges faced by researchers from various communities in developing perceptually consistent retrieval systems. Based on the review of perceptual approaches to texture retrieval, we can conclude that very few psychophysical studies have been undertaken to identify perceptual dimensions for texture retrieval. Additionally, the existing studies were all performed using texture images obtained under unknown illumination and viewpoint conditions.

The visual appearance and the numerical values of common texture features can be dramatically affected by changes in direction of illumination, hence

- texture features should be computed (where possible) from height information or, if this is not possible, from consistently illuminated samples,
- observers should be provided with images obtained under consistent illumination conditions for psychophysical studies.

Perceptual grouping is an easy and intuitive way to capture human judgments. It is also a practical approach for deriving similarity data using a reasonable number (circa 100) of textures.

Hierarchical Cluster Analysis is a useful technique for the analysis and visual inspection of similarity data whereas Multidimensional Scaling is a commonly used technique for dimensionality reduction of large perceptual spaces (i.e. high-dimensional similarity data).

Chapter 3

Design and Implement the Psychophysical Experiment

3.1 Introduction

The previous chapter established that the number of psychophysical surface texture experiments reported in the literature is low. Most papers cite Rao's work [Rao93a] as the main source for texture dimensions and many of them use Rao's data to train their retrieval or classification systems. However, as it has been pointed out in Chapter two, all the psychophysical experiments performed so far have presented human subjects with texture images whose illumination and viewpoint information are unknown, and uncontrolled. As explained previously, this is likely to have biased the calculation of texture features and provided potentially confusing stimuli to the observers. However, these papers have shown that 'perceptual grouping' is a useful tool for economically producing similarity data.

Thus, the main objectives of this chapter are:

- (a) to develop two databases that are suitable for psychophysical experiments and that in particular use images that have been captured under controlled and known conditions,
- (b) to design and perform perceptual grouping experiments that will use these data, and
- (c) to derive similarity matrices from these experiments.

Figure 3.1 shows a breakdown of the steps considered necessary for performing such experiments.

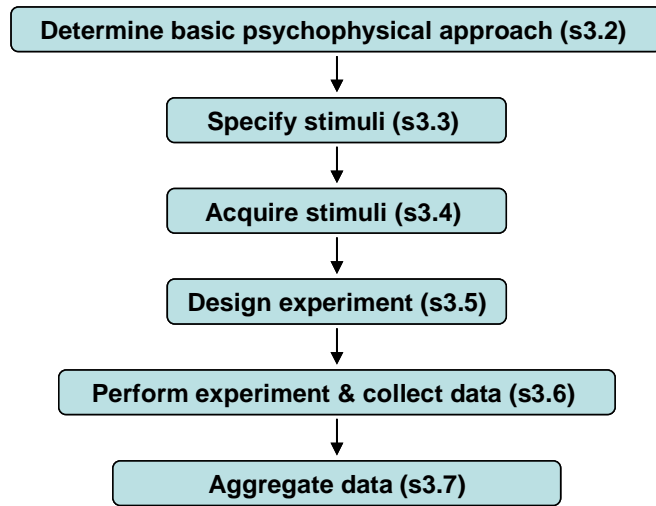


Figure 3.1 - Steps required for psychophysical experiment setup

The first step, discussed in Section 3.2, considers the overall approach to the experiment from the pragmatic point of view of how to determine similarity data using a reasonable sized database of around 100 samples. Sections 3.3 and 3.4 specify the necessary characteristics of the stimuli and determine suitable sources. The next section considers the detailed design of the experiments while Section 3.6 covers its implementation and Section 3.7 describes the aggregation of the resulting data.

3.2 Basic Approach

The objective of this thesis is to develop perceptually relevant texture retrieval systems. Ideally this would make use of an exhaustive number of example retrievals performed by a reasonable number of observers on a database that was representative of all possible surface textures. Clearly this is not realistic given the resources available to the individual researcher. The collection of suitable samples is surprisingly time-consuming and so the use of databases containing approximately one hundred textures was considered to be an obtainable goal. Unfortunately even for this number of samples it is unrealistic to expect observers to provide error free retrievals of say 30 ordered textures using each of the 100 samples as a query within the 30-40 minutes that it was thought they could maintain their concentration. However, a pilot ‘grouping’ experiment showed that it was feasible for an observer to sort this number of textures into an arbitrary number of perceptually similar groups. Furthermore, if such a task is performed by several observers then a similarity score between any two samples can be estimated by counting the number of times that observers have placed the pair within

the same group. The ‘ideal’ retrieval for any query may then be estimated by identifying the remaining textures in the database in order of their similarity scores.

Originally it was intended to provide observers with moving imagery of each surface – however, it was quickly realised that even the relatively large TFT displays available today do not contain sufficient pixels for the display of the number of textures required at reasonable screen resolutions. Thus the overall approach decided upon comprised:

- capture or obtain images of around 100 sample textures obtained under controlled and known illumination conditions;
- perform experiments in which observers were asked to group photographs of the samples into an arbitrary number of groups, the only criterion being that the members of a group should be perceptually similar;
- construct similarity matrices from the observers’ groupings.

3.3 Specification of the Stimuli

Most researchers agree that texture is a highly complex phenomenon. However, Chapter 2 reported that the number of psychophysical studies on the psychophysical aspects of surface texture retrieval reported in the literature is surprisingly low. It was therefore decided to keep stimuli as simple as possible, to control environmental conditions as far as was practicable, and to focus purely on the core issue: surface texture retrieval. With this strategy in mind the following criteria for the stimuli were drawn up:

- a) Two databases would be used: a general dataset covering as wide range of surface textures as was practicable, and a more specialised set covering a particular application domain. These datasets are referred to as Tex1 and MoMA.
- b) The datasets should contain around 100 samples or less – so as to allow observers to complete the grouping task within thirty minutes.
- c) The datasets should consist primarily of *surface* textures and not contain confusing surface markings – and so at least Tex1 should contain purely monochrome, constant albedo, lambertian surfaces.
- d) Samples should contain a single homogeneous texture, so as to ensure that observers were not using different parts of a sample to find a ‘match’ with different textures.

- e) The samples should be of approximately the same granularity (or scale) and roughness – as these two texture dimensions have already been investigated in depth.
- f) The datasets could be generated synthetically, or be captured from real surfaces, but in either case the observers must be able to envisage the imagery as being of *believably real* surfaces.
- g) The imagery must be of sufficient resolution and size for the observers to be able to perceive the characteristics of the surface texture, and yet small enough, such that they could manipulate and view all of the samples simultaneously on a large table.
- h) The imagery presented must be generated under a single set of illumination conditions.
- i) Height data should be available for at least Tex1 to allow the generation of texture features unbiased by illumination conditions.
- j) The height-data requirement could be relaxed for MoMA providing that uniform imagery illumination conditions is used so that *biasing* of the texture features would be consistent across the dataset.

3.3.1 Dataset consisting mainly of believable surface textures

In order to get an understanding of how humans categorise texture surfaces, it is important to present them with imagery that they can envisage as being of real surfaces. A simple definition of “believable textures” would be: textures that originate from our environment, or could be thought of as originating from our environment. The main reason for this criterion is that the human visual cortex is likely to be highly non-linear, and tuned or optimised for such type of visual stimuli. Interpolation between non-ecologically valid stimuli is therefore not guaranteed to produce consistent results.

3.3.2 Granularity and Roughness

An initial pilot study demonstrated that humans tend to group together all fine-granule textures, independent of the structure of the texture elements or even their placement within the surface plane. However, granularity (or scale) has been identified in all of the previous studies as a major dimension of texture, and as the datasets were to be limited to around 100 samples it was decided to focus on the more challenging characteristics of surface texture. Likewise for roughness. It was therefore decided to try and limit the range of roughness and granularity exhibited by the samples.

3.3.3 Controlled Illumination and viewpoint conditions

This can be considered as the most important design issue in selecting or generating samples for perceptual grouping. Chapter two has shown that all psychophysical experiments performed so far for texture surface perception have considered only texture images generated under unknown illumination and viewpoint conditions. However, Figures 2.1 and 2.2 clearly show that:

- (a) illumination can dramatically bias the outputs of common texture features, and
- (b) the perceived qualities of a surface can change significantly with changes in illumination direction.

3.3.4 Variety of texture samples

As discussed previously, the failure to generate sufficient samples for the psychophysical experiment can seriously bias the results of the experiment. When texture surfaces are considered, the issue is not only about having enough samples to perform the experiment, but also about having sufficient categories of textures covered by the experiment. The studies performed by Rao *et al.* [Rao93a & Rao93b] have been of enormous help in achieving this objective. Even if the dataset used by Rao *et al.* was not generated under controlled conditions, it is nevertheless valuable in the variety of textures it presents.

For Tex1 we have therefore tried to obtain as wide a variety of surfaces as was feasible. This requirement was relaxed for the MoMA dataset which is application specific and was taken as provided by domain experts.

3.3.5 Matte surfaces

Gloss is a specific surface appearance property that is described in terms of the reflectance of a material surface. In the case of specular reflection light is directed at an angle opposite to the incident light where as for lambertian reflection, light is diffused equally in all directions. Figure 3.2 illustrates how an incident light is scattered on two different surfaces.

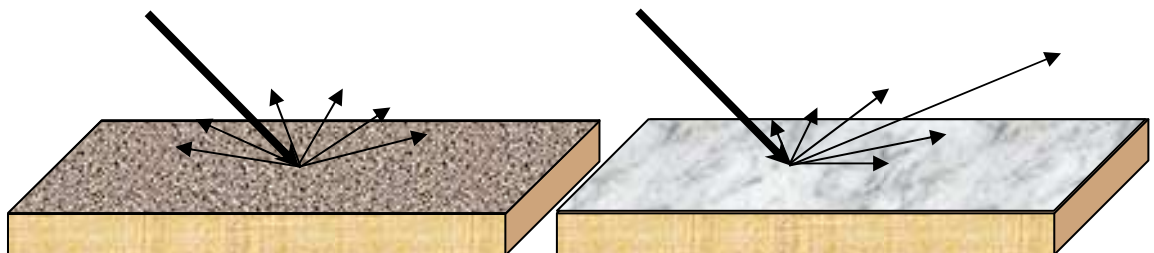


Figure 3.2– Lambertian reflection (left) and specular reflection (right) on somewhat matte and mirror-like surfaces respectively

It has been shown that the degree of ‘glossiness’ significantly affects our perception of surface characteristics [Ho08] and the presence of gloss can influence our perception of the global structure of surfaces by making them appear more curved [Todd97 & Ming86]. It is important that the grouping experiment is performed without any bias resulting from the surface properties themselves. Given that in the first instance we are primarily interested in how humans perceive and categorise surface textures, the surfaces used for Tex1 are rendered using a simple Lambertian reflectance model (i.e. considering only matte surfaces).

3.3.6 Constant albedo surfaces

Albedo information characterises the reflectance of a given surface and it basically represents the amount of light which is scattered from that surface when an array of light is incident on it. Thus, areas of high albedo on a surface would reflect most of the incident light and they look brighter than areas where the albedo is low due to absorption of light energy by the surface [Lin99]. As in the case of surface reflectance, natural surfaces, in practice, are composed of patches that have different light energy absorption capabilities. Thus, the variation in absorption level means varied brightness level across the same surface. This phenomenon can influence the judgments of human observers comparing samples. Therefore for Tex1 we rendered all surfaces as having constant albedo.

3.3.7 Size of the datasets

In order to obtain a fair and unbiased judgment from the human subject, it is crucial that the decision for grouping textures is performed under no influence of fatigue, fluctuations in mood or even boredom, therefore, the size of the dataset cannot be too large. We found that datasets of around 100 samples could be grouped by the average observer in 20-30 minutes.

3.3.8 Size and resolution of images

As mentioned in Chapter two, most psychophysical experiments have used digitised versions of the Brodatz album. The latter normally consisted of photographs occupying a picture area of about 19.5 by 24 cm. The digitised versions used by psychophysical studies varied in the number of pixels and range of grey levels used. Resolutions of 384 by 384 pixels with 256 grey levels and 256 by 256 pixels with 64 grey levels are most common in literature [Tamura78, Amadasun89 & Rao93a].

Naturally the resolution required is a function of the viewing distance and the size of the photographs. In [Rao93a], Rao *et al.* used a 4 by 5 inch picture area to represent the

Brodatz textures for their psychophysical studies. We have also found that this kind of size of photograph gives a good compromise between providing a large enough sample for assessment purposes, but is still small enough that it is practicable to manipulate and layout around 100 photographs on a conference table. We have therefore generated all pictures at 4 by 4 inch. Given this size of photographs it was found that 512x512 eight bit images gave sufficient resolution when laid out on a table and viewed from a standing position.

3.4 Acquisition of Stimuli

As previously discussed it was decided to obtain two datasets:

- Tex1: the primary dataset comprising at least one hundred samples drawn from as wide a range of surface textures as was practicable together and
- MoMA: an application specific dataset defined by domain experts and focusing on a narrow set of texture types.

It was decided that the exacting height-map requirement would only be applied to Tex1 and that images taken under controlled and consistent conditions would suffice for the MoMA set.

These sets both had to comply with the criterion detailed in the previous section. These are summarised in Table 3.1 for convenience.

<i>C1. Number of samples</i>	50-150
<i>C2. Surface reflectance characteristics</i>	Lambertian, Monochrome, constant albedo
<i>C3. Homogeneity</i>	Single level homogeneous textures required
<i>C4. Realism</i>	Imagery should believably represent real surfaces
<i>C5. Resolution and size</i>	4" x 4" at least 512x512x8bit
<i>C6. Height data</i>	Should be available for at least Tex1
<i>C7. Consistent environmental conditions</i>	Illumination conditions and viewing geometry must be consistent throughout each dataset
<i>C8. Roughness and granularity</i>	Approximately constant throughout each data set

Table 3.1– Summary of criteria for dataset selection

3.4.1 Sourcing the Datasets

Chapter 2 surveyed the publicly available datasets and identified two possibilities: CURET and Outex. Unfortunately neither of these datasets forms suitable candidates for Tex1 as neither contains height data, or suitable image sets for reliably deriving height data. Neither are they suitable as our second application specific dataset as they contain a wide range of randomly collected textures. Furthermore, many of the samples in both of these databases violate criteria C2, C3, and C8.

A) The application specific dataset

A chance contact from the Museum of Modern Art, New York revealed that they had a collection of photographic papers that they had collected and imaged using a flatbed scanner in order to emphasise their surface texture. As these textures covered a narrow range of texture types and as they were obtained under identical imaging conditions it was decided to use this source for the second dataset. Furthermore, staff and associates of MoMA were very keen to participate in the psychophysical experiments. Images of the MoMA dataset are shown in Appendix A (Figure A.4 and Figure A.5).

B) The Tex1 dataset

Reluctantly it was decided that for Tex1 the data would have to be generated by the author. Originally the aim was to obtain height maps of real samples using photometric stereo – a cheap and fast method of obtaining texture height data. However, it was found that the number of *homogeneous* texture samples of the appropriate scale, size and roughness that could be brought into the laboratory was surprisingly low. Texture synthesis was therefore used to augment this dataset with the proviso that the resulting imagery should be as believable as that obtained from a real surface (i.e. meet criteria C4).

The rest of this section describes the acquisition of height information from real samples where this was possible and the generation of synthetic data when capture was not feasible. The complete Tex1 dataset consists of fifty-two natural textures and sixty-eight synthetic ones, for a total of hundred and twenty surface textures. The Tex1 dataset is provided in Appendix A (Figure A.1 to Figure A.3). The Tex1 textures are denoted in the following way: “T” + index (for example T31). The labels of the Tex1 textures in Appendix A are subscripted with an “n” or a “t” to distinguish between the natural and synthetic textures.

3.4.2 Tex1 – Capturing height-maps of real surfaces

The use of surface height maps instead of images as the primary source provides two main advantages:

- (i) texture features calculated directly from height maps are *independent* of any imaging conditions used to view the surfaces (unlike those features shown in Figure 2.2);
- (ii) if height maps of glossy textures with varying albedo have been obtained then they may be rendered to meet criterion C2.

The method selected for surface height capture was photometric stereo [Woodham80]. It is a cheap, fast method that has been used in the Texture Lab over the last ten years [McGunnigle01 & Dong05]. The theory assumes:

- an orthographic projection system with the camera axis being perpendicular to the plane of the surface,
- the light vector and intensity that is constant over the surface,
- that shadowing and occlusion are negligible, and the surface is Lambertian.

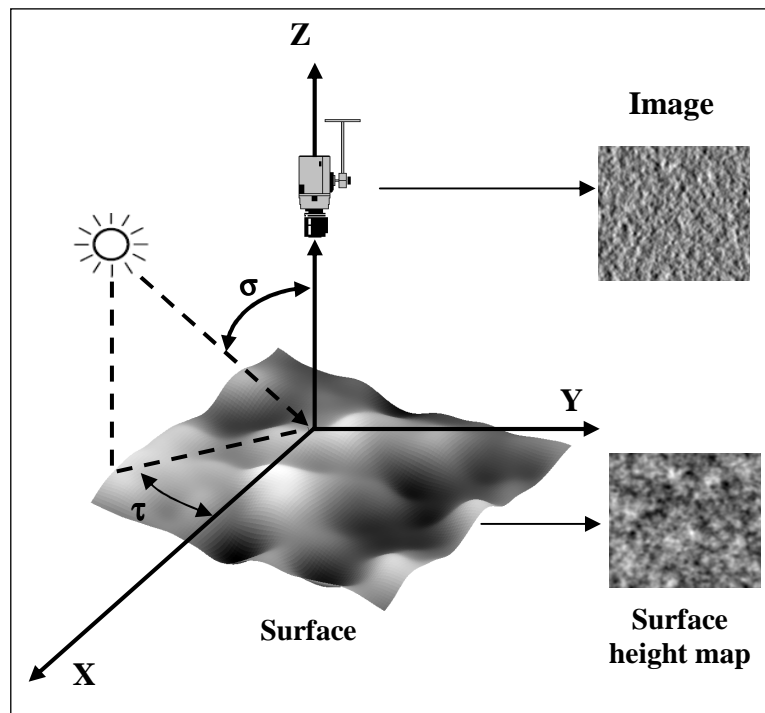


Figure 3.3 –Capture geometry

Figure 3.3 illustrates the relevant geometry. The camera's optical axis is along the Z axis and the texture surface lies in the plane X-Y. The light source is placed at a distance far from the surface relative to its size in order to approximate orthographic light projection. τ denotes tilt angle and represents the angle that an illuminant vector

projected onto the surface plane makes with X axis, whereas the slant angle, σ , represents the angle the illuminant vector makes with the Z axis. Based on the above assumptions the image function is defined as follows:

$$i(x,y) = \frac{-p(x,y) \cos \tau \sin \sigma - q(x,y) \sin \tau \cos \sigma + \cos \sigma}{\sqrt{p^2(x,y) + q^2(x,y) + 1}} \quad (3.1)$$

The setup shown in Figure 3.3 results in only one image. To recover the surface topology of a given texture material we require at least three images taken at different (non coplanar) angles. This provides three simultaneous equations of the form of equation (3.1) which can be solved to provide an estimate of the per-pixel scaled surface normals. These in turn can be used to derive the unit surface normals and albedo values.

Hence Photometric stereo has been employed to recover the real world textures to be used in the psychophysical experiment. Figure 3.4 shows a set of photometric images taken at tilt angles 0° , 90° and 180° . It is important that the three photometric images provide enough change in illumination gradient so that the partial derivatives for the surface can be estimated.

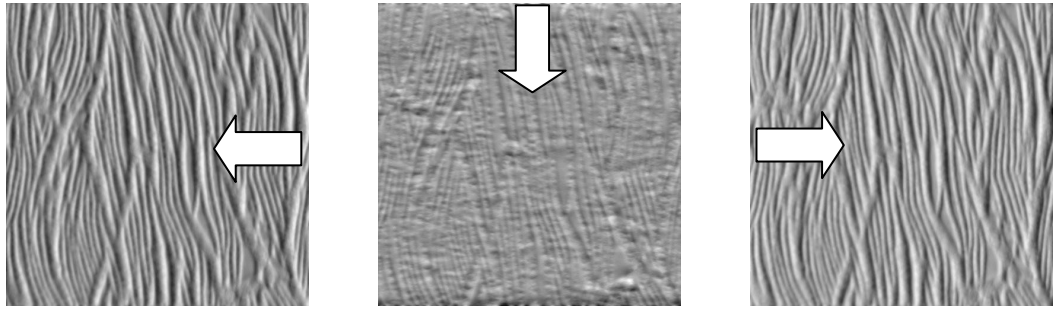


Figure 3.4 – Photometric images for the anaglypta surface taken at a fixed slant of 45° and tilt angles 0° , 90° and 180°

A summary of the mathematical derivation follows: the reflectance functions for the three images are given, in matrix format, by:

$$i_d(x,y) = \rho(\mathbf{l}_d \mathbf{n}) \quad \forall d \in \{1 \dots r\} \quad (3.2)$$

$i_d(x,y)$ represents the reflectance intensity at point (x,y) for a given illumination vector \mathbf{l}_d , r represents the number of photometric images required (3 in our case), $\mathbf{n}(x,y)$ is the surface normal unit vector at a given position (x,y) in the surface plane and is defined as

$$\mathbf{n} = \frac{(p, q, -1)^T}{\sqrt{p^2 + q^2 + 1}} \text{ with } p \text{ and } q \text{ being the partial surface derivatives.}$$

Assuming (i) the surface albedo, ρ , to be uniform and (ii) a non-singular illumination matrix \mathbf{L} where $\mathbf{L} = (\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)^T$, then the scaled surface normals are estimated by taking the inverse of the illumination matrix as follows, $\mathbf{m} = \mathbf{L}^{-1} \cdot \mathbf{I}$, with \mathbf{I} being the reflectance matrix derived from the photometric images. The vector $\mathbf{m} = (m_1, m_2, m_3)^T$ is then used to derive the partial derivatives of the surface being recovered with $p = \frac{m_1}{m_3}$ and $q = \frac{m_2}{m_3}$. Further details can be obtained from [Gullón03, Dong03, & Wu03].

Once the partial derivatives have been estimated the surface is recovered using a non iterative version of Frankot and Chellapa's integration method [Frankot88]. Figure 3.5 shows height maps of some natural surfaces that have been considered in this thesis.

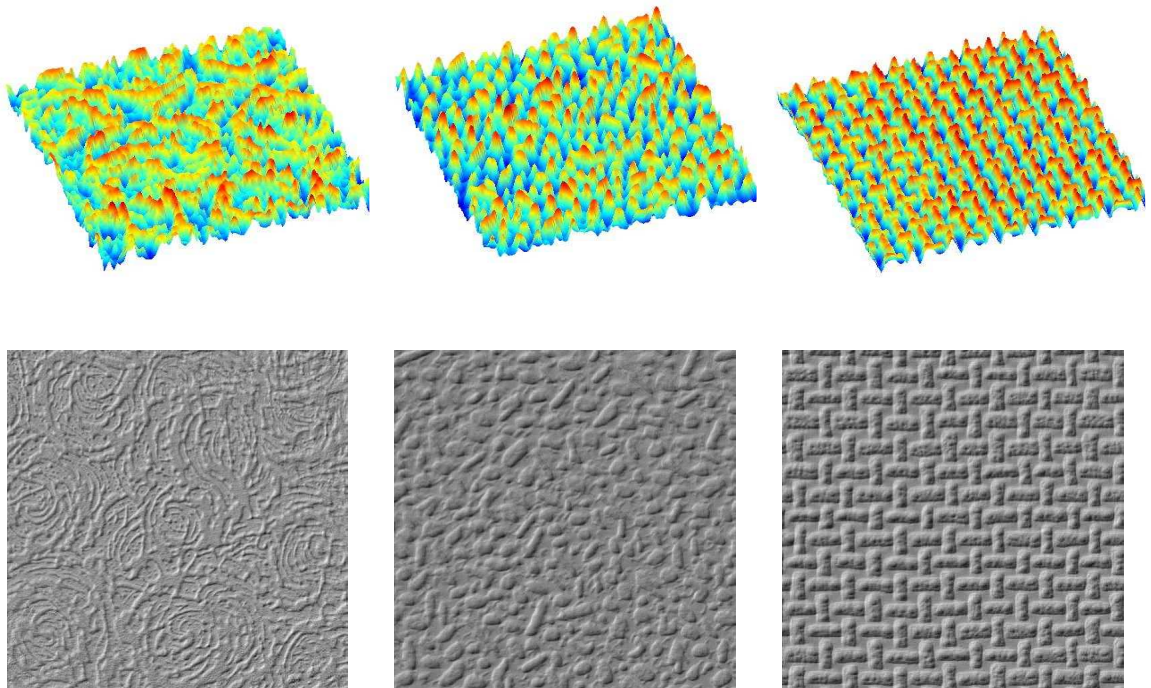


Figure 3.5 - Row 1 shows the height maps of some real world surfaces and row 2 shows their corresponding images

3.4.3 Tex1 – Synthetic texture generation

Synthetic textures have been intensively used in understanding the mechanisms of texture segregation. Pioneering research undertaken by Julesz *et al.* [Julesz81 & Caelli78] focused on the use of randomly placed texture elements in order to identify which pair of texture elements would segregate easily. The micropatterns or texture elements that were used to generate the synthetic textures consisted mostly of dots, lines and stimuli in the form of Ls, Xs, and Ts.

Even though the computational theories put forward to examine the discriminability of such textures have been seen to work well, synthetic textures constructed on the basis of micropatterns are not representative of natural textures that we encounter in day to day life and so violate criterion C4 concerning realism. The challenge, therefore, is to come up with synthetic textures that are similar to real examples.

Texture synthesis has been used extensively to generate realistic texture images or surfaces. 2D texture synthesis methods have mostly been categorised as being parametric or non-parametric. Parametric synthesis methods perform synthesis by matching global or local statistics between a sample image and result images in a feature space. Some well-known parametric methods have been implemented by Gagalowicz and Ma [Gagalowicz85], Heeger and Bergen [Heeger95], De Bonet [DeBonet97] and Zhu *et al.* [Zhu98]. Most parametric methods are based on multiresolution approaches that use a bank of filters and sampling strategies for the statistical encoding of textures. A well referenced approach is the one proposed by Portilla and Simoncelli [Portilla00].

Non-parametric methods are sometimes referred to as ‘synthesis by example’. Paget and Longstaff [Paget96 & Paget98] proposed a non-causal synthesis algorithm based on Markov Random Field (MRF) to model arbitrary textures. Others such as Efros *et al.* [Efros99] and Wei *et al.* [Wei00] used a neighbourhood search strategy to synthesise textures. Efros and Freeman [Efros01] also proposed image quilting which consisted of synthesising new textures by stitching together small patches of existing images. Other patch-based approaches were proposed by Liang *et al.* [Liang01] and Kwatra *et al.* [Kwatra03].

Even if all the methods available in literature perform well when tested with specific texture samples they have not been generalised to work on a large variety of textures. Additionally methods such as the one based on MRF, which can be used to synthesise a large variety of natural textures, are normally computationally expensive. Patch-based approaches have the disadvantage that patches selected to synthesise a texture contain limited amount of information and cannot provide for good statistical description of real-world textures that normally contain features at widely varying scales. Due to those issues, we have decided against using the synthesis methods available in literature and instead we have used some basic approaches that (1) are inexpensive to implement and execute and (2) are more suitable to generate synthetic textures that are closer in appearance to the existing natural textures. The approaches correspond to frequency domain synthesis of textures and are presented in the rest of this subsection.

A) Frequency domain synthesis – $1/f^\beta$ noise

One-over-fBeta-noise (random phase fractals) have been widely used to model textured surfaces given that they can be easily generated and at the same time the surface relief produced appears to be that of natural surfaces[Pentland84]. These surfaces are fully represented by the power spectrum, and hence are easily generated by synthesising a suitable power spectrum function:

$$S_f(w, \theta) = \frac{k_f}{w^\beta} \quad (3.3)$$

and combining it with random phase.

$S_f(w, \theta)$ represents the power spectrum of the fractal surface, k_f is a constant that controls the surface variance and β is the power roll-off factor [Kube88]. Other variations of this model have been defined in literature that split the power spectrum into two fractal dimensions. Examples are the Mulvaney and the Ogilvy models. Surfaces generated using the Mulvaney model have a flat spectrum at lower frequencies and a roll-off value of 3.0 at high frequencies whereas the Ogilvy model allows the generation of directional surfaces with different power spectrum characteristics for different directions [Gullón03 & McGunnigle01]. Figure 3.6 illustrates two texture images obtained after rendering a Mulvaney surface and an Ogilvy surface. While the Ogilvy surface seems to contain some directional information, the surface relief for both surfaces are very isotropic.

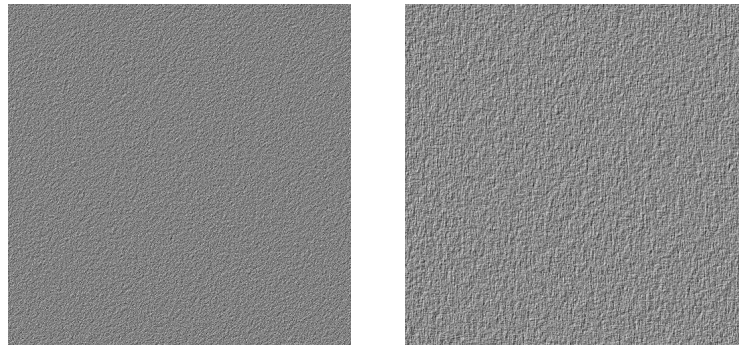


Figure 3.6 – (a) Mulvaney surface (b) Ogilvy surface

B) Other frequency domain functions

Gluckman [Gluckman05] created visual patterns based on a combination of discrete frequencies in the power spectral domain in order to examine the ability of using filter based statistics to discriminate between an arbitrary set of visual stimuli.

Thus the power spectrum, $F(w)$, of a given pattern is a combination of a set of functions (f_1, \dots, f_n) , with each function f_i being a set of integer frequency pairs $\{(u_1, v_1) \dots (u_m, v_m)\}$. Figure 3.7 illustrates this approach; the first row presents different power spectral functions generated by combining two functions f_1 (blue spots) and f_2 (black squares), whereas the second row provides the corresponding patterns generated using those power spectrums.

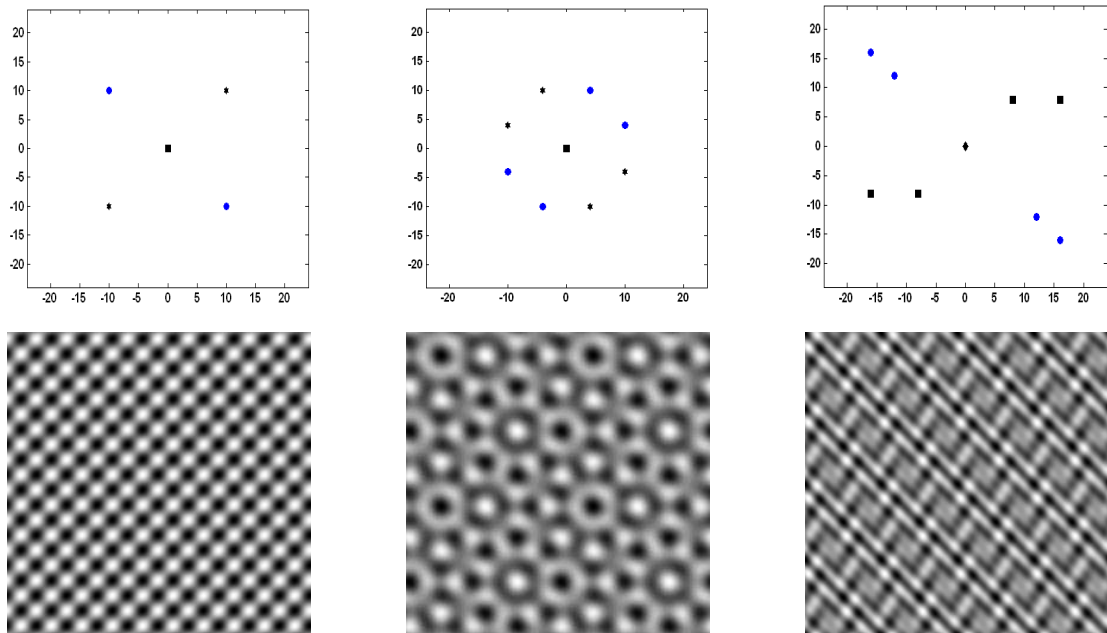


Figure 3.7– Row 1 shows different distribution of frequencies and row 2 shows resultant patterns

Such methods can be extended to placing 2D Gaussians in the power spectrum, which, when coupled with random phase can produce sandwave type surfaces. While the results generated using these methods do not appear very realistic on there own, they can be used for the basis of generating realistic looking surfaces.

C) Generating structural information

Even if a limited number of “natural looking textures” can be generated using power spectrum approaches it is evident that they lack the “structural” characteristics of many man-made and biological (as opposed to mineral) surface textures. Hence we need to generate synthetic textures whereby structural information can be clearly perceived.

Figure 3.8 shows some textures captured using photometric stereo that contain significant phase information.

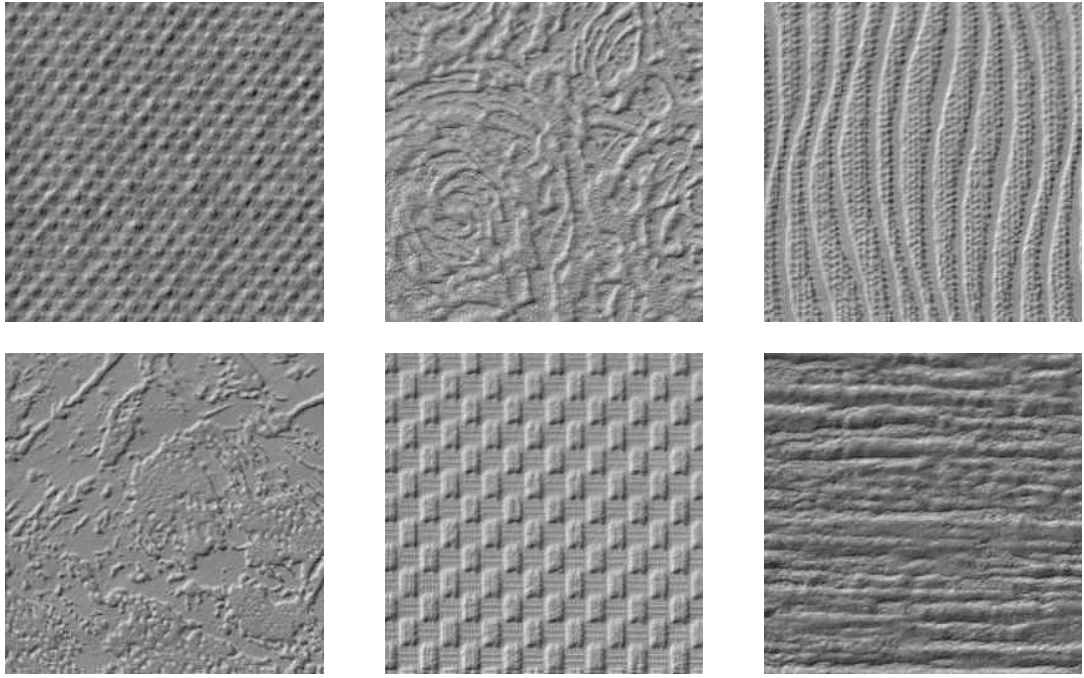


Figure 3.8 – Phase rich textures

We now introduce three techniques that can be applied in combination or individually, that introduce phase information into textures defined in the power spectrum: folding, thresholding and placement.

Folding

The term “folding” is used in this thesis to denote the process that allows a non-linearity to be introduced on surfaces that have been generated based on specific Power law parameters. The folding process operates by first drawing a line/plane across the height distribution of a given surface (or signal) based on an input threshold value. All the values below the folding line are mapped above it (i.e. for all height values lower than the threshold). The mapping is basically a reflection about the folding line. Figure 3.9 illustrates how the folding is performed on 1D and 2D signals. The amount of folding controls the amount of non-linearity introduced in the height map generated. The function used to generate the height maps has been modelled as follows

$$H(x,y)= H(x,y) + 2*(f_{xy}-H(x,y)) \quad \text{if and only if} \quad H(x,y)<f_{xy} \quad (3.4)$$

It is easily implemented using mean shift and absolute functions. After performing folding of the height map using equation (3.4), the dynamic range of the surface is normalised in the range [0, 1].

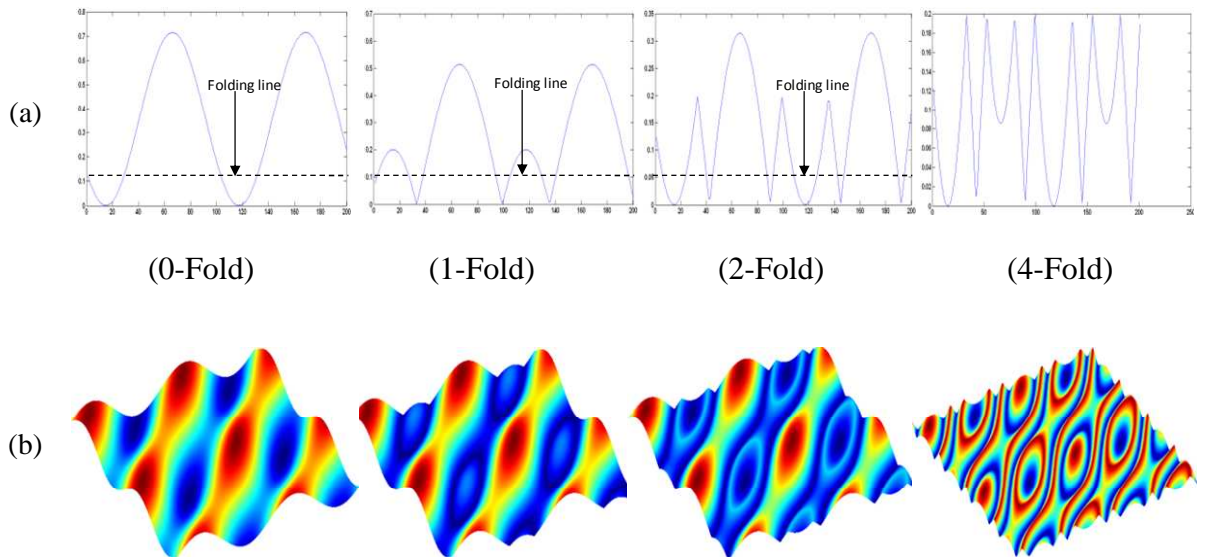


Figure 3.9– (a) 1D signal folding (b) corresponding folding within a 2-D plane representing the same signal

Thresholding

Folded, or unfolded power spectrum surfaces may be used as a probability map to control the placement of simple texture elements. The simplest approach being to threshold such ‘control surfaces’. Random or semi random placement of texture elements then only occurs in areas in which the control surface is greater than the threshold.

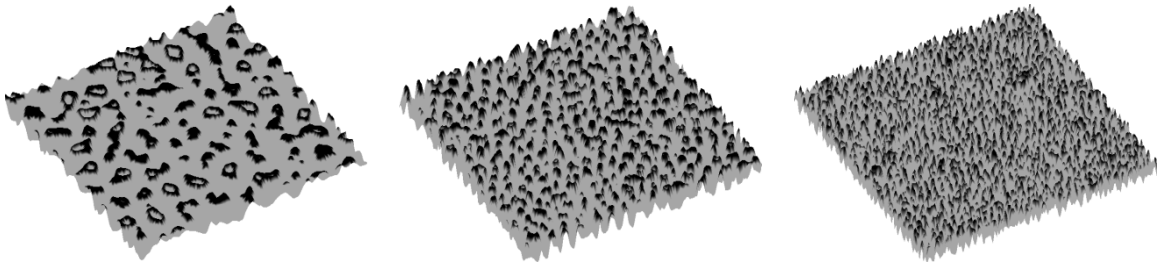


Figure 3.10 – Dark regions identify the placements on surfaces with varied high frequency information

Figure 3.10 shows example surfaces that have been generated by keeping different amounts of high-frequency information (a circular filter is applied to the magnitude information to achieve this). The dark regions on the surfaces represent the placement regions for primitive mapping and are given by retaining height values above a certain cut-off value.

Texture primitive placement

Once placement rules have been determined as described above, they may be used for placing or generating simple texture elements. The texture elements used within this research are either half hemispherical textons to provide for more general structural appearance or otherwise they comprise ellipsoidal textons that provide some directionality. Where primitives overlap we take the maximum of the height of any primitive at that x - y position.

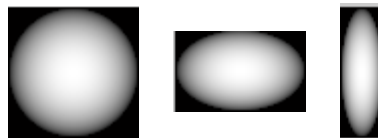


Figure 3.11 – Texture primitives

Figure 3.11 shows some primitives used in generating the synthetic textures. Different sizes of these primitives oriented at different angles are used to provide for diverse varieties of texture surfaces.

D) Summary of structural information generation

The different stages required to create synthetic textures are summarised in Figure 3.12. Figure 3.13 shows height maps of natural textures together with height maps of synthetic textures generated to resemble the natural samples.

Figure 3.14 shows images of the real and synthetic height-maps when illuminated using the lambertian model and with uniform albedo.

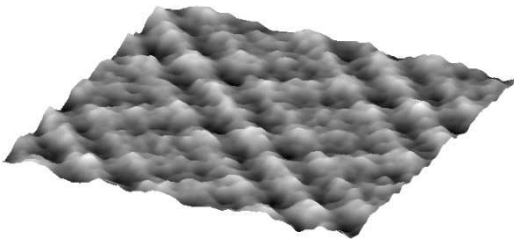
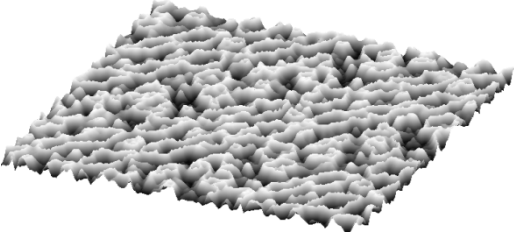
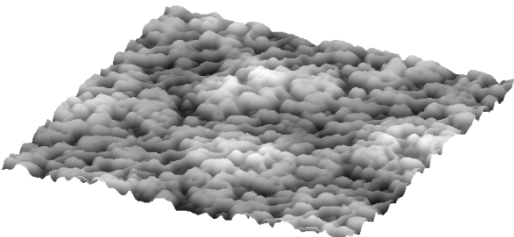
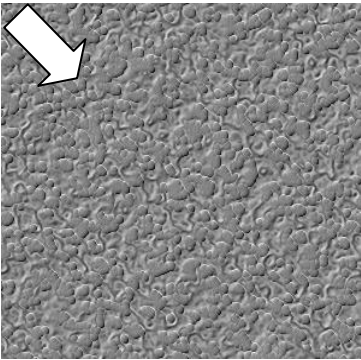
Stage	Height Map	Description
(I)		Original height map is generated either for random placements or regular placements.
(II)		This stage shows the resultant surface after the folding function in equation (3.4) has been applied to the fractal surface generated in stage (I).
(III)		The surface generated after thresholding the folded surface in stage (II) and mapping primitives.
(IV)		Texture image generated by rendering the resultant surface at a slant angle of 45° and a tilt angle of 135° (from the top left corner) and with uniform albedo.

Figure 3.12 – Stages (I) to (III) illustrate the complete process of applying folding, thresholding and primitives mapping to generate a synthetic surface. Stage (IV) represents the illuminated surface.

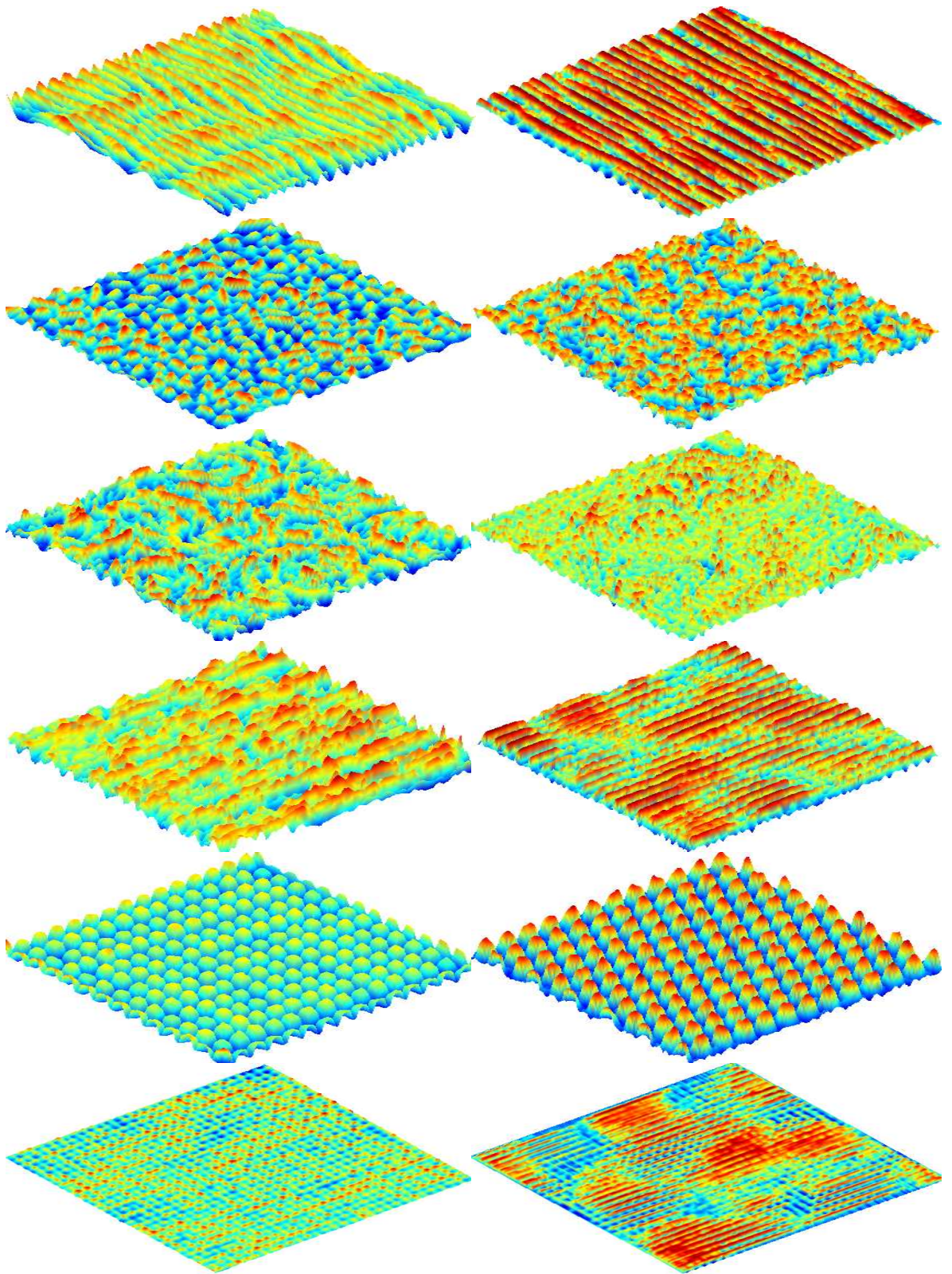


Figure 3.13 – (column1) natural surfaces recovered using photometric stereo, (column2) visually similar synthetic surfaces.

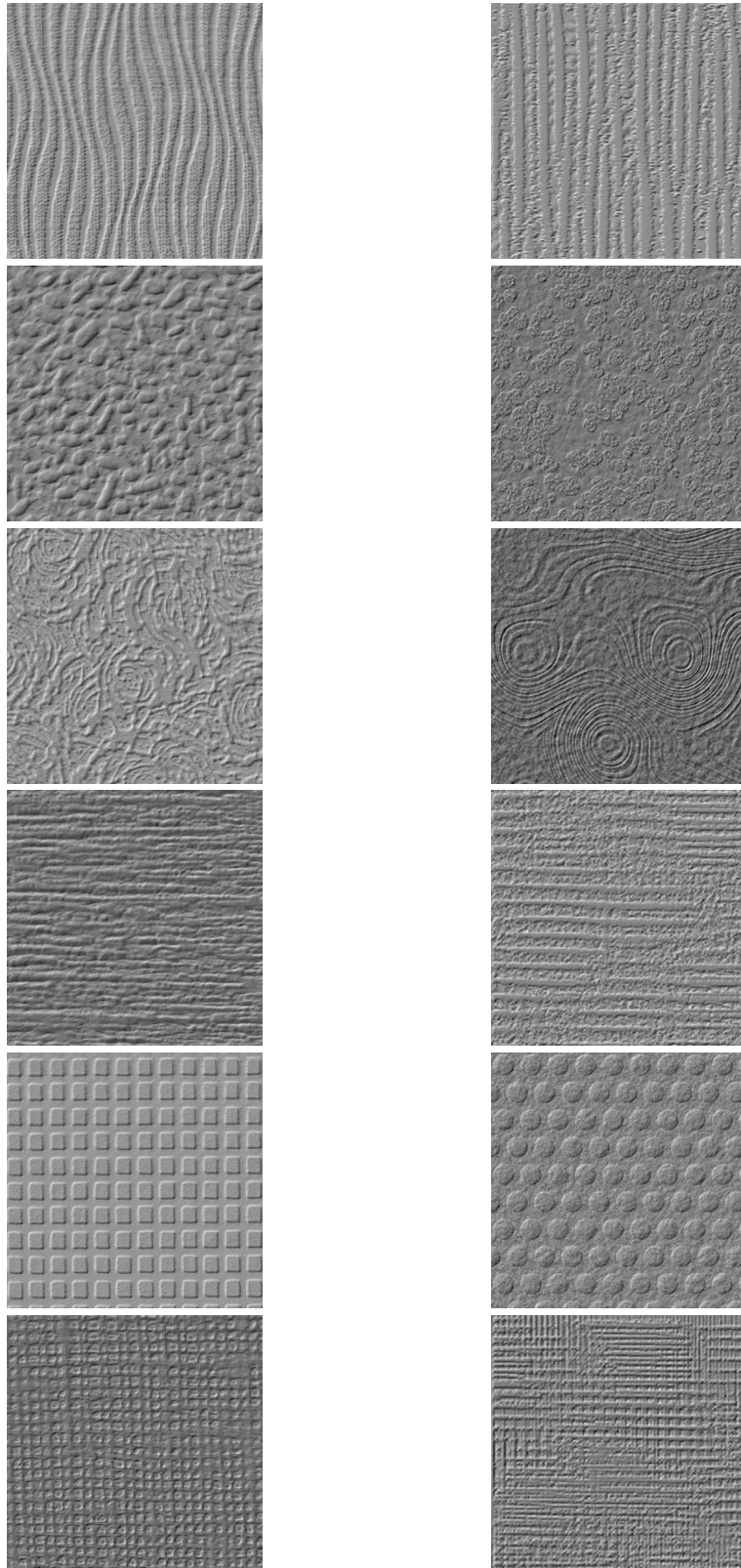


Figure 3.14 -(column1) natural texture images and (column2) visually similar synthetic texture images

3.4.4 Preparation of samples for experiment

The psychophysical studies, performed within the context of this thesis, employ texture samples that are presented to human subjects in the form of photographic prints. Each print occupies an area of 4 by 4 inches and was printed at a resolution of 512 by 512 pixels using a laserjet printer. An example is shown Figure 3.15.

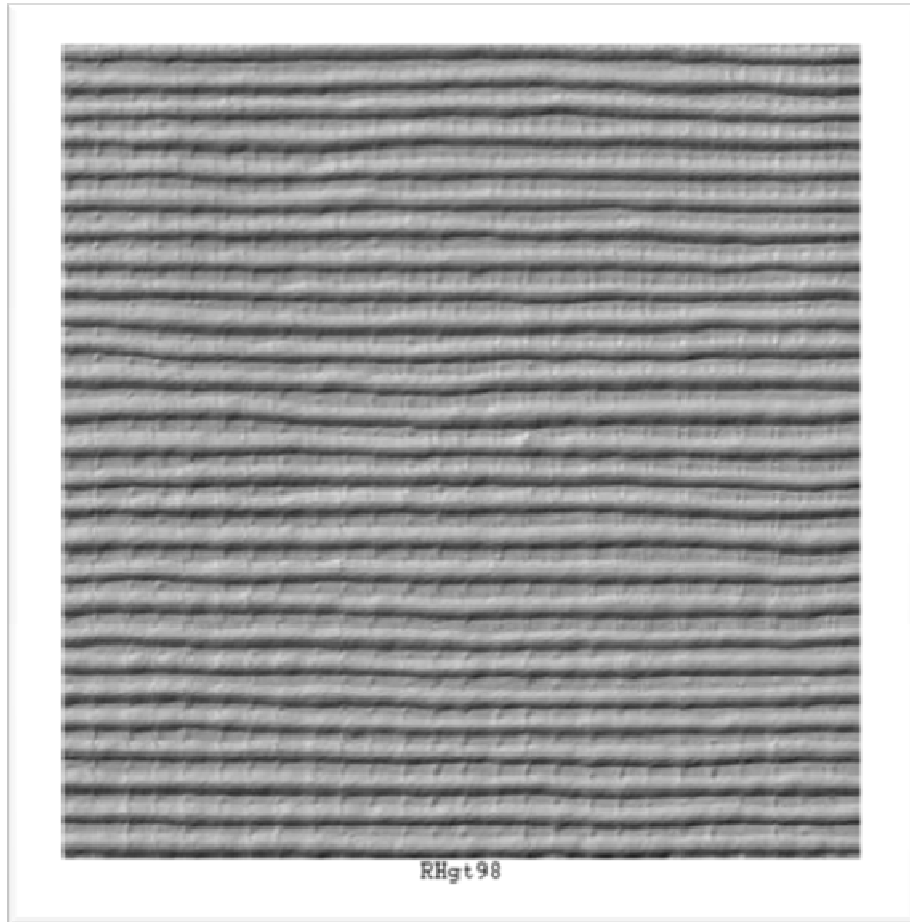


Figure 3.15 – Tex1 sample photograph provided to observers for psychophysical experiment

The use of photographs facilitates the grouping task as it allows observers to have a snapshot of the whole texture dataset thus allowing them to situate themselves within the context of the experiment.

3.5 Experiment Design

One of the fundamental tasks in setting up the psychophysical experiment was to determine the assessment method to be employed by the human subjects to group the textures in both the Tex1 and MoMA datasets. Chapter two summarised the different assessment techniques that have been utilised so far by researchers.

3.5.1 Grouping task

As discussed in Chapter two, perceptual grouping principles have been employed to solve a number of practical vision problems, and the advantages that motivated us to choose this method are summarised below.

- a. Perceptual Grouping has already been used successfully in the field of texture perception to determine perceptual dimensions.
- b. It does not require any complex setup and can be applied to match or compare any size of texture images.
- c. By providing the observer with a view of all the images in the database, it allows the observer to relate to the context in which the experiment is performed.
- d. As compared to other well known techniques such as Pairwise comparison, the user does not need to remember previous judgments as all the images remain in his/her field of vision, whether grouped or ungrouped.
- e. Finally, the technique allows observers to carry out the experiment in 30-40 minutes: thereby preventing undue fatigue.

3.6 Implementation of experiment

The grouping experiment was performed using the two different datasets of texture surfaces, namely the MoMA dataset and the Tex1 dataset. Eight observers were asked to perform the experiment for the Tex1 dataset. This group consisted of both naïve and expert observers. For the MoMA dataset, nineteen subjects were used. In this particular case all nineteen subjects were experts given that they all, in some way or another, had knowledge about the type of samples that they were going to be exposed to.

The same procedure was followed by all the observers to perform the grouping task. The steps are summarised below.

- a) Observers were presented with samples in the form of photographs. The samples were randomly located on a flat surface so that the subjects were able to view all of them at the same time. All photographs were orientated in the same direction, so that the illumination under which they had been captured was consistently portrayed and where possible this was aligned with the room illumination. A table was used in the case of the Tex1 samples and, given that larger photographs were used for the MoMA dataset, these samples were placed on the floor.
- b) The observers were asked not to rotate the photographs but they could pick them up and move them around.
- c) Observers were asked to create as many groups as they wanted by physically moving the photographs into groups.
- d) No group size constraint was imposed on the observers, i.e., they were free to create as many groups as they felt necessary and the groups could be of any size including singletons. They were only requested not to create any “oddball” or “left-over” group, rather that such samples should be left as singletons.
- e) Once the grouping was completed, the observations were registered by the experimenter.

No similarity criteria were imposed on the subjects and they were free to take as long as they required to perform the grouping task. Appendix B shows the instruction sheet that was provided to subjects.

3.6.1 Comments collected on the experiment

After the grouping experiment was performed, observers were asked to provide feedback on the criteria they used to group the textures presented to them. Their remarks are summarised below.

A) Tex1 grouping

The expert subjects who performed the experiment looked for some perceptual cues from the texture samples. Perceptual attributes such as directionality and regularity of texture elements were considered. Additionally, the subjects identified textures with respect to the type of texture primitives.

Naïve subjects were also influenced by the structural information of the texture samples. However, their main criterion of grouping the textures was the type of material that the

samples represented, such as fabrics, wood, rock, wallpapers and consequently their application like wallpapers for kitchen walls.

The eight subjects took on average 35 minutes to group the 120 texture photographs presented to them. An average of 22 groups was created by the 11 subjects, with a minimum size of 8 and a maximum of 35.

B) MoMA grouping

Grouping for the MoMA textures were performed by subjects who have substantial knowledge on the textures and the type of material being considered. The subjects are mainly conservators of photographs, with some of them being paper, sculpture and painting conservators. Since the MoMA textures are mainly photographic papers, the first criterion used by the subjects was the effect of using the papers for photographs, printing or painting purposes. Another equally important criterion employed by some of the subjects was the source of the photographic paper itself.

Subjects could identify different papers with respect to their knowledge about different manufacturers such as Kodak, Agfa, Unicolor and others. The 19 subjects for the MoMA grouping experiment took on average 50 minutes to create texture groups from 81 samples. The size of the groups created ranged from 10 to 43, with an average of 24 groups being created by the subjects.

3.7 Aggregate Data (generate similarity matrices)

The output of the experiment was a set of groups created by each observer. These data need to be aggregated so that they could be used:

- to check the consistency of the experiment,
- to discover any structural information in the data to provide insights, or methods for the generation of a retrieval system, and
- to determine ‘ideal’ retrievals for performance analysis.

The procedure was as follows:

- $N \times N$ occurrence matrices were created for each observer. N is the number of samples in a dataset: for MoMA N is 81 where as for the Tex1 N dataset is 120. The occurrence matrix, as illustrated in Figure 3.16, is a binary matrix whereby a ‘1’ represents whether any texture T_i is grouped together with any other texture T_j .
- Two similarity matrices, \mathbf{S} , were created (one for each dataset) by aggregating all of the occurrence matrices for each dataset. A cropped representation of the similarity matrix for Tex1 textures is presented in Appendix C.
- The similarity matrices were normalised in the range 0 to 1 so that they are independent of the number of subjects.

	T_1		T_i		T_N
T_1					
T_j			1		
T_N					

Figure 3.16– Occurrence Matrix

The values composing \mathbf{S} , are referred to as the similarity coefficients with the coefficient $S(i,j)$ indicating the similarity of texture samples, T_i and T_j . The higher the value of $S(i,j)$ the more alike the textures T_i and T_j are judged to be. The values of the matrix can be summarised by the following conditions:

$$(S1) \quad 0 \leq S(i,j) \leq 1$$

$$(S2) \quad S(i,i) = 1$$

$$(S3) \quad S(i,j) = S(j,i)$$

3.8 Conclusions

Chapter three has presented the procedure by which similarity matrices have been produced for two collections of surface textures:

Tex1 - a general dataset containing 120 samples drawn from a wide variety of texture types, for which the height maps have been obtained or generated, and

MoMA - a specialised collection of 81 samples of photographic papers collected under consistent illumination conditions

To the author's knowledge this is the first time that such data have been produced using homogeneous textures that were collected under known and consistent conditions.

Chapter 4

Psychophysical Data Analysis

4.1 Introduction

The objective of this thesis is to produce methods for developing perceptually relevant surface texture retrieval systems. The previous chapter used psychophysical experiments to derive similarity matrices for two surface texture datasets: Tex1 and MoMA. In their current form these data are represented in high dimensional spaces that are unsuited for either visual inspection, or the practicable computation of appropriate feature sets. Indeed, without a method for inspecting these data it is difficult to determine whether or not the psychophysics has produced non-random results.

Thus the objectives of this chapter are:

- i) to inspect the similarity data for evidence of structure that indicates that the psychophysical experiments have produced non-random results;
- ii) to inspect these data for evidence of natural groupings;
- iii) to investigate the number of dimensions that these data can be represented in; and
- iv) to examine the major dimensions of the data for obvious traits that could be useful for the design of retrieval systems.

Chapter two identified Hierarchical Cluster Analysis (HCA) and Multidimensional Scaling (MDS) as suitable tools to be used for analysis purposes. Section 4.2 describes the format in which the similarity matrices are use by HCA and MDS. In Section 4.3 we use HCA to investigate issues (i) and (ii). Section 4.4 uses MDS to investigate the dimensionality of the datasets and Section 4.5 uses both techniques to look for obvious traits in the main dimensions of the data.

4.2 Data representation

HCA requires that the input data be presented in the form of an N by N matrix of pairwise dissimilarities, $\mathbf{D} \equiv (d_{ij})$, where each element d_{ij} represents the dissimilarity between the i^{th} and the j^{th} textures and N is the number of textures available in the dataset. Any dissimilarity value d_{ij} needs to satisfy certain minimum conditions as follows: $d_{ij} \geq 0$, $d_{ii} = 0$ and $d_{ij} = d_{ji}$.

MDS requires ‘proximity’ data that defines the ‘nearness’ in space of a pair of textures. Ideally the proximity measure should be a distance measure with a value close to zero representing textures with similar characteristics and vice versa for large values. Thus, since MDS is to be applied to our perceptual data it is again necessary to use dissimilarity matrices \mathbf{D} derived from psychophysical data. However for presentation purposes (e.g. in dendrograms) we often scale these data by the number of observers that have taken part in the experiment for ease of interpretation.

The similarity values are converted to dissimilarities using the transformation $d_{ij} = 1 - S_{ij}$. Similarity coefficients from all the psychophysical results presented in this thesis are scaled in the range (0, 1) and this range is preserved for dissimilarity coefficients when the transformation is applied.

When dealing with psychophysical data, the assumption that dissimilarities behave as distances is not valid. For dissimilarities to be considered as metric data, they should satisfy the following conditions: (1) $d_{ij} = 0$ if and only if $i=j$, (2) $d_{ij} = d_{ji}$ for all $1 \leq i, j \leq N$, and (3) $d_{ij} \leq d_{ik} + d_{kj}$ for all $1 \leq i, j, k \leq N$. However, the dissimilarities do not satisfy condition (3) which is the transitivity condition for metric data. Thus, the dissimilarity matrix used within the context of this thesis is non-metric.

4.3 Cluster Analysis

Results from hierarchical clustering algorithms are generally presented in the form of tree diagrams. The dendrogram is the cluster analysis tool that is used to investigate the clustering tendency for the Tex1 and MoMA datasets within this thesis and is presented in Section 4.3.1.

To analyse the clustering represented by the dendrograms, we can either partition the dendrograms at a given dissimilarity level or otherwise partition them with respect to the number of groups required. Section 4.3.2 and Section 4.3.3 analyse the groups formed when the dendrograms for the Tex1 and the MoMA datasets are partitioned into six groups.

4.3.1 Dendrograms

The use of dendrograms allows us to easily partition the dataset into as many clusters as we would like and to visualise the relationships (similarities) between these groups. This allows us to qualitatively assess the consistency of the results of the psychophysical experiments and to check that there is apparent structure within the data. The only disadvantage of using dendrograms that should be noted here is that they are normally crisp approaches to clustering (i.e. texture can be allocated to only one group at a time), however since our objective is purely visual inspection for consistency, rather than discrimination between groups, it is not an issue here [Baker75 & Ding02].

Together with displaying grouping information, dendrograms have an additional property: the height information shows level of similarity or dissimilarity between clusters.

Definition: A dendrogram is an n -tree on a set of objects $\Omega \equiv \{O_1, O_2 \dots O_n\}$ and is given by a set T of subsets of Ω satisfying the following conditions [Gordon87]:

- I. $\Omega \in T, \emptyset \notin T, \{O_i\} \in T$ for all $\{O_i\} \in \Omega$
- II. $A \cap B \in \{\emptyset, A, B\}$ for all $A, B \in T$
- III. $A \cap B \neq \emptyset, h(A) \leq h(B) \leftrightarrow A \subseteq B$ for all $A, B \in T$
- IV. $h_{ij} \leq \max(h_{ik}, h_{jk})$ for all $O_i, O_j, O_k \in \Omega$

Conditions I and II specify a hierarchically nested set of subsets, with each subset representing a class of similar samples. Condition III provides information for the height h associated with each internal node such that for each pair of objects (O_i, O_j) , h_{ij} represents the height of the internal node specifying the smallest class to which objects O_i and O_j belong to.

The height of the internal node represents how similar the objects within the groups are. Thus, the smaller the value of h_{ij} , the more similar objects O_i and O_j are. Condition IV is an ‘ultrametric’ condition implying that the height of any class to which two objects O_i and O_j belongs to is less than the height of the same objects associated with any other object O_k not in the class.

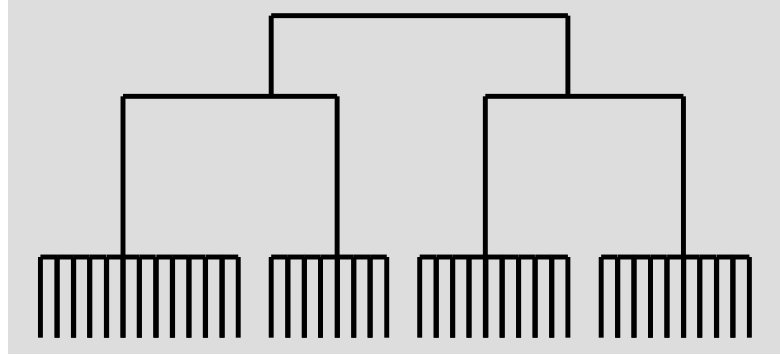


Figure 4.1- Ideal dendrogram

If the dissimilarity data represents a number of clearly defined groups then the dendrogram should provide clear indication of this as illustrated in Figure 4.1. Note however, that Tex1 samples were chosen in order to maximise the apparent variability within this dataset.

In order to construct the dendrogram from the dissimilarity matrices, agglomerative clustering has been applied. Agglomerative clustering uses a bottom-up approach to create clusters. Starting from N singletons, a recursive procedure is used to merge a pair of clusters C_i and C_j based on a pairwise linkage function [Ding02].

As the data is non-metric, only a selected number of linkage functions can be used to create the dendrogram. The most common of these are the single linkage, the complete linkage and the group average linkage functions. Based on a general definition provided in [Gordon87], these three functions can be defined as follows:

$$d(C_i \cup C_j, C_k) = \alpha_i d(C_i, C_k) + \alpha_j d(C_j, C_k) + \gamma |d(C_i, C_k) - d(C_j, C_k)| \quad (4.1)$$

where $d(C_i, C_j)$ represents the dissimilarity between any two classes C_i, C_j , α and γ are parameters whose values identify the clustering strategies. Table 4.1 shows the parameter values for each clustering algorithm. w_i represents the number of samples that a particular class C_i contains

Linkage function	α_i	γ
Single link	$1/2$	$-1/2$
Complete link	$1/2$	$1/2$
Group average link	$\frac{w_i}{(w_i + w_j)}$	0

Table 4.1- Linkage functions and associated parameter values

The linkage functions differ in the way they characterise the similarity between a pair of clusters (or singletons). For single linkage, the distance between two clusters is given by the minimum distance between all pairs of textures drawn from two clusters where as for complete linkage it is the maximum. Group average linkage uses the distance between the centroids of a pair of clusters. The complete linkage function produces tightly bound and compact clusters [Jain99] as compared to the single and group average linkage and has also been successfully used by Rao *et al.* [Rao93a] to generate dendrograms for psychophysical data. Thus a complete linkage function is used in this thesis.

4.3.2 Analysing the Tex1 Dataset

The dendrogram illustrated in Figure 4.2 shows how the texture samples from the Tex1 dataset are clustered into different groups. The leaves from the dendrogram represent the individual textures. As we move up the dendrogram, groups of textures are formed based on the data available from the dissimilarity matrix. The following observations can be made.

- (a) The topology of the dendrogram shows no distinct set of dominant groupings – indicating that the samples are reasonably well distributed across the range of textures considered.
- (b) The dendrogram ‘breaks’ into a significantly large number of small groups below a dissimilarity level of 5 indicating that the number of groups created by the average number of subjects is quite large.

For visual inspection purposes, a clustering containing the six main perceptual groups has been obtained by placing a cutting line just below dissimilarity level 6 (Figure 4.2).

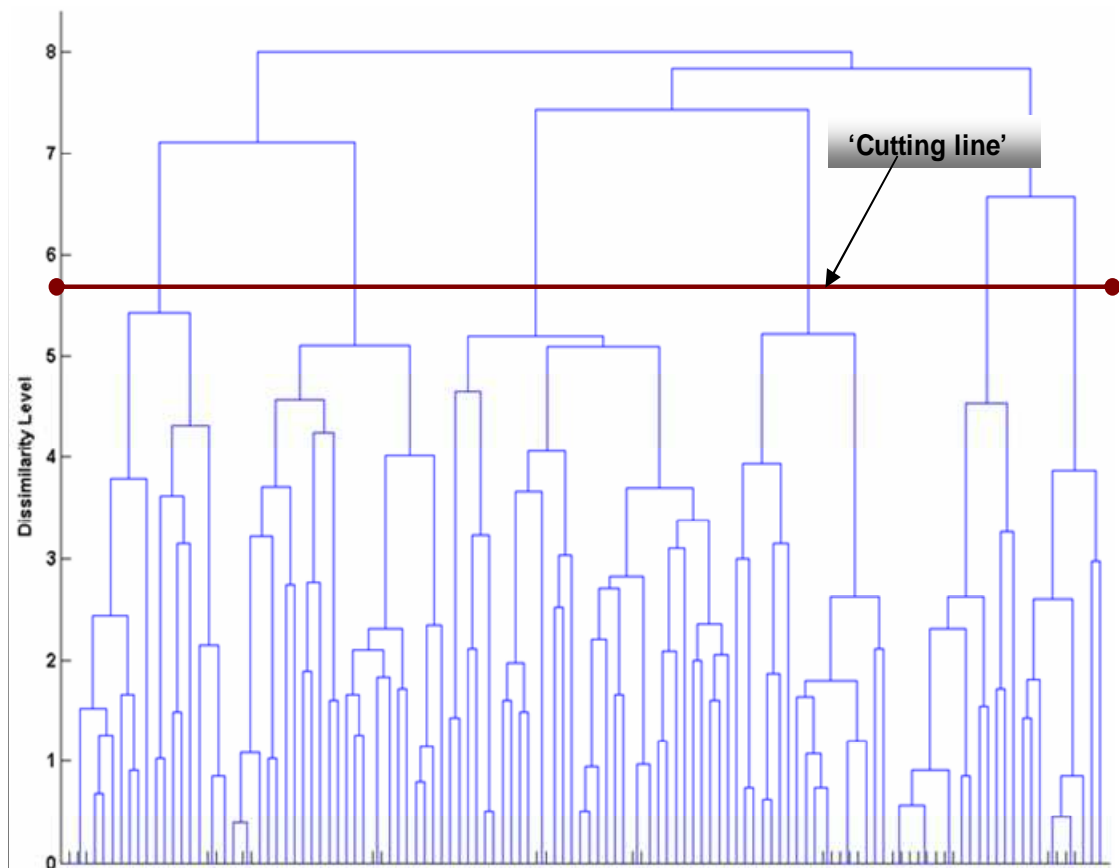


Figure 4.2- Complete dendrogram with leaves being texture samples used in the psychophysical experiment. Dissimilarity level units correspond to number of observers.

The line crosses the subtrees of the dendrogram where a considerable number of subjects have agreed on the groups formed (indicated by the significantly larger heights as compared to the rest of the dendrogram). Figure 4.3 illustrates the reduced dendrogram.

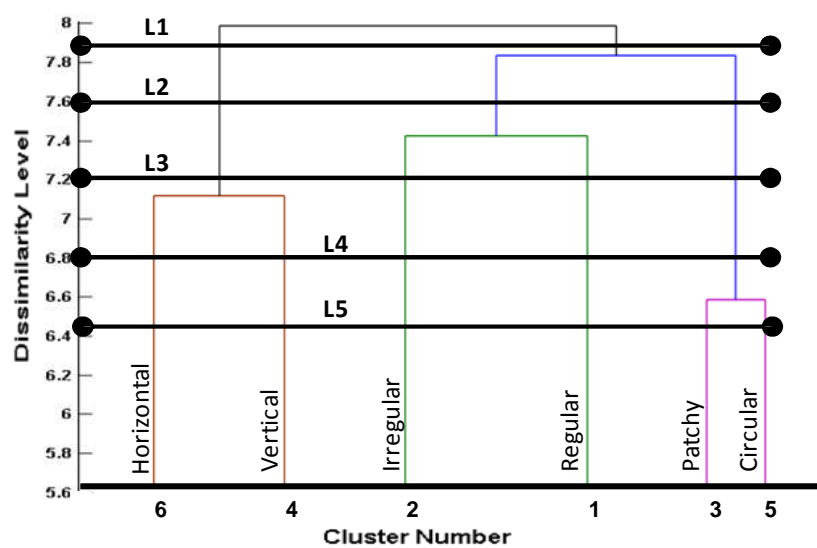


Figure 4.3- Dendrogram showing the dissimilarity level when 6 'major' clusters are considered

Randomly chosen samples from each cluster are shown in Figure 4.4. Each column represents a different group. The representative samples shown are chosen randomly from the clusters created, with some clusters containing a considerably larger number of textures than others. The complete groups are provided in Appendix D (see Figure D.1 to Figure D.6). Moving from left to right in Figure 4.4 leads to the following observations.

- (1) The internal members from each group show perceptually similar characteristics, i.e. they are visually consistent. It seems unlikely that these groups could have been created at random and therefore it is probable that the psychophysical experiments have produced meaningful results.
- (2) The six main groups displayed in Figure 4.4 show that the Tex1 dataset comprises mainly of horizontal¹ textures (group 6), vertical textures (group 4), regular textures (group 2), irregular textures (group 1), patchy textures (group 3) and circular textures (group 5).
- (3) Elements from various groups appear to be distinctively different even though some overlaps may exist between the groups. For example overlaps exist between the regular and irregular groups or between the patchy and circular groups.
- (4) Reducing the six groups to create larger texture sets imply that we would need to merge groups 3 and 5 together, then group 6 and 4 and groups 1 and 2.

The dendrogram in Figure 4.3 is “parsed” to investigate whether there exists any kind of visual consistency when the groups from Figure 4.4 are merged together to create the whole Tex1 dataset. To do so several cutting lines are applied to the reduced dendrogram at levels where a split occurs. The cutting lines are referred to as levels L1 to L5 and would help to visually assess how the six perceptual groups merge to form the complete Tex1 dataset.

Level L5

This level shows the six groups discussed in observation (4). In addition to height which shows the level of observer disagreement, the separation of the groups along the cutting lines also indicates how cohesive pairs of groups are and their tendency to merge. Thus the patchy textures (group 3) and circular textures (group 5) are most similar to each other and provide the first pair of candidates for merging if required.

¹ Note that the names of these groups have been chosen by the author purely to facilitate discussion – they have not been identified using psychophysical experiments.

Level L4

At this level the patchy and circular textures have merged. Visual inspection of the samples of the two groups indicates the presence of global information such as large and randomly placed patches from the samples in the patchy group and large and randomly placed circular structures for the textures in the circular group.

Level L3

Group 6 (horizontal textures) and group 4 (vertical textures) merge at this level to create one main group of unidirectional textures.

Level L2

Line L2 from Figure 4.3 shows that the whole Tex1 dataset can be represented into only three groups with groups 1 (regular) and 2 (irregular) merging to form one larger group of textures. Visually inspecting the pairs of groups formed at this level, we can summarise the dataset into the following categories: structured (regular and irregular), unstructured (patchy and circular) and unidirectional (vertical and horizontal).

Level L1

The two supergroups created by merging groups 1 and 2 and groups 3 and 5, merge to create an even larger group. The resultant group consists of the structured and unstructured textures as presented at Level L2. We can observe that the merging occur at a relatively high level of dissimilarity.

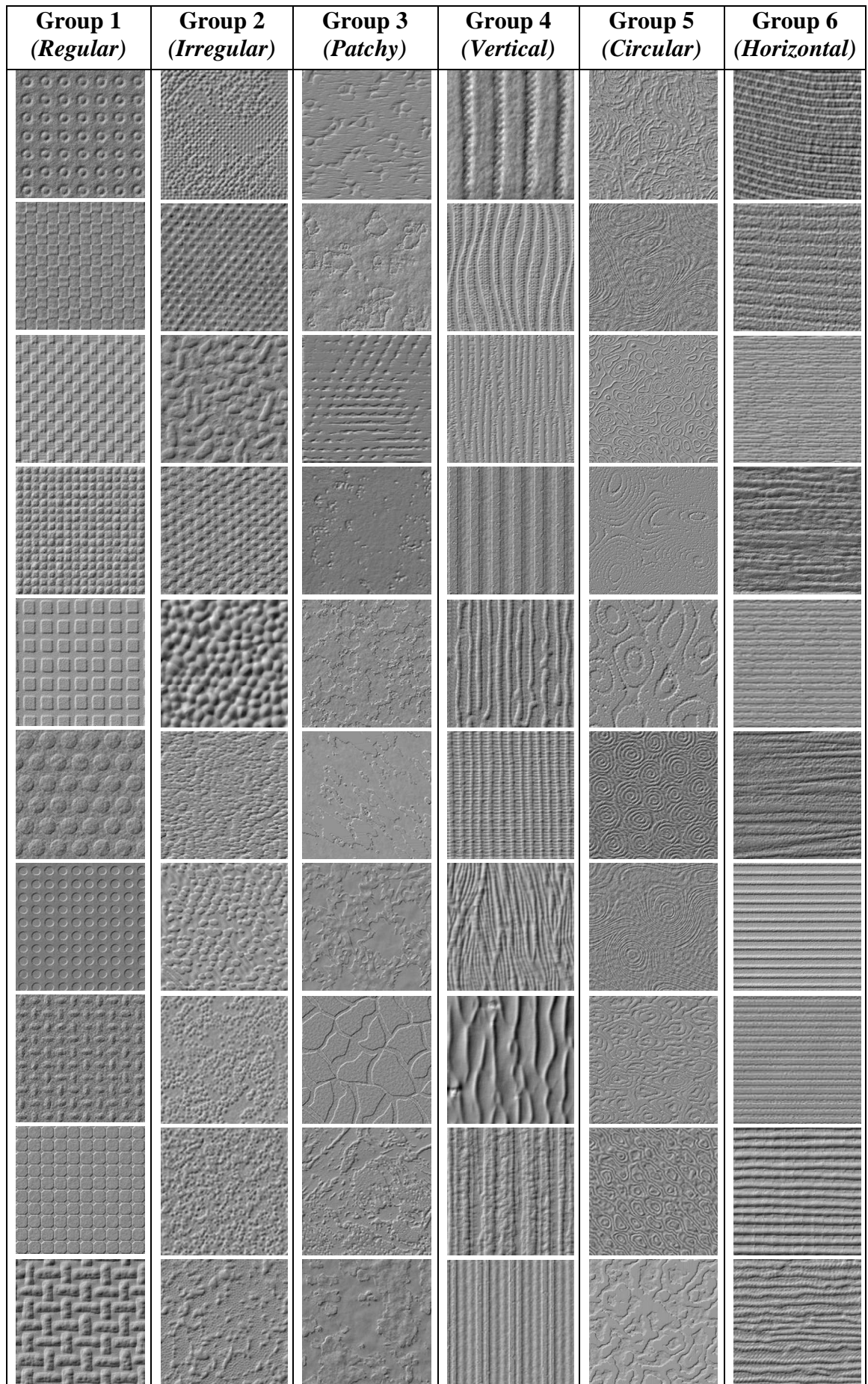


Figure 4.4 - Groups 1 to 6 from dendrogram represented by columns 1-6

4.3.3 Analysing the MoMA Dataset

As described in Chapter three, compared with Tex1, the MoMA database is more specialised, and moreover the number of subjects who participated in the grouping experiment was higher (recall from Chapter 3: 19 subjects). Figure 4.5 shows the dendrogram obtained when a complete linkage clustering is performed on its dissimilarity matrix.

Analysing the dendrogram in Figure 4.5 in a top-down manner (i.e. starting from the root) shows that five out of the nineteen observers could not decide how the whole MoMA dataset could be split into more than two groups. This is indicated by the level of indecision in the dendrogram of Figure 4.5. At level 1 all 19 observers agreed that all the MoMA textures could not be placed in only one group whereas level 2 shows that approximately 14 agreed that the MoMA textures could be placed into more than two groups.

As compared to the dendrogram for MoMA dataset, the one the Tex1 dataset (shown in Figure 4.2) splits very quickly into two or more branches indicating that the observers could easily perceive the variability among the textures in the dataset. These observations from the two dendrograms (in Figure 4.2 and Figure 4.5) provide further indication that the MoMA dataset is more compact than the Tex1 dataset.

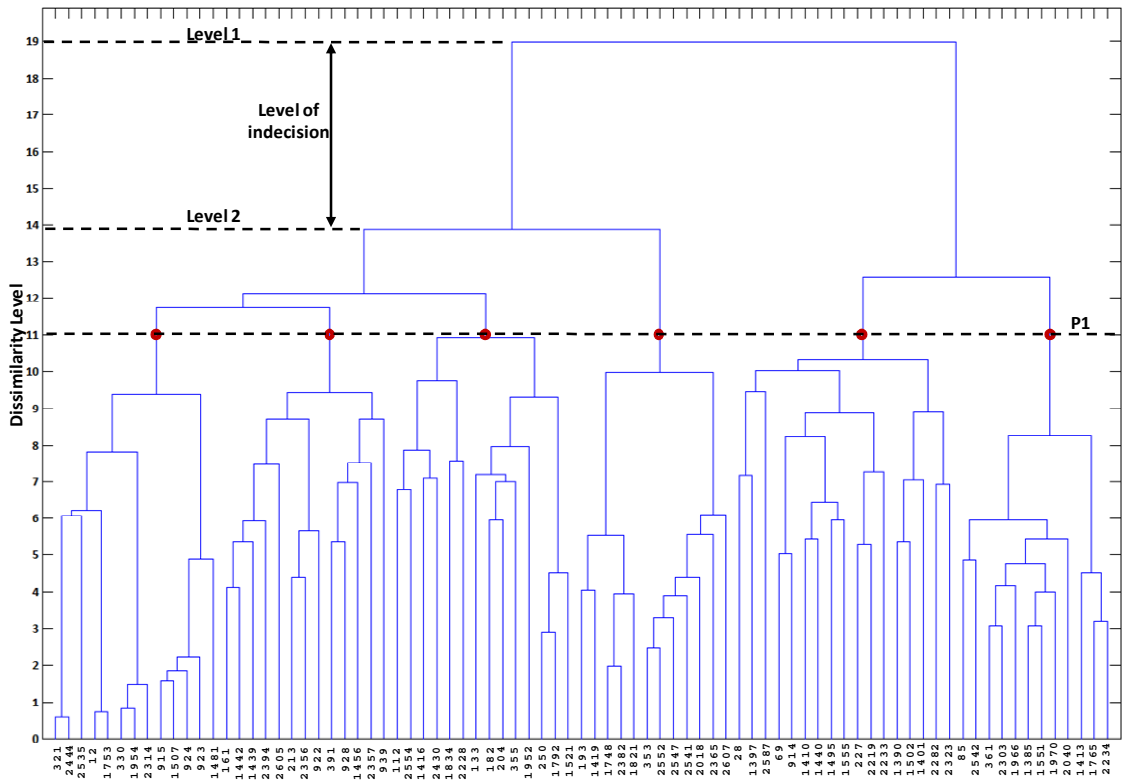


Figure 4.5 – Dendrogram representing linkage between MoMA textures to form groups

A cutting line (denoted as P1 in Figure 4.6) shows the level of dissimilarity at which the dendrogram for the MoMA dataset could be split to obtain six main groups. As indicated by P1 only eight out of the nineteen subjects agreed that MoMA textures could be placed into six groups. The remaining eleven could find enough variability in the textures in order to split them into smaller groups.

Figure 4.6 shows the reduced dendrogram after applying cutting line P1. The six groups obtained are again used to investigate the visual consistency of the data. Randomly chosen textures for the six groups are shown in Figure 4.7. With the MoMA texture representing only photographic paper surfaces, any distinction among the six groups is more subtle. From Figure 4.7 only group 1 appears to contain textures with regularly arranged elements, whereas the other groups appear to differ based on the roughness of the surfaces.

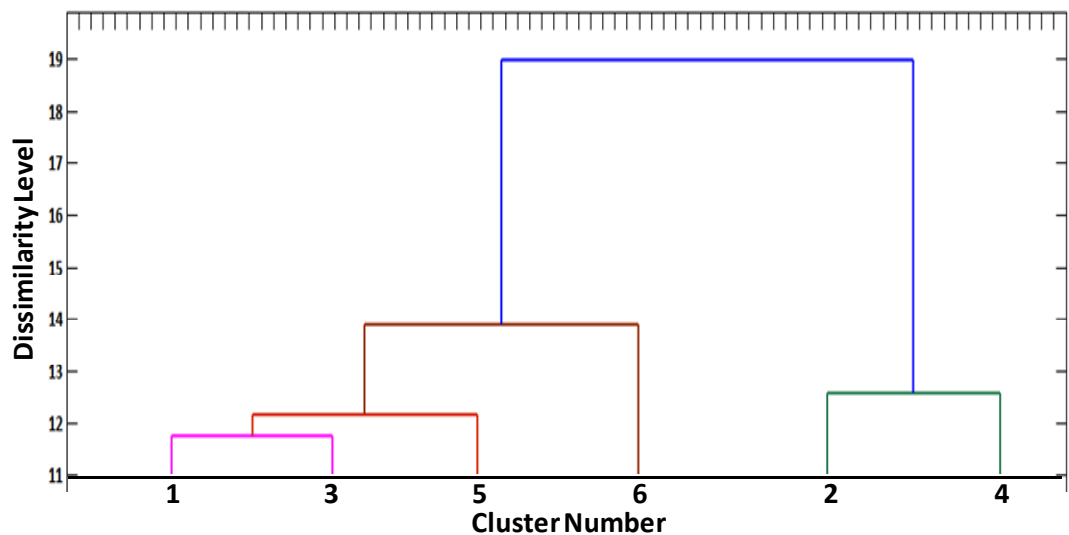


Figure 4.6– Reduced dendrogram for MoMA textures to show six main groups

The complete groups for the MoMA dataset are provided in Appendix D (see Figure D.7 to Figure D.12)

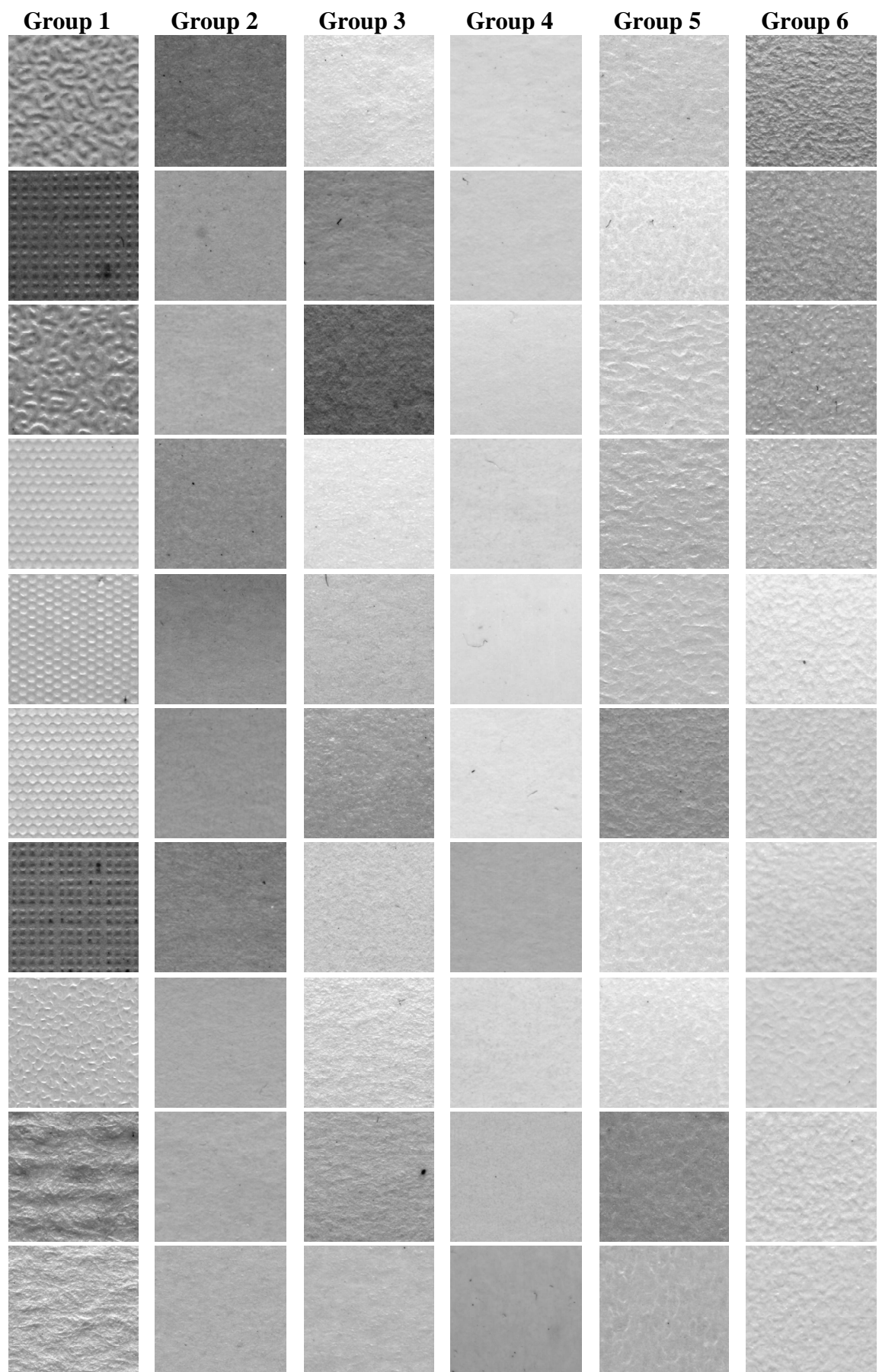


Figure 4.7– Groups 1 to 6 for MoMA database

4.4 Dimensional Analysis

The dendrograms presented in the previous section provide an effective way of visualising the psychophysical data and show that they do appear to contain meaningful (or at least non-random) data. However, they provide few insights into the number of dimensions that these data can or should be represented in. Such a property facilitates practicable computational approaches to coding, classification, segmentation, and the subject of this thesis: retrieval. Hence in this section we apply Multidimensional Scaling to the dissimilarity matrices in order:

- (a) to determine whether there is evidence that our perceptions of surface texture can be encoded into a particular (and low) number of dimensions;
- (b) to investigate the effect that reducing the number of dimensions has on the variability encoded in the data; and
- (c) to investigate whether or not any of the major dimensions derived have an obvious interpretation.

Recall that the data is purely ordinal and thus non-metric MDS must be used. This can be considered as a two-fold optimisation process that first finds an optimal monotonic transformation of the dissimilarity data and secondly derives an optimal configuration to represent the data such that the dissimilarities between the points in that configuration match as closely as possible the scaled dissimilarity values.

Two metrics are commonly used to assess how well different configurations produced by MDS represent the original data: Alignment Error and Stress. We use these measures to investigate issues (a) and (b) above, while visual inspection is used to address (c).

4.4.1 Alignment error

The “alignment” of the two matrices is performed using a technique called Procrustes analysis. The latter is normally used to compare MDS results from two different configurations. However, within the context of this thesis it is employed to compare the full and reduced representations. The Procrustes analysis dilates, translates, reflects and rotates the distances from the chosen configuration in order to match the dissimilarity values [Cox00]. Alignment error is thus the sum of squared errors that result while performing those linear transformations and is computed as follows

$$\text{Alignment error} = \sum_{r=1}^N (y_r - x_r)^T (y_r - x_r) \quad (4.2)$$

y_r and x_r in the equation above are the off-diagonal elements from the Procrustes and dissimilarity matrices. We use the term “Procrustes matrix” to represent the distances between texture pairs in the reduced dimensional space which is subject to Procrustes analysis. With a set of affine transformations performed on the Procrustes matrix, the mapped distances are given by:

$$x'_r = \rho \mathbf{A}^T x_r + \mathbf{b} \quad (4.3)$$

where ρ is a scaling coefficient, the matrix \mathbf{A} accounts for rotation and \mathbf{b} is a rigid translation vector. Taking into account those transformations, the alignment error is now given by:

$$\text{Alignment error} = \sum_{r=1}^N (y_r - x'_r)^T (y_r - x'_r) \quad (4.4)$$

Thus the alignment error is the error that remains after the rotation (\mathbf{A}) and translation (\mathbf{b}) transformations have been applied to the Procrustes matrix.

Alignment error – results

Figure 4.8 shows the alignment error between the full and reduced perceptual spaces as a function of the number of dimensions (of the reduced perceptual space) for both the Tex1 and MoMA datasets.

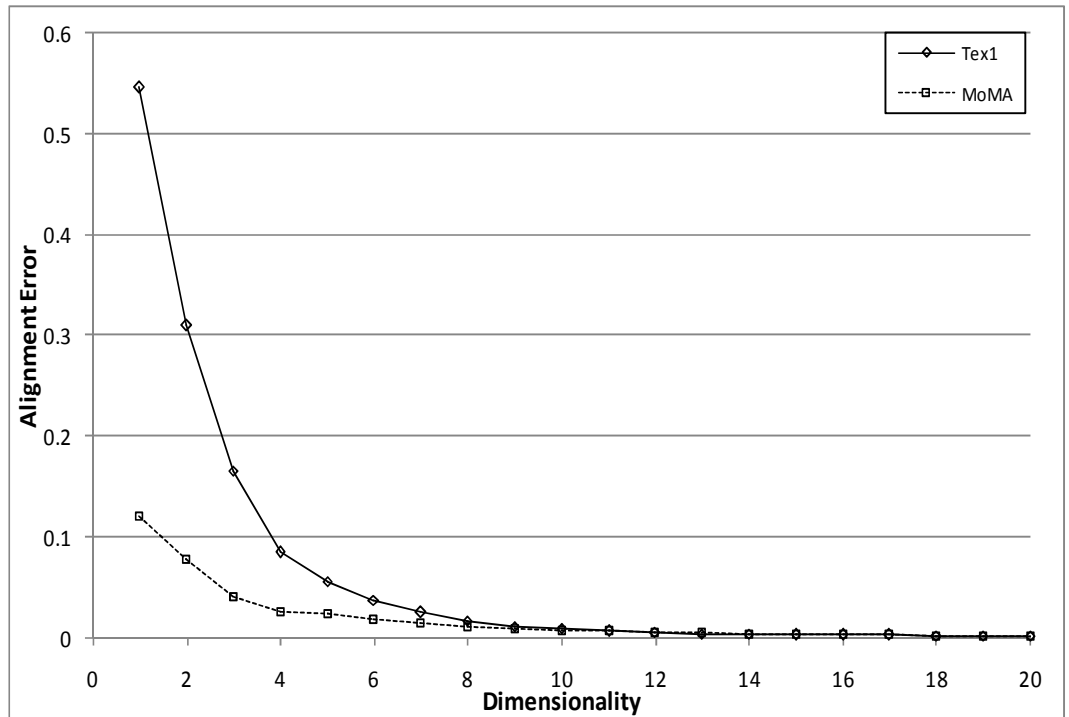


Figure 4.8– Graph showing how alignment error decreases as dimensionality increases

The decline in the alignment error of the two datasets is relatively smooth from four dimensions onwards and there are no obvious breakpoints at which to curtail the dimensionality for use in practical applications. It is noticeable that at lower dimensionalities the alignment errors of the MoMA spaces are less than those of Tex1, indicating that the former can potentially be represented using fewer dimensions. In addition we see that a dimensionality of ten is sufficient to encode most of the original information for either dataset. Ten dimensions however, is still a relatively unwieldy number for either visualization or for the automatic development of retrieval systems.

4.4.2 Stress

Stress is the term coined by Kruskal [Kruskal64b], to denote the loss function used to minimise nonmetric MDS models. Stress, S , is defined as

$$\text{Stress} = \sqrt{\frac{S^*}{T^*}} \quad (4.5)$$

S^* is called the raw stress of the configuration tested and T^* is a normalising factor that allows the stress value to be dimension free. Both terms are defined as follows

$$S^* = \sum_{r,s} (d_{rs} - \hat{d}_{rs})^2 \quad (4.6)$$

$$T^* = \sum_{r,s} d_{rs}^2 \quad (4.7)$$

\hat{d}_{rs} represents the dissimilarity values defined on an N by N dissimilarity matrix such that the mapping is always monotonic where as d_{rs} represents the distances computed from points in the spatial configuration being considered. Since its conception by Kruskal, stress has been widely used as a measure for the goodness of fit of a chosen configuration. Guidelines to judge the goodness of fit are summarised in Table 4.2 [Kruskal64a].

Stress value	Goodness measure
Above 0.20	Poor
0.10	Fair
0.05	Good
0.025	Excellent
0.0	Perfect

Table 4.2– Stress values with corresponding goodness of fit interpretation

Stress analysis – results

Candidates for the number of dimensions, d , suitable for representing a perceptual space are most commonly determined by identifying “elbows” in Scree plots (see Figure 4.9). A Scree graph is a plot of stress values against dimensions.

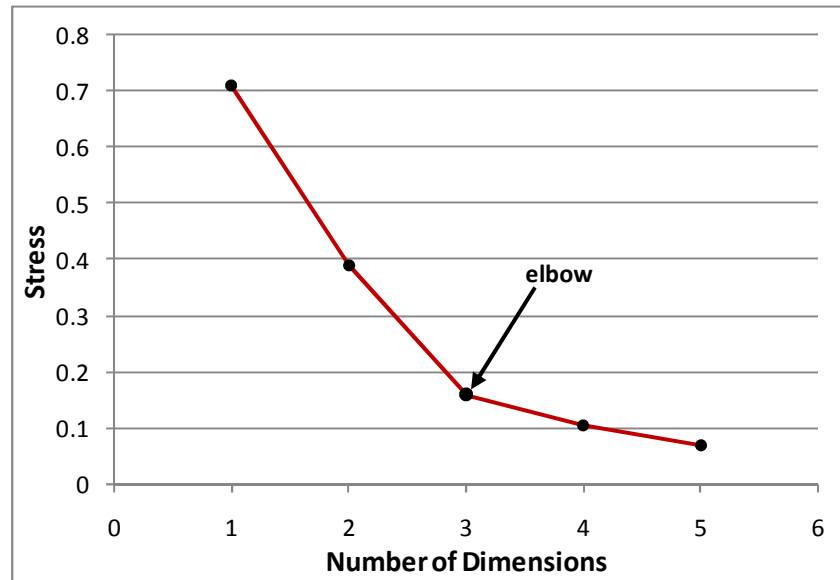


Figure 4.9 - Scree plot showing “elbow” effect

The ideal “elbow” is a sharp drop in error values followed by relatively small decrements. For instance if the data were derived from colour experiments, we would expect a significant ‘elbow’ at $d = 3$. The identification of an “elbow” is accomplished by visual inspection of the Scree plot.

To determine the number of dimensions for the MoMA and the Tex1 datasets, only the first twenty dimensions are considered. Figure 4.10 illustrates the Scree plots for both datasets. The behaviours of the stress values, for both datasets, show no significant change in the pattern of the goodness of fit measures. This observation confirms the behaviour of the alignment errors as presented in Figure 4.8 and strengthens the case that there is no clear cut dimensionality that can be chosen to represent the MoMA and the Tex1 datasets.

However, Figure 4.10 shows that from dimensionality four onwards the stress values are less than 0.1 – indicating a “good” fit in the case of Tex1. For MoMA, the stress values indicate a “very good” fit even at lower dimensions and we can observe that the decline of the stress for the number of dimensions greater than three is very smooth. Hence a three-dimensional perceptual space may be enough to capture the information within the dataset required for this application domain.

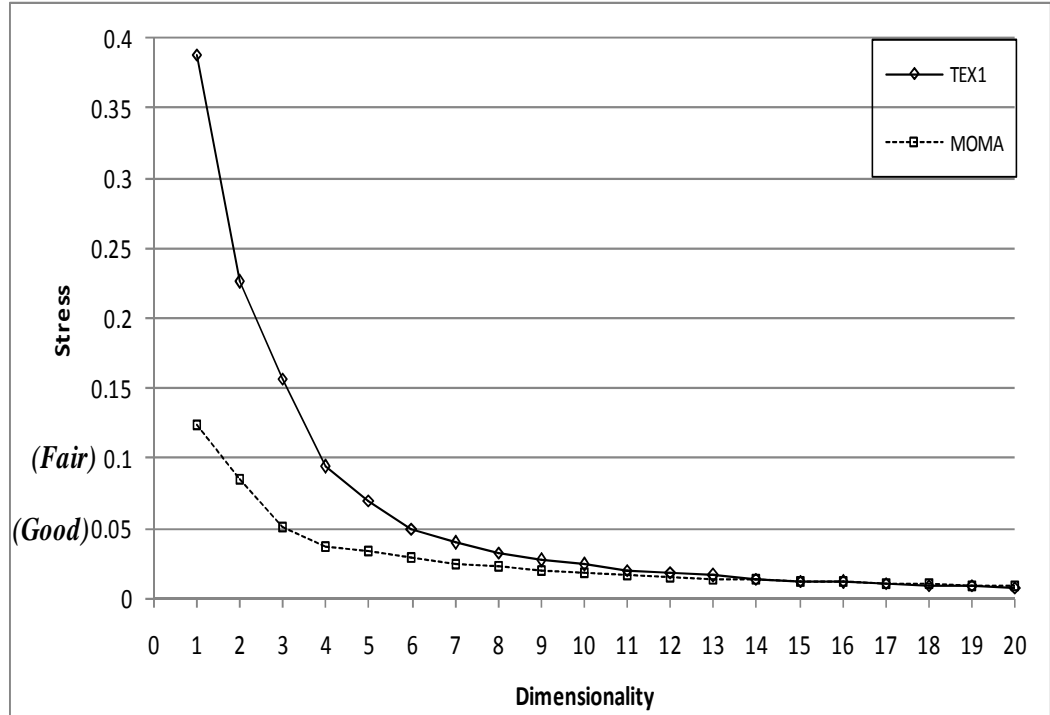


Figure 4.10- Normalised stress for the first 20 dimensionalities used to represent the MoMA and Tex1 datasets

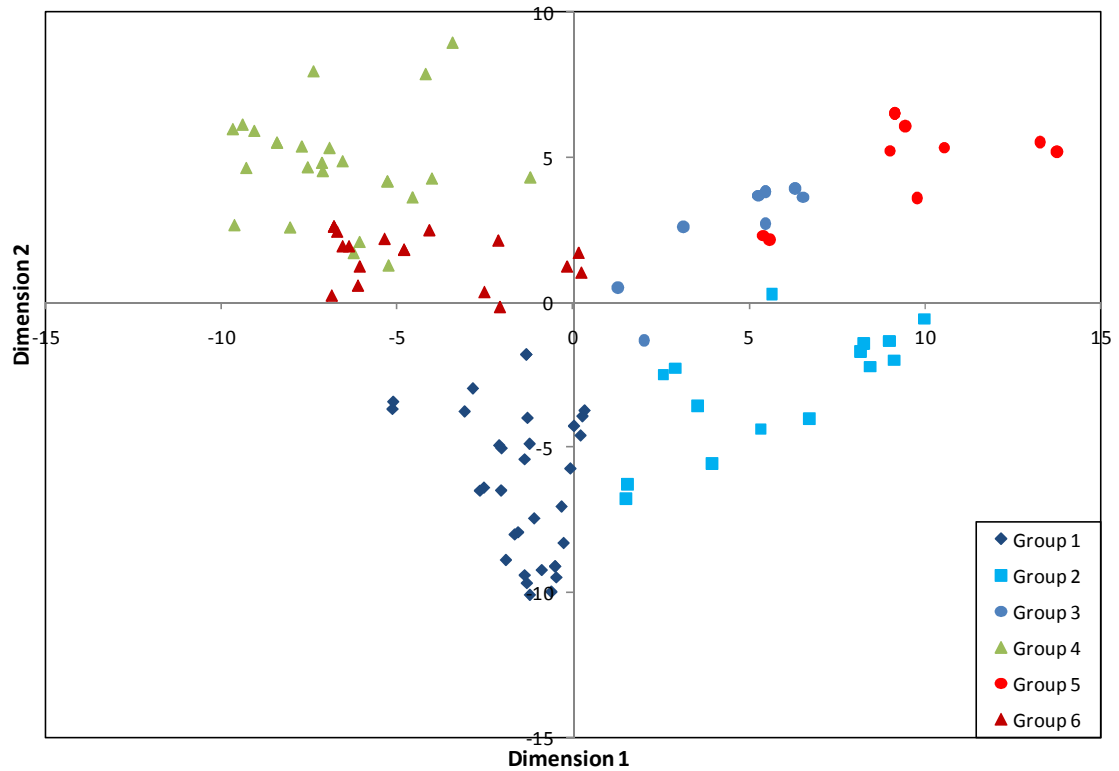
Unfortunately neither plot exhibits a clear-cut elbow and hence it is not possible to rely on the stress values for dimensionality selection. However, it seems likely that a perceptual space of between 4 and 10 dimensions would be adequate for the majority of texture processing applications.

4.5 Traits in major dimensions

This section uses both the MDS and clustering results to examine if there are any major traits within the main dimensions of the data. Four-dimensional perceptual spaces were used for this purpose as they represented a good trade-off between manageability and the use of the majority of the variability in the data.

4.5.1 The Tex1 dataset

Figure 4.11 shows the arrangement of the Tex1 textures within the first two dimensions of a 4D reduced perceptual space. The colour coding indicates the grouping obtained from the dendrogram in Figure 4.3 and shows that these major groups are well represented within the first two dimensions (albeit with some overlap).



*Figure 4.11 – Spatial representation of dataset after 120*120 Similarity Matrix is reduced to a 2D space using MDS*

Figure 4.12 provides more details of these data: moving around the quadrants shows that the shift from one texture category to another is fairly consistent. In the top left quadrant we have all unidirectional textures, albeit divided into 2 groups: mainly horizontal and vertical textures. Except some minor overlaps we can observe that even the horizontal and vertical textures are quite well separated. As we move anticlockwise from the top left quadrant we notice that the unidirectional textures are followed by bidirectional, structured textures in the bottom left quadrant.

Moving towards the bottom right quadrant the primitives in the textures become less apparent. The quadrant contains some coarse textures with primitives at relatively high frequency but still apparent to the naked eye. The remaining textures are generally fine and isotropic. The last quadrant (top right), contains some irregular textures consisting of large patches or circular structures. In this case the global information (or longer range structure) is more apparent.

While the transitions appear obvious to the eye it is not immediately apparent how they might be easily incorporated into the automated design of retrieval systems.

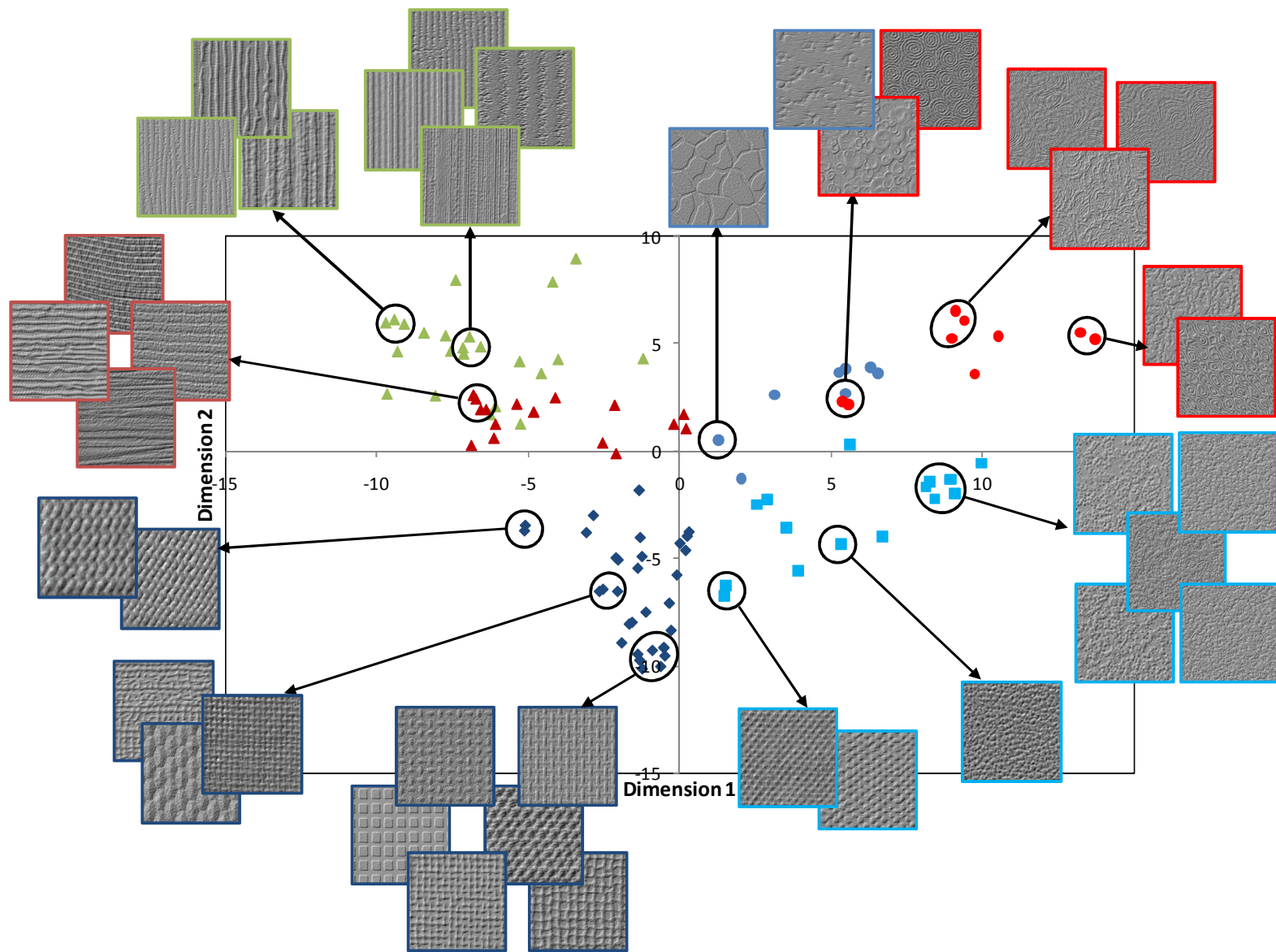


Figure 4.12 – Span of different categories of textures when the first two dimensions are considered

4.5.2 The MoMA dataset

Figure 4.13 shows the spatial distribution of the MoMA textures within the first two dimensions of a 4D perceptual space. The six groups presented have been obtained after cluster analysis of the psychophysical data and correspond to the groups shown in Figure 4.7.

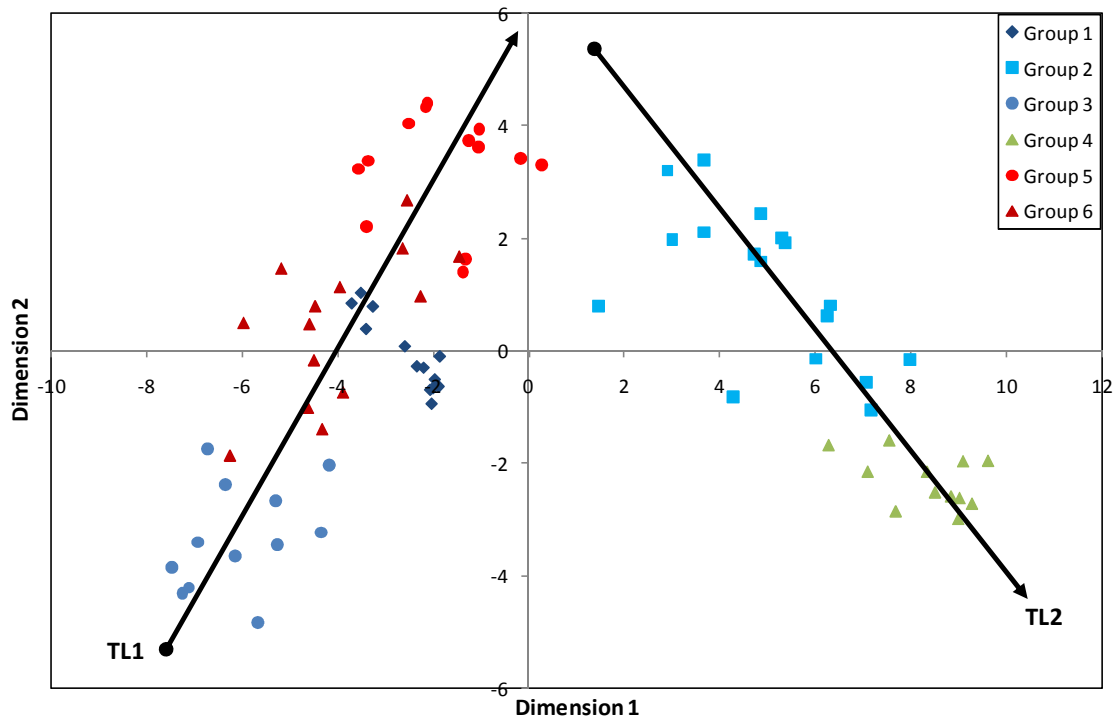


Figure 4.13– Spatial arrangement of MoMA textures in a 2D plane

Again we can see that the first two dimensions show the dendrogram groups as distinct clusters. However, unlike Tex1 the MoMA textures form more compact groups and are less scattered in the 2D plane. This is likely to be because the MoMA dataset contains a smaller range of more specialised textures.

Although the groups appear to be less cohesive, indicating that the groups can be easily broken into smaller and more compact groups, there are fewer overlaps among the groups than in the case of the Tex1 dataset. Overlaps are present mainly in the case of groups 1, 5 and 6. Increasing the number of dimensions and splitting the datasets into more groups contribute in increasing the compactness and separation levels of the groups.

Trend lines TL1 and TL2 have been investigated for any obvious visual traits. The results of moving along the trend lines TL1 and TL2 are displayed in Figure 4.14 and Figure 4.15 respectively. As discussed in Chapter 3, the MoMA dataset does not

contain many texture samples that exhibit much structure (non-random phase artefacts). It can be noticed, while moving along TL1 in Figure 4.14 that those textures are clustered close to rough textures containing some structural information. As we move up TL1, we get smoother textures.

As the coloured outlines of the thumbnails show, texture samples from different groups do overlap, but when inspected visually, we notice that the samples are very consistent in appearance.

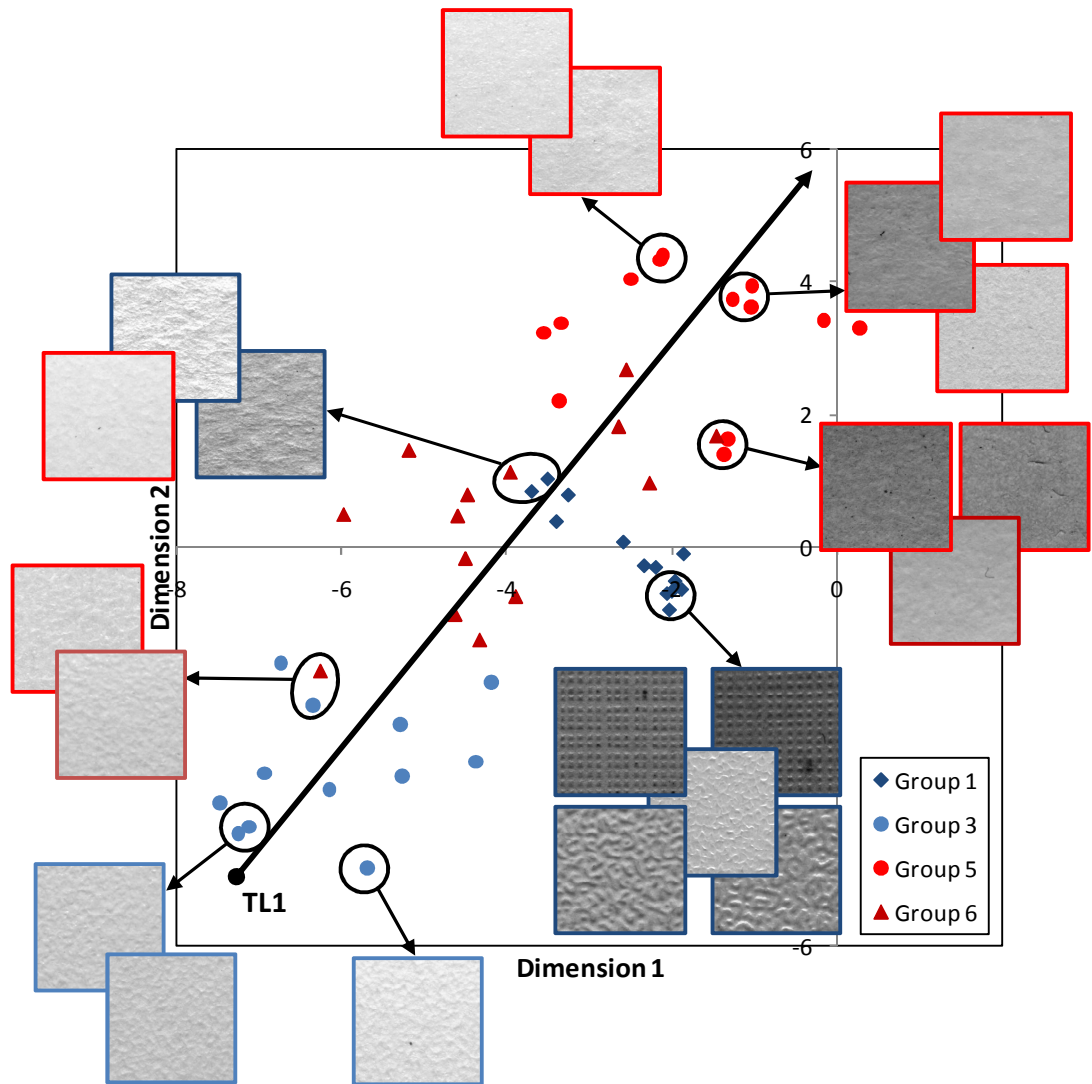


Figure 4.14-Variation of different texture categories along trend line TL1

The groups of texture samples that occur along trend line TL2 appear to be more compact than the ones along TL1.

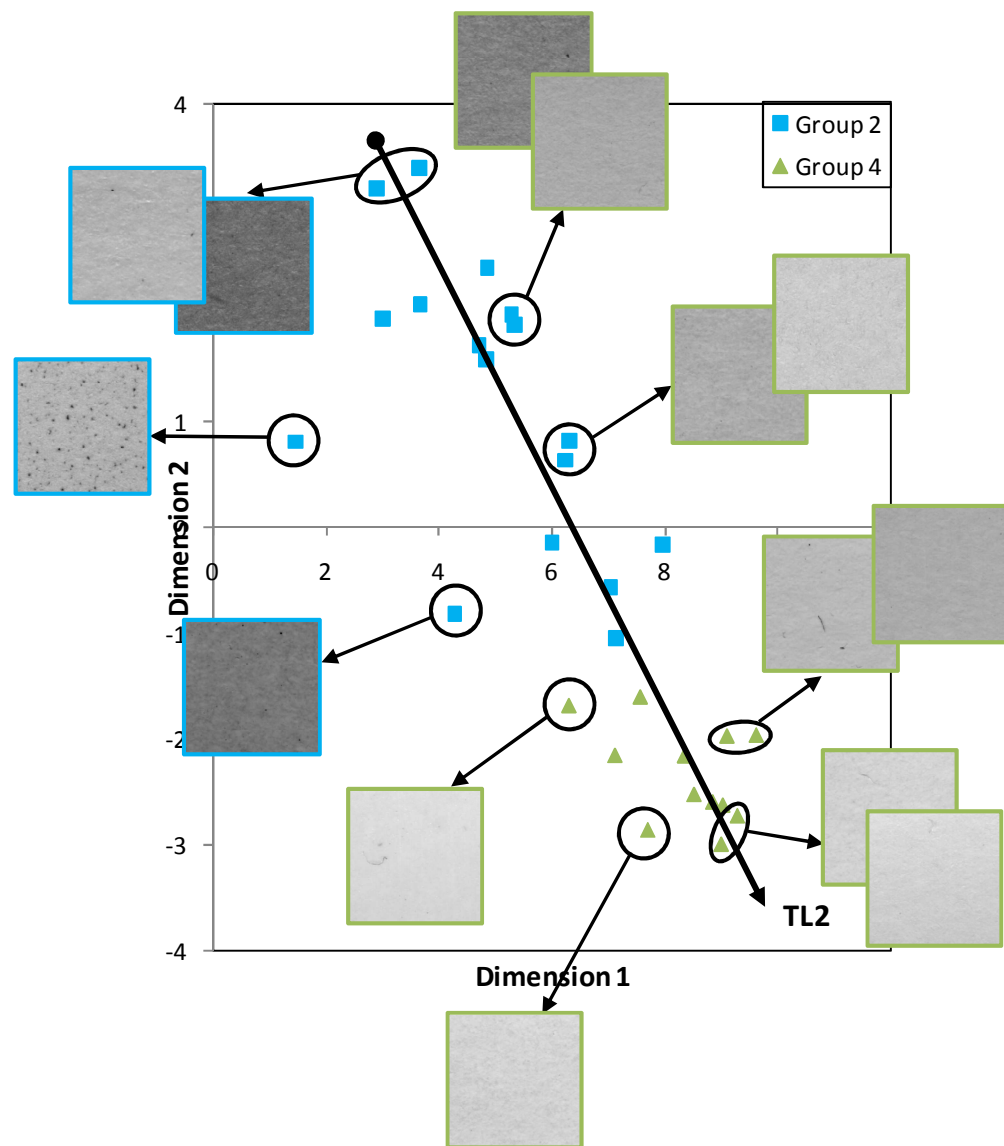


Figure 4.15- Variation of different texture categories along trend line TL2

Again however, there is no immediately obvious way of exploiting these trends for the design of a retrieval system.

4.6 Conclusions

This chapter has used hierarchical clustering, dendrograms, and Multidimensional Scaling to investigate the nature of psychophysical data derived from two different sets of surface textures: Tex1 and MoMA. In summary the results from this analysis are:

- (a) that the dendrograms show that there is obvious structure in the similarity matrices and they are certainly non-random;
- (b) that there is a wide spread of texture types in Tex1 and to a lesser extent in MoMA but that in both cases there is no obvious number of groups that should be extracted;
- (c) that these data can be represented well between four and ten perceptual dimensions; but that
 - (i) there is no ‘elbow’ in the stress graphs that suggests that any particular number of dimensions is *the* dimensionality that should be used for all applications,
 - (ii) these dimensions do not have any simple interpretation that indicates a particular feature set will provide the optimum or near optimum retrieval performance.

Thus while the similarity matrices have been shown to encode useful information, and MDS has been shown to be a valuable tool for reducing the complexity of the problem there is no really concrete evidence that suggests how many dimensions should be used or which features should be employed. This may be a consequence of using a relatively low number of texture samples dictated by the psychophysical approach selected to probe what may be an exceptionally complex perceptual space. For application specific problem, the number of dimensions can be selected using domain criteria. Examples are provided in chapters six and seven.

Chapter 5

Identifying features for texture retrieval

5.1 Introduction

The previous two chapters described the psychophysical studies performed to capture how observers perceive and group together surface texture. The next step involves creating a corresponding feature space for texture retrieval.

The dimensionality analysis performed in Chapter four demonstrated that, although the first two dimensions do cover the majority of the variations in the dataset, no obvious texture characteristics were apparent that could easily be directly exploited for feature selection.

In similar research Petrou *et al.* [Petrou07] concluded that the best approach was to perform automated feature selection on a large (i.e. several thousand) set of features. Here we also follow this approach. Thus our first criterion for feature selection is the availability of a large feature set.

In this chapter we examine four feature families in detail and select one for use in the texture retrieval experiments. However, before we discuss these features, we first present the major criteria used in this selection.

5.2 Feature Selection Criteria

The perceptual dimensions investigated in Chapter four did not lead to any dominant texture feature set that can be exploited for texture representation and mainly retrieval. In order to avoid any bias in selecting texture descriptors and to allow any retrieval model the freedom to select its own relevant features, a large soup of features is more appropriate for representing the texture samples being investigated.

Moreover, the feature set used within the scope of this thesis needs to be one that has already been investigated and well described in literature, i.e. no new feature description method will be investigated. In order to come up with some potential feature descriptors, a number of feature selection criteria are developed. These are explained in the remainder of this section.

5.2.1 Phase sensitive features

It has already been demonstrated that the phase spectrum contains most or all of the structural information in an image. The phase spectrum is very important in determining the placement of bright and dark spots in images. The complex cells in the primary visual cortex are very sensitive to this kind of information [Hubel68 & Haynes04], thus the phase information contributes immensely in helping people to recognise and interpret objects within an image. This can be illustrated by Figure 5.1, which shows a checkerboard image together with its magnitude only and phase only reconstructions. Even if the magnitude only image has the same variance as the original image, it appears to be visually different from the original checkerboard image, whereas the phase only image can be “visually” classified as in the same category as the original one due to the main structural information still being present.

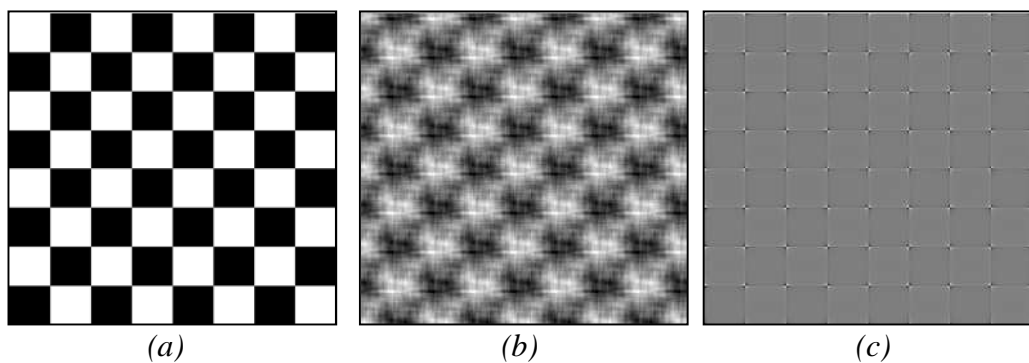


Figure 5.1 - Magnitude only, (b), and phase only, (c), representations of a checkerboard, (a)

In order to be able to reliably discriminate between different texture images or surfaces, it is important that the features or feature set chosen to represent those textures be able to encode not only the magnitude information in the image but also the phase information.

5.2.2 Power Spectrum sensitive features

The power spectrum has been extensively used to represent textures since different textures normally generate different energy distributions in the frequency domain, and that variation can be very easily and efficiently captured within the power spectrum, which represents the strength of each spatial frequency. Hence, if the spatial frequency domain is sliced appropriately, it is possible to represent different textures using different spectral energy signatures.

5.2.3 Position independent features

One type of invariance that we require from the features used is position invariance. Position or translation invariance has mostly been associated with features for object recognition. The human visual system already possesses a highly developed ability to fixate objects of interests and hence influence similarity judgements whenever position independence is concerned. Thus, when human subjects are presented with two samples composed of the same texture primitives but displaced by a certain amount, they can readily associate the two samples as being similar, as long as the texture primitives are not distorted and the placement rules are preserved. Making such a judgment requires no such effort on behalf of a human subject, however finding invariant descriptors to mimic human behaviour remains a constant struggle for researchers.

5.2.4 Generating large pool of features

Using learning models to build retrieval systems is fairly new to the CBIR community and has really gained interest within the last five years or so. However, the use of computational features to describe textures can be traced back more than two decades. Most of the features applied in the field of texture processing can be related to segmentation of either stationary or non-stationary texture images. Due to the subjectivity of human perception, we noticed the association of low-level features with high level descriptors such as directionality, coarseness, regularity and so on. However, this has not helped to reduce the semantic gap that exists between how humans perceive different categories of textures and the way a computational model emulates the human judgment. One of the main reasons lies in the fact that we are not able to understand the

mechanisms that the human brain employs in describing and discriminating textures. Likewise, associating high-level features with a certain type of texture helps only in limiting the way of “computationally describing” that texture [Petrou07]. Thus, the smaller the number of features employed, the greater the prejudice or bias in representing the different texture categories. In order to reduce those prejudices, we start from the idea that we don’t know what the high-level descriptions are for the available textures. We assume that any low-level features required to represent these high-level descriptions would be made available from a very large pool of features extracted from the texture samples. Hence any learning model applied would use this huge set of features to train a retrieval system that provides the same decisions as humans do.

5.2.5 Avoiding Redundant Features or Feature Sets

Approaches that use combinations of feature extraction techniques are common. However, this strategy can introduce irrelevance and redundancy within the set of features available [Yu03]. Irrelevance as defined in [Yu04] can only be determined when learning the retrieval system, and cannot be avoided at the stage of feature extraction if a large set of features is required. Redundancy, on the other hand can either be considered as the presence of highly correlated features, or otherwise the presence of two or more relevant features that contribute in the same way to describe a given texture characteristic. Redundant features not only increase the computational complexity of a retrieval system, but also degrade the performance of the system. Ideally we want features that are orthogonal such that changes in any particular feature do not trigger a change in any other feature selected to represent the texture samples.

5.2.6 Features that are inexpensive and simple to compute

This characteristic is required because of the previous criterion that we require large feature sets.

5.3 Investigating Feature Extraction Methods

Using the criteria described in the previous section, we now investigate four feature sets in more detail. These are:

- (1) Local Binary Patterns,
- (2) Gabor wavelets,
- (3) Synthesis features of Simoncelli and Portilla, and,
- (4) Trace Transform features.

Local Binary Patterns were selected because of their popularity in the literature and their non-linear characteristics [Ojala96].

Gabor wavelets were selected, again because of their extensive use in the literature, and for the more limited use of ‘Gabor phase’ [Bovik90 & Oppenheim91].

Simoncelli and Portilla’s features were selected because of their excellent performance for synthesis of phase rich imagery [Simoncelli95 & Portilla00].

Trace transforms were selected because they have been used in similar work reported by Petrou *et al.* [Petrou07].

5.3.1 Local Binary Patterns

Local Binary Pattern (LBP) operators generate binary codes that describe how the local texture pattern is built and was first introduced as a complementary measure for local image contrast [Ojala96]. LBP operators are very popular because they are fast to compute and are also invariant to monotonic changes in grey-scale.

LBP operators label each pixel of an image by using an N by N mask to threshold the neighbourhood around each pixel. The result of this thresholding operation is a local binary pattern which is interpreted as a binary number. Occurrences of different patterns are aggregated into a histogram and this forms the texture descriptor.

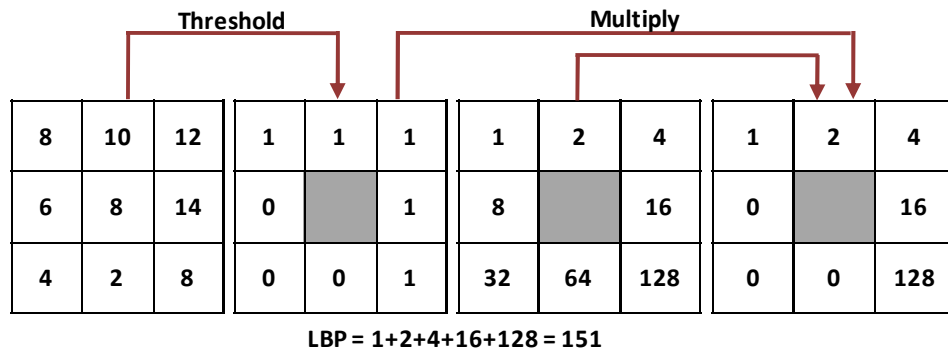


Figure 5.2 - Original LBP operator with associated weights

Figure 5.2 shows how the binary codes are computed when a 3*3 mask is applied. A predetermined mask of weight is applied to the binary code obtained from the thresholding process to generate a unique LBP feature. In order to provide for rotation invariant LBP features [Ojala02], circular symmetric neighbourhood sets are used as shown in Figure 5.3 .

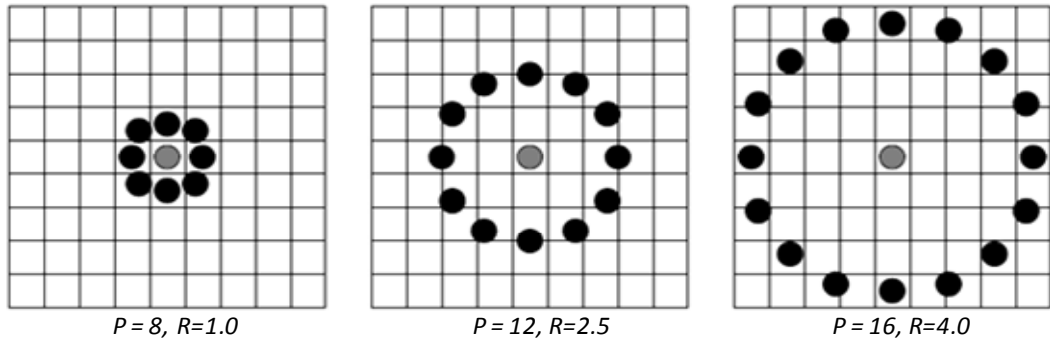


Figure 5.3 - Circular symmetric neighbour sets used generated at three different scales

In addition to providing rotation invariance, varying the number of samples, P , and the radii, R , of the neighbour sets allow the extraction of multiscale LBP codes. The operators used to generate multiscale codes are denoted by $LBP_{P,R}$. The use of multiscale neighbour sets can result in LBP histograms of containing very large numbers of bins.

To reduce the size of the histograms Ojala *et al.* [Ojala02] proposed the use of the so-called “uniform patterns”. These are based on a measure of uniformity that depends on the number of 1/0 or 0/1 transitions in the patterns.

Strengths of LBPs

LBP operators are simple to design and implement. The operators can be specifically tuned by the use of “uniform patterns” to represent different image primitives such as lines, corners, joints etc...More importantly, they are computationally cheap.

Weaknesses of LBPs

- 1) If used in its original form, i.e. features extracted on a 3 x 3 neighbourhood, LBP operators cannot capture large-scale features within textures.
- 2) The operator is also not very robust to local changes in texture, such as those originating from variations in illumination directions; however, this limitation can be ignored since we deal with a controlled illumination environment.

- 3) The use of multiscale LBP operators may result in sparse sampling of a 2D texture plane, which may not result in an adequate representation of the texture.
- 4) Moreover, sampling, as exploited by LBP operators may result in aliasing effects.
- 5) Using the full LBP patterns can result in histograms with a large number of bins (e.g. 2^{16} bins for a 16 bit pattern).
- 6) Uniform patterns containing 2 transitions have proved to be successful in the literature for texture analysis however, with the size of the histograms reduced considerably; the feature vector may not be large enough for retrieval purposes.

There is little information available in the literature concerning how LBP operators react to changes in phase and position. We have applied the $LBP_{8,1}$ operator to a randomly selected Tex1 texture (T89 selected here) and the histograms are recorded for (1) the original texture (2) with its phase randomised and (3) with the intensity values translated in a circular manner. The results for 10 bin LBP histograms generated by the operator $LBP_{8,1}$ are presented in Figure 5.4.

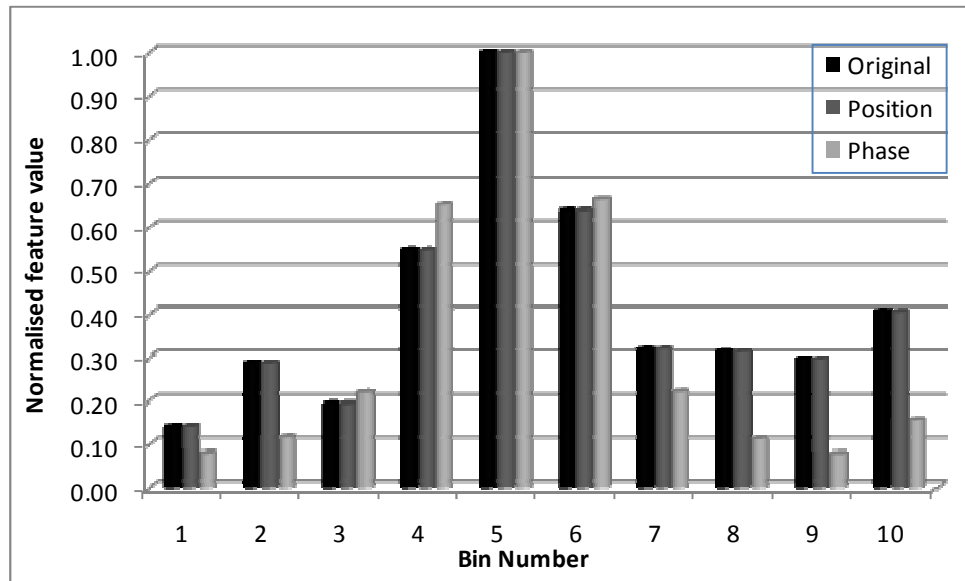


Figure 5.4 – Variation of histogram values for when operator $LBP_{8,1}$ is applied to texture T89

We notice that this LBP operator is both position independent and phase sensitive. This makes LBPs a potential candidate.

5.3.2 Gabor features (Phase and power spectrum features)

Gabor wavelets have been extensively used in texture segmentation due to the fact that they allow multi resolution (or multi spectral) decomposition through proper tuning of their orientations and radial frequencies due to their localisation capabilities both in the spatial and spatial frequency domains. Thus they can be designed to be highly selective in both position and frequency. However Gabor filters gained much more importance with research showing that the Human Visual System processes images by decomposing them into a number of subbands. They provide similar characteristics and allow them to “mimic” the Human Visual System [Claus00]. Multiresolution filtering techniques have extended the use of Gabor wavelets in order to cover areas such as texture image retrieval and classification [Porat89 & Bovik90].

Given their joint spatial/spatial-frequency localisation capabilities, sets of Gabor filters have been used both in the spatial domain and the spatial frequency domain. Figure 5.5 shows representations of Gabor filters in both the spatial and frequential domains.

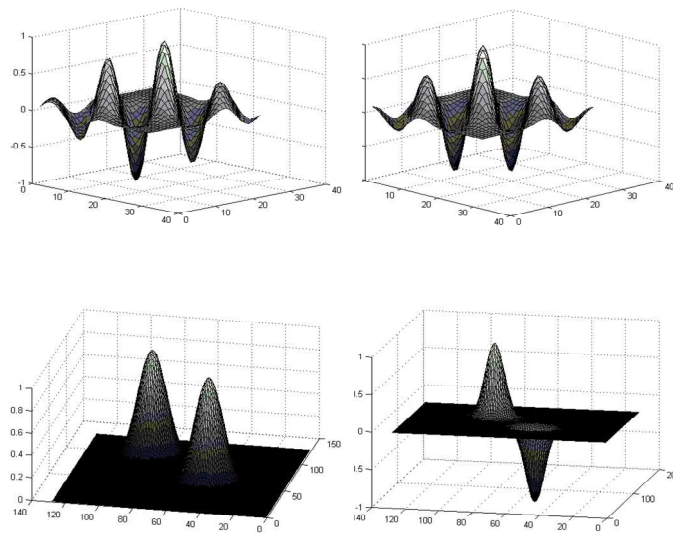


Figure 5.5 - (top row) spatial representation of Gabor wavelet pair, and (bottom row), corresponding frequency domain representation

Gabor filter outputs have been used in different ways to provide texture features. While some people have computed the moments of the distribution of the responses in the spatial domain, others extracted features by creating geometrical (and central) moments based in the spatial-frequency domain [Bigun94]. Texture features have also been computed from the magnitude response of Gabor quadrature filters [Bovik90], or just from the real component [Jain91].

In order to achieve optimal separation of texture features and thus provide for robust texture representation, multichannel texture processing has been proposed in the literature [Randen99]. Filter banks that allow the frequency spectrum of any textured image to be decomposed into a given number of subbands resulting in different feature signatures for different textures have been heavily exploited. The most popular filter banks encountered in literature are the dyadic Gabor filter banks. An example of such a filter bank is illustrated in Figure 5.6.

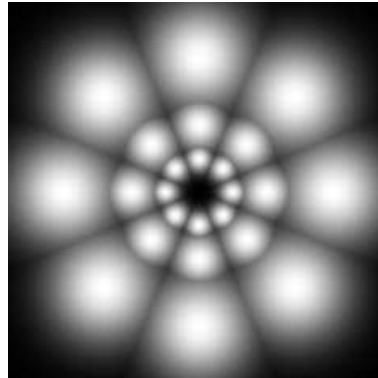


Figure 5.6 - Dyadic Gabor filter bank with 4 orientations and 3 frequencies

Strengths of Gabor features

- 1) They appear to share common Human Visual System properties and have been exploited to simulate the way it functions.
- 2) They are localisable both in space and frequency and have been exploited heavily in detecting approximate basis sets for the representation of textures.
- 3) They can be oriented and tuned such that they act as edge and line detectors. Hence their significant use in texture segmentation.

Weaknesses of Gabor features

In their basic form they only extract power spectrum information. Gabor phase [DuBuf90, DuBuf91 & Oppenheim91] has been used successfully for segmentation purposes. However, their use for classification (and retrieval) purposes suffers from position sensitivity.

5.3.3 Simoncelli's features

The model presented by Portilla and Simoncelli [Portilla00] follows a number of texture models that are based on the application of oriented linear kernels at multiple spatial scales for representation and synthesis. However, the use of orthogonal separable wavelet decompositions for texture analysis and synthesis exposed a number of limitations. These were the inability of the wavelets to capture extended contour information and large scale structures.

To overcome those limitations, Simoncelli *et al.* presented their universal parametric model for texture representation. The latter model is based on the use of directional derivative operators of any desired order in the form of steerable filters [Simoncelli95]. The most important characteristic of the steerable pyramids is that through polar-separable decomposition in the frequency domain they allow for independent representations of scale and orientation. Figure 5.7 illustrates the steerable filters at different scales and orientations.

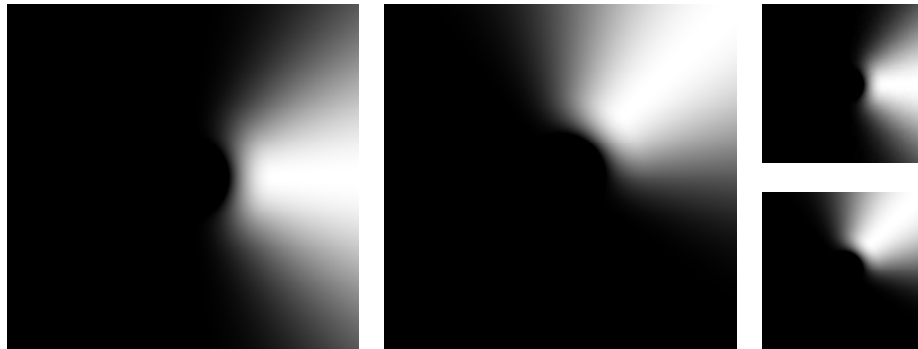


Figure 5.7 - Steerable filters at 2 different orientations and scales

Simoncelli *et al.* compute the features of their parametric model on the response images obtained after applying a pyramid of steerable filters to the original image. The responses can be represented as N pyramids of response images where N represents the number of orientations at which the filters were applied. Each pyramid, in turn, is composed of M images at different scales.

The pyramidal representation of the responses is performed in the spatial domain where images at each scale and orientation are complex, with real and imaginary parts being quadrature pairs (due to the application of a Hilbert transform while computing the filtered image in the frequency domain).

Simoncelli *et al.*'s features are derived from fixed overcomplete multi-scale complex wavelet representations. The features are based on the pairs of wavelet coefficients computed for adjacent spatial locations, orientations and scales. The features are computed either on the local magnitude information or on the real and imaginary coefficients of quadrature pairs in the spatial domain.

The features used by Simoncelli *et al.* in the analysis and synthesis of textures [Portilla00] are as follows:

- 1) Global Marginal Statistics extracted from the intensity information of images or height information of surfaces. Feature vector comprises of mean, variance, skewness, kurtosis and range of distribution.
- 2) Mean magnitude information of filtered images at N scales and M orientations.
- 3) Autocorrelation features based on the subband decomposition of each subsampled image.
- 4) Cross Correlation of magnitude information at different scales.
- 5) Cross Correlation of phase information at different scales

Strengths of Simoncelli features

The use of oriented linear filters at multiple spatial scales allows the encoding of maximum information within the spatial domain.

Additionally, the model can generate large feature sets based on the raw wavelet coefficients computed at varying scales and orientations.

Weaknesses

Simoncelli's features are based on overcomplete wavelet representations and contain considerable redundancy. The set is non-adapted to specific texture categories and they are not able to explain which feature set is more dominant for which category of texture.

The multi-scale wavelet representation to compute the feature is a complex and inappropriate when texture retrieval from large datasets is considered. However, the biggest disadvantage for use here is that many of the phase sensitive features are also position sensitive.

5.3.4 Trace Transform features

The Trace Transform is a procedure through which a triple transformation applied to a surface, S , within a given coordinate system, C , results in a scalar value that acts as a signature for that surface. This transform originates from the fact that a 2D function can be fully reconstructed if knowledge of its integrals along straight lines defined in the spatial domain representing the signal is available. The Trace Transform is in fact a generalisation of the Radon Transform that has been successfully applied in the field of computer tomography [Kadyrov01]. Figure 5.8 (a) and (b) illustrate the geometry for both the trace and the radon transforms.

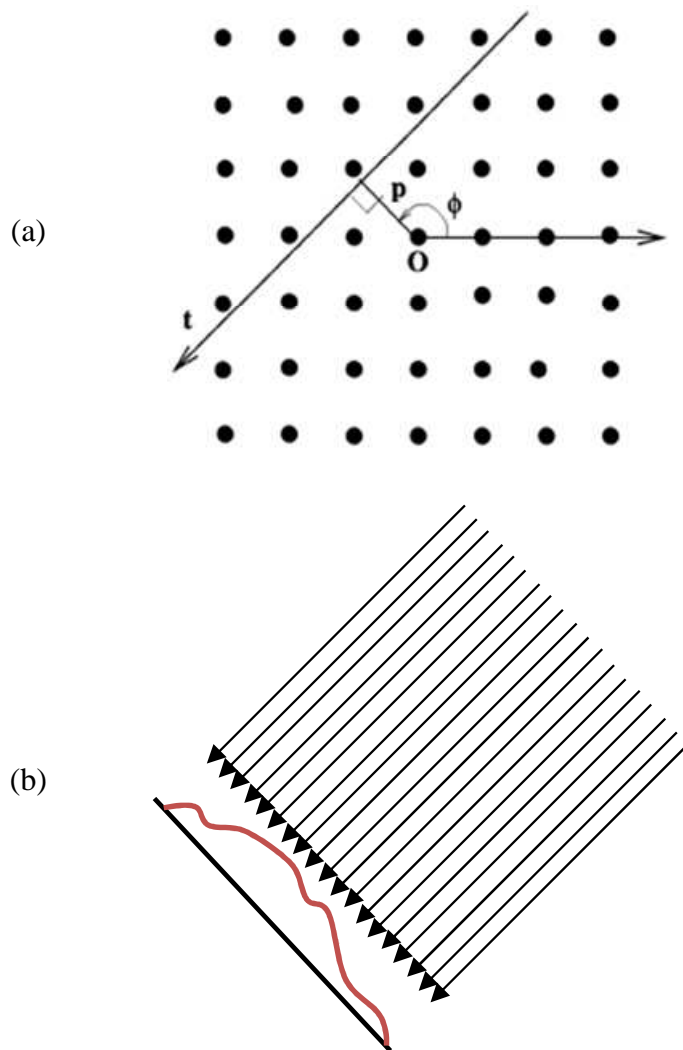


Figure 5.8- (a) Parameters associated with a tracing line, (b)
Converting 2D surface to a 1D function

Tracing lines, t are drawn at different values of ϕ and p . ϕ represents the angle that the normal, joining the tracing line from the origin, makes with the horizontal axis; whereas

p is the length of the normal that joins the origin and the tracing line. The functionals applied along the tracing lines are known as the *trace functionals*.

In the case of the Radon Transform, the trace functionals are only integrals over the parameter t , whereas for the Trace Transform, these functionals can take any form. p is sampled in such a way that the tracing lines are parallel to each other when a particular value of ϕ is considered and the result, after applying a given functional, is a 1D function as shown in Figure 5.8(b). Applying the Trace Transform for different values of ϕ leads to a 2D function of variables ϕ and p . ϕ is sampled in the range $[0, 2\pi]$ and p lies within the range $[-p_{max}, p_{max}]$, where p_{max} is limited to half diagonal length of the surface considered.

Given that we are dealing with discretised values only, the 2D function resulting from the Trace Transform can be represented as a 2D matrix with the change along the columns representing the change in p and the change along the rows representing the change in ϕ . Another functional, P , can then be applied along the columns of the 2D function which results in a 1D function representing only the changes in the value of ϕ . A third functional Φ along the resultant 1D function generates a scalar value which is used as a feature for the surface considered.

The whole process is presented as the Triple Feature construction by Kadyrov *et al.* [Kadyrov01 & Kadyrov02]. In addition to the trace functionals, T , the P functionals have been referred to as the *diametric functionals* and the Φ functionals are referred to as the *circus functionals*. Assuming that different functionals can be applied for T , P and Φ , the scalar value generated can be represented in the form

$$f^{ijk} = \Phi^i(P^j(T^k(S(C; \phi, p, t)))) \quad (5.3)$$

Figure 5.9 shows the process of computing feature f^{ijk} for texture T76 (Figure 5.9(a)) from the Tex1 dataset. The trace functionals applied for illustrative purpose in this case corresponds to integrals over the parameter t , finding the maximum over parameter p , and finding the integral over ϕ .

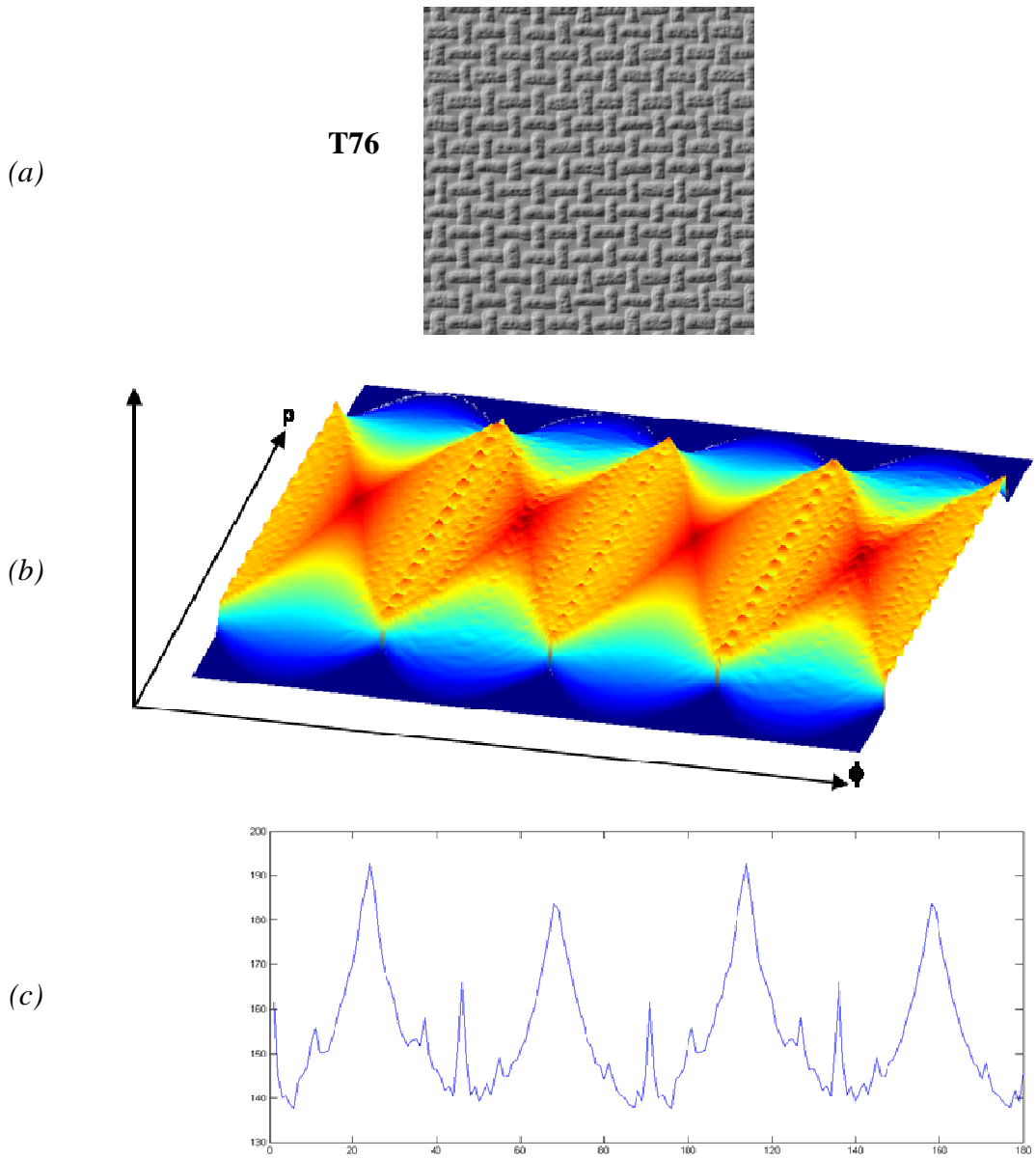


Figure 5.9 - Triple feature construction from original texture (a), to2D function (b), and transformed to a 1D function (c). The final result is a scalar obtained from (c)

Strengths of the Trace Transform (TT) features

The main advantage of the TT feature set is that it allows the generation of thousands of features by varying the type of functionals used for the T , P and Φ transforms. Furthermore, TT has already been utilised in a perceptual context to learn how human rank different categories of textures [Petrov07].

Moreover, functionals T , P and Φ can be selectively chosen so that the resultant TT features are invariant to rotation, translation and scaling transformations [Kadyrov01].

Weaknesses

The capability, by the TT, to generate thousands of features comes with the drawback of large memory utilisation while computing for very large feature sets and significant storage on hard disk for later use.

Kadyrov *et al.* provide functionals that cater for both phase sensitivity and position dependence [Kadyrov01].

5.4 Feature Set Selection

The investigation performed in the previous section elaborated on the strengths and weaknesses of some commonly used texture feature extraction approaches. Table 5.1 summarises the characteristics of each feature set with respect to the selection criteria presented in Section 5.2.

			Criteria					
			Position independent	Phase sensitive	Power Spectrum	No Redundancy	Inexpensive and Simple	Large Pool
Features	Multi-scale LBP		✔	✔	✔	✘	✔	✘
	Gabor	Power	✔	✘	✔	✘	✔	✘
		Phase	✘	✔	✘	✘	✔	✘
	Simoncelli		✘	✔	✔	✘	✘	✔
	Trace Transform		✔	✔	✔	✔	✔	✔

Table 5.1 – Eligibility of selected features with respect to chosen criteria. ✓ means eligible and ✗ means ineligible.

LBP features are good candidates for texture representation and have been successfully applied for segmentation and classification purposes. However, the use of uniform patterns reduces the size of the LBP histograms and thus results in feature vectors too small for texture representation. If the full LBP code is used as a separate signature, then the size of the histograms becomes too big (say for 16 bit codes) and the representation is too sparse. Thus LBP codes have been not been considered within this thesis.

Gabor phase and power features are still very popular, however the phase sensitive features are position dependent and cannot be used here.

Simoncelli features have also been rejected because they do not provide position independent features.

Finally, we have selected Trace Transform features to represent the textures. In addition to being able to generate considerably large feature sets, the TT features are easy to compute and depending on the size of the feature vector, they are relatively inexpensive. Functionals that capture power and phase spectrum information are provided in the literature [Kadyrov01] and the TT features provide invariance to several affine transformations [Petrou04], of which position independence is one. A limited set of

functionals based on the ones proposed by Kadyrov *et al.* [Kadyrov02] is provided in Appendix E.

5.4.1 Feature Normalisation

Dealing with large a pool of features or sets of features inevitably leads to variation in the span of the different features considered. Given that, within the scope of this thesis, the Similarity between two textures is obtained by applying a distance function, it is important that the distance value computed is not dominated by features with wide value ranges. Thus, feature normalisation is performed on all the extracted features such that each feature contributes more or less equally to the final distance measure. After considering several normalisation procedures [Aksoy01], the one that has been applied linearly transforms all the features to have zero mean and unit variance. Furthermore, assuming the features to be normally distributed, we perform an additional scaling and shifting of the features values such that all the features are found within the range [0, 1]. The transformation is performed as follows:

$$\tilde{x} = \frac{(x - \mu)/3\sigma + 1}{2} \quad (5.4)$$

5.5 Conclusion

The objective, in this chapter, was to select a candidate set of texture features to be used to map the MoMA and Tex1 textures from the 4D perceptual spaces derived in Chapter four to a 4D feature space that could be exploited for texture retrieval.

An investigation into four commonly used texture feature extraction approaches was performed and we arrived at the conclusion that the Trace Transform features were the most appropriate texture features for use in the research described in this thesis.

Chapter 6

Surface Texture Retrieval

6.1 Introduction

The goal of this chapter is to use the psychophysical data presented in Chapter four together with the pool of Trace transform features identified in Chapter five to develop retrieval systems for both the Tex1 and MoMA databases.

Given the high dimensionality of the original psychophysical data the first task is to reduce this “Full Perceptual Space” to a more manageable “Reduced Perceptual Space”. Chapter four described the use of MDS for this purpose, however, the resulting stress graphs showed a gradual degradation, making selection of an optimal dimensionality difficult. It concluded that the selection of dimensionality is best determined with reference to the application. The first part of this chapter therefore investigates the effects of reducing dimensionality on optimal retrieval performance.

Having determined the complexity of the space required for a given retrieval capability the task is to map these Reduced Perceptual Spaces onto a set of texture features that can be used to perform automated retrieval. After analysing the characteristics of the resulting “MDS derived Feature Spaces” (MFS), a series of retrieval experiments was conducted and their performances evaluated.

A simple alternative to this dimensionality reduction approach would be to use the Full Perceptual Space directly.

Thus the objectives of this chapter are:

- i) to determine the dimensionality required of perceptual texture spaces for a given retrieval performance level,
- ii) to map these Reduced Perceptual Spaces onto suitable feature spaces,
- iii) to evaluate the performance of retrieval systems that exploit these feature spaces, and,
- iv) to compare this approach with a more direct method that employs the Full Perceptual Space.

PART I: MFS Based Texture Retrieval

6.2 Overview of the development and retrieval processes

This section provides an overview of the four stages involved in the development of the retrieval systems and their operation as shown in Figure 6.1. This is provided for reference throughout the rest of Part I.

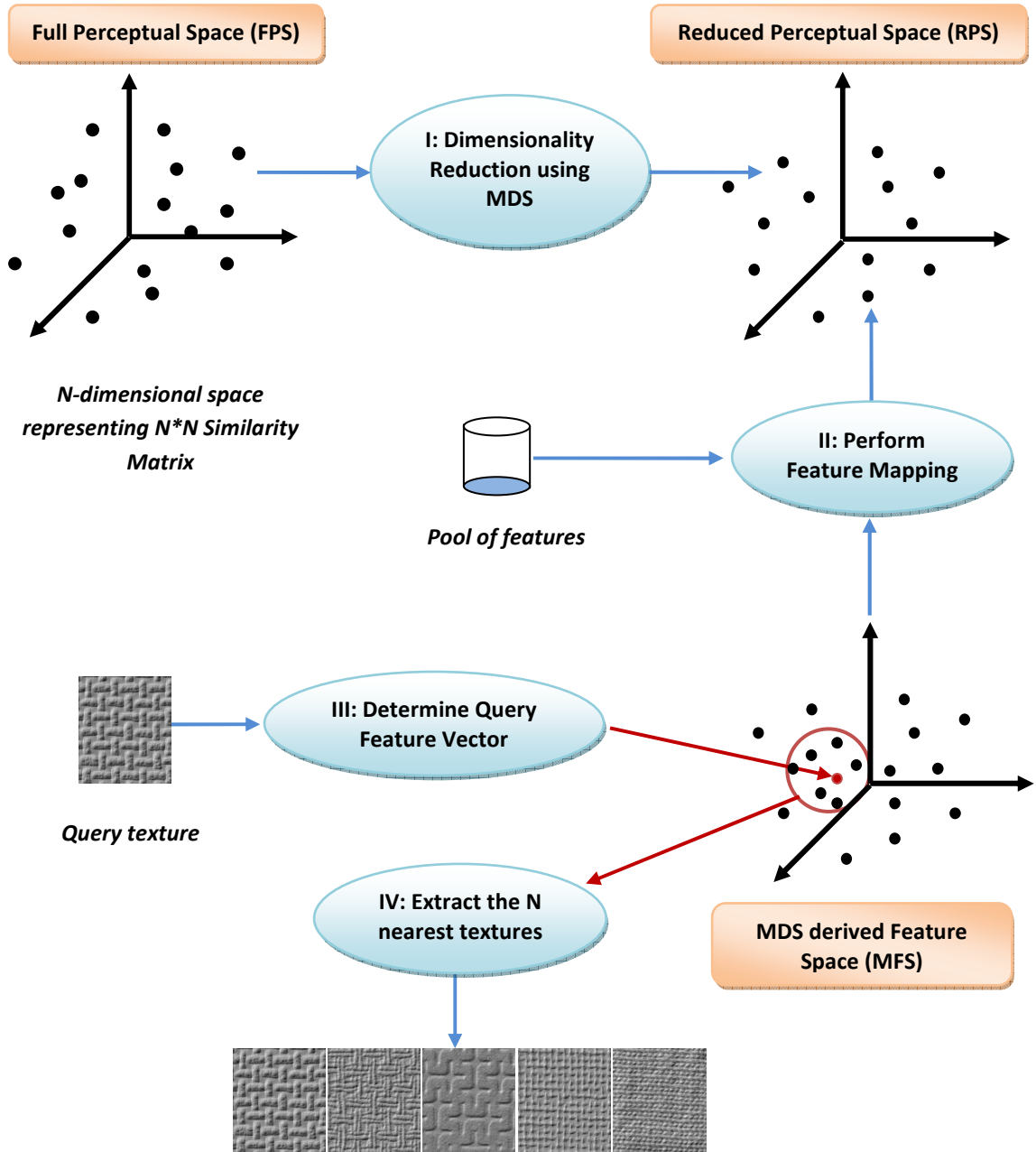


Figure 6.1–The four main stages of the proposed MDS based Retrieval Model

The four stages of are listed below.

- (I) MDS is used to reduce the dimensionality of the ‘Full Perceptual Space’ (FPS) obtained from the psychophysical experiments to produce a ‘Reduced Perceptual Space’ (RPS).
- (II) Feature selection and regression analysis are used to produce a feature space that approximates the RPS. This may be thought of as a ‘prediction model’ i.e. given the height map of a texture, can we use this feature space to predict the texture’s position in RPS. We term this space the ‘MDS derived Feature Space’ or MFS for short.
- (III) In order to process a query, the feature vector of the query surface is calculated in this new feature space (the MFS).
- (IV) The query’s feature vector is used to retrieve the n nearest textures in MFS.

6.3 Stage I – Determining the number of perceptual dimensions to be used for retrieval

In chapter four we used MDS to reduce the Full Perceptual Spaces (which comprised high-dimensional sparsely sampled similarity matrices) to Reduced Perceptual Spaces. As demonstrated by the low stress values in Figure 4.10, the mapping allowed most of the relevant information to be captured within relatively low-dimensional spaces. However, due the smoothness of these stress graphs it was difficult to pick a single dimensionality that would best represent the datasets and we concluded that this choice is generally application dependent.

The purpose of producing a low-dimensional RPS is to be able to fit a prediction model (the MFS) to that space that would allow us to perform effective retrieval. If the number of dimensions is too large, the data samples within that space become too sparse and it becomes difficult to fit a reliable prediction model. On the other hand, if the number of dimensions is too small then there is the danger that we will discard pertinent information and reduce the performance of the retrieval system to below that deemed to be acceptable.

Our strategy is therefore to assess the suitability of different RPS directly by using them for retrieval. We make use of ‘precision’ as a measure of their retrieval performance and choose the RPS dimensionality that provides suitable retrieval rates. Of course as the ‘prediction model’ represented by MFS is an *approximation* to the RPS, the automated retrieval system constructed using a particular RPS is likely to produce a lower performance than one that would be provided by the RPS itself. However, the MFS has the advantage that it may be used with query textures that were not included in the original psychophysical experiments whereas the RPS cannot.

Precision [Salton68], within the Information Retrieval community, is normally defined as follows:

$$\text{Precision} = \frac{\text{No. of Relevant Samples Retrieved}}{\text{No. of Samples Retrieved}}$$

Although we will use precision to evaluate retrieval performance at a later stage in this chapter, we are presently going to use it to assess how change in dimensionality of RPS can affect retrieval rate.

The precision for each query object (or texture) O_i within the Tex1 or MoMA dataset $\Omega = \{O_1, O_2, \dots, O_N\}$ is computed separately and the average precision is computed for the RPS dimensionality considered. The process is repeated for increasing dimensionality of RPS.

The following procedure was used to derive the average precision of retrieval for RPS over a range of dimensionality d .

1. For every object (texture) O_i from Ω , extract all of the remaining textures in order of their proximity to O_i within the FPS to provide the ordered retrieval set R_p^i . Repeat using all members of Ω as the query i in turn to provide the ‘gold stand’ retrieval R_p^i for each texture.
2. Initialise d to 1,
3. Use MDS to compute the RPS (i.e. the distance matrix of the objects $\{O_1, O_2, \dots, O_N\}$) of dimensionality d .
4. Repeat step 1 for the RPS of dimensionality d to produce the set of retrievals R_d^i .
5. Compare all R_d^i with corresponding R_p^i to obtain the number of common textures for specific retrieval numbers and divide by number of retrievals considered,
6. Compute the average precision for the all the objects in Ω .
7. Increment d and repeat from step 3.

6.3.1 Results of the analysis of Reduced Perceptual Space (RPS)

Figure 6.2 (a) and (b) show the retrieval precisions for the Tex1 and MoMA datasets respectively. Precision rates were determined using the RPS for each dataset, with dimensions of the RPS ranging between one and fifteen. Precision values were computed for retrievals of the first 10, 20 and 30 samples for each dataset.

We can observe, from Figure 6.2, that the precision rates converge quickly to 1 for both the Tex1 and the MoMA datasets. They show that, except for Tex1 ten sample retrieval case, we can achieve precision rates above 90% using a four-dimensional RPS (i.e. $d = 4$).

Additionally, the rate of convergence is faster for the MoMA dataset and this can be explained by the fact that there are fewer variations in the texture samples for that dataset. Consequently a 3D RPS would be equally meaningful for the MoMA dataset i.e. when $d = 3$. However for the analysis and evaluation of the MoMA and Tex1 datasets, a common dimensionality, i.e. $d = 4$, is used throughout this chapter.

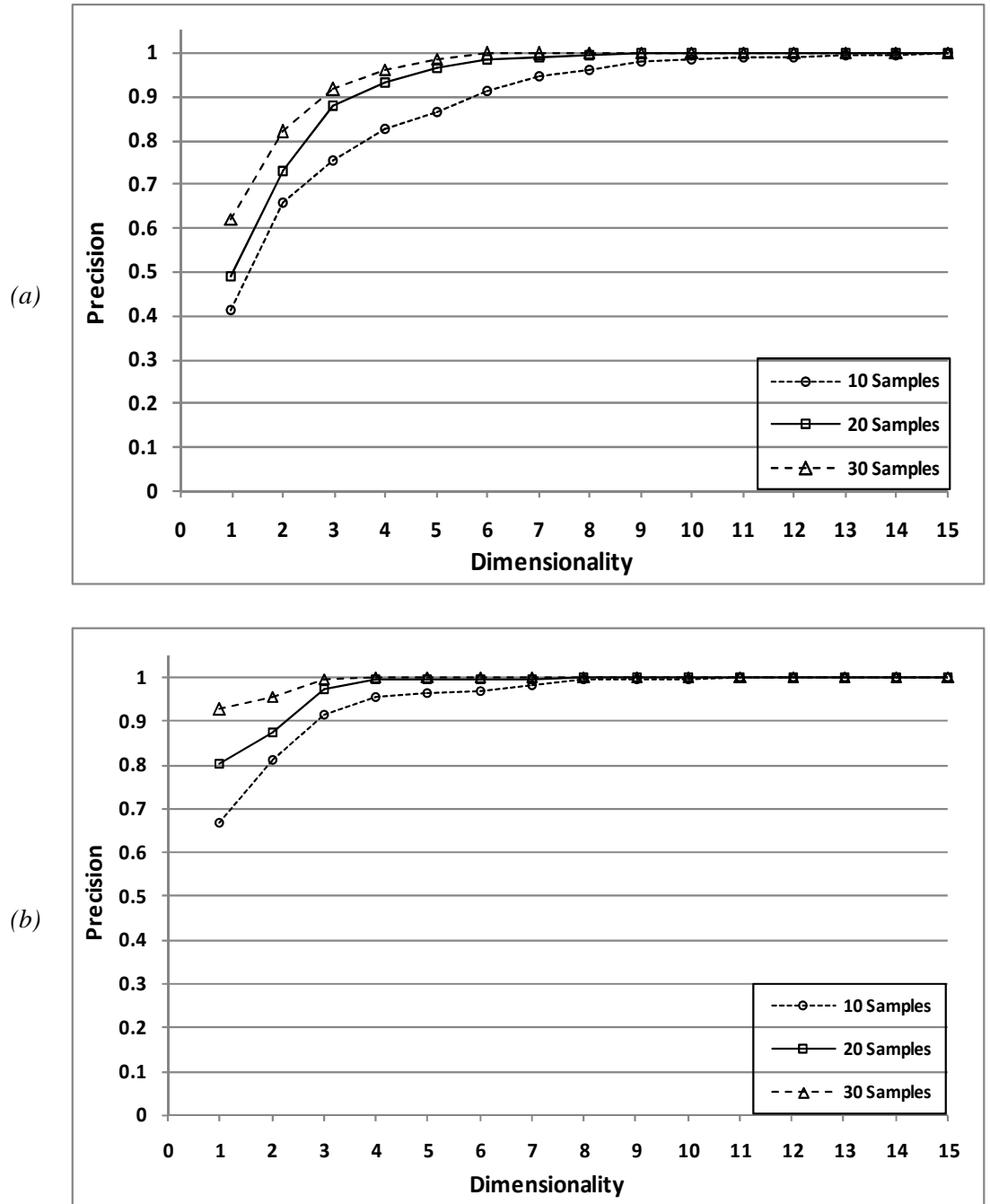


Figure 6.2- Graph of precision versus dimensionality, (a) precision for Tex1 (b) for MoMA dataset

6.4 Stage II – Producing an RPS to MFS mapping using regression analysis

Recent developments in fields of machine learning, computer vision and cognitive science have seen the use of thousands of features, extracted on relatively few samples, to build predictors or classifiers that could be used to predict outcomes of future or unknown observations. This concept, mostly applied in fields such as genomic studies, is of particular interest within the scope of this research given the similar challenges presented [Molinaro05]. As in the case of genomic studies [Molinaro05], we have also considered that a very large pool of features will enable a better mapping of the perceived similarities represented in the RPS. However, the number of texture samples used in the psychophysical experiments had to be kept low due to the practicality of experiments with human observers.

Hence the retrieval framework presented in Figure 6.1 views the retrieval process as a prediction problem where a query texture is regarded as an ‘unknown’ observation and the outcomes of the prediction model are considered to be the retrieval results in terms of decreasing similarity. The mapping of the texture samples to the feature space (MFS) is performed using the reduced dimensionality RPS. This section thus presents stage II from the retrieval framework (see Figure 6.1).

6.4.1 The Prediction Model

To model the prediction problem, we assume the N texture samples used in the psychophysical study are characterised by a set of N feature vectors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$. Each feature vector contains f feature measures that have been computed directly from the texture samples and are represented as $\mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^f)$.

The prediction problem can thus be modelled in the form of:

$$\mathbf{y} = g(\mathbf{X}) + \varepsilon \quad (6.1)$$

where ε is the error and $\mathbf{y} = (y_1, y_2, \dots, y_N)$ provides the axes of the MFS. The problem therefore is to determine the function $g(\mathbf{X})$ such that the resulting MFS(Y) approximates its RPS.

The formulation provided in equation (6.1) is a standard regression problem. Regression analysis has been used extensively for prediction purposes in the field of sales forecasting, dendroclimatology and others. Within the context of texture retrieval,

work that has used regression analysis to train a retrieval model and for the prediction of retrieval results is scarce. Long and Leow [Long01] used a hybrid model based on a combination of neural networks and Support Vector Machines to perform invariant and perceptual mapping of textures. Long and Leow's model involves a complex setup and their training algorithm is not clearly defined. The authors do not provide any indication of how an unknown texture sample could be mapped within their invariant space first and thereafter to the perceptual space. Moreover, the psychophysical experiment performed by Long and Leow is similar to the one performed by Rao and Lohse [Rao93a], in that they also make use of Brodatz texture images. We have already discussed in Chapter two that the use of Brodatz textures is not likely to produce either reliable psychophysical results or reliable feature vectors. Hence the validity of Long and Leow's model must be questioned.

Given that MDS produces linear uncorrelated axes, we have decided to model the relationship between \mathbf{y} and \mathbf{X} one dimension at a time and to do this using a simple linear regression function. Thus we assume that the coordinate y_i of a given texture T_i in a particular axis of the MFS can be modelled using a linear combination of features as in equation 6.2:

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_f x_i^f \quad (6.2)$$

where $\boldsymbol{\beta} = \{\beta_0, \beta_1, \dots, \beta_f\}$ is the vector of regression coefficients.

Using a least squares loss function, the regression coefficients are optimised by minimising the square error function:

$$R(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^N (y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))^2 \quad (6.3)$$

Thus:

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (6.4)$$

The regression coefficients, computed using equation (6.4), are stored and used to estimate the location of the dataset textures within the MFS. The same coefficients are used to estimate the location of a query texture within the MFS, with different values for the selected features. The retrieval process is just a matter of finding the nearest neighbours of the query texture within the MFS.

As we described in Chapter five, we have chosen the Trace Transform (TT) as the source of our features as it provides a pool of several thousand measures (we use 3136). The reason is that a large set is likely to increase the probability of being able to produce an accurate mapping of the MFS to the RPS.

However, it will also include many irrelevant features and this brings two practical problems:

1. including irrelevant features in the learning process increases the computational cost, and
2. irrelevant features detract from the accuracy of the prediction model as they introduce more noise.

The feature selection process is discussed in detail in the next section.

6.4.2 Feature Selection

Feature selection is a central problem in machine learning and statistics and is being actively researched mainly due to its importance in the area of data mining. In the context of regression and specifically within the scope of this thesis, feature selection will be applied for the following reasons:

- a. to perform retrieval at a lower computational cost by retaining only ‘relevant’ variables,
- b. to enhance predictive accuracy of the retrieval model by eliminating irrelevant features, and,
- c. to eliminate the effects of correlated features within the regression model by removing features which are highly correlated with more dominant ones.

Literature provides three different categories of feature selection methods, namely (1) filter methods, (2) wrapper methods, and (3) embedded methods [Blum97 & Liu05]. Filter methods basically use some intrinsic property of the data (textures) in order to select features and do not require knowledge of the learning algorithm to be applied. Wrapper methods, on the other hand, apply the learning algorithm to each feature or feature set and then use the estimated accuracy of the learning algorithm to select relevant features. Finally, embedded methods, as their name suggests, integrate the feature selection process inside the learning algorithm where some features are preferred instead of others and possibly not including all the features available in the learning process [Liu05].

With the high number of features involved a brute-force selection that exhaustively evaluates all possible combinations of the input features to find the best subset can be ruled out straightaway. A wrapper approach is more appropriate for the current research since we want to select features based on how well they can approximate perceptual

similarities. Two selection procedures that are commonly used for feature selection are (1) the forward selection algorithm and (2) the backward elimination algorithm. Forward selection algorithms proceed by adding, at each stage, a feature to a set of features selected at previous stages such that the prediction error is minimised. On the other hand, backward elimination starts with a full set of features and eliminates one feature at each successive stage in order to reduce the prediction error.

Given the large number of features involved, forward selection has been preferred given that the first stage already provides the most dominant feature and fewer iterations would be required to fit the model. Additionally, Cross Validation has been used to alleviate the problem of overfitting.

6.4.3 Cross-Validation

Overfitting occurs when the learning algorithm fits the dataset too well, resulting in poor predictions of unknown samples. To avoid overfitting a hold-out strategy has been commonly used in literature [Kohavi95]. The hold-out strategy involves setting aside instances of a dataset that are not shown to the learning algorithm and using the remaining subset to fit the prediction model. The subset that is set aside is called the test or hold-out set and the one used to fit the model is called the training set.

Cross Validation (CV) is a hold-out approach that has proved reliable in different areas of classification and regression [Kohavi95] and will be considered within the scope of this thesis. Several CV techniques have been detailed in literature; however, we have investigated two specific techniques: ‘*K*-fold CV’ and ‘Leave-One-Out CV’.

K-fold Cross Validation

In *K*-fold CV the dataset is divided into *K* subsets of roughly the same size. The training set is then assigned to all bar one of the *K* partitions with the omitted one attributed to the test set. In such a case the prediction model is tested with data unknown to the training set, but that follows the same distribution as the training set. This process is performed *K* times with each one of the *K* partitions acting as the test set. Hence *K* models are generated with each model producing a different prediction error value. The model with the lowest error is used to represent the dataset.

The prediction error is computed as follows:

$$R(\beta) = \frac{1}{K} \sum_{k=1}^K \frac{1}{M_k} \sum_{i \in S_k} \left(y_i - f(\mathbf{x}_i, \beta^{(-S_k)}) \right)^2 \quad (6.5)$$

where: s_k represents the test data at the k^{th} step (i.e. when the k^{th} partition is used).

M_k is the size of the test set whereas $\beta^{(-s_k)}$ is the set of regression coefficients estimated using the remaining $K-1$ partitions used as training data, i.e. omitting the sample s_k .

The feature that allows the prediction model to generate the smallest error is selected. Using the forward selection process, the K -fold CV is used iteratively to determine a new relevant feature at successive iterations. The learning process is aborted when the addition of new features does not reduce the prediction error by a significant amount. The variant of K -fold CV that is quite popular in literature is the 10-fold CV and is the one that has been implemented here.

Figure 6.3 shows the way in which the dataset is split and also how the training and test data sets are created. The split is done in a sequential manner, independent of the position of the samples in the RPS. Each partition is used as the test set and consequently the remaining partitions are merged to give the training set. In this way, all the samples are used both for training and testing purposes, however no sample is considered for both processes at the same time.

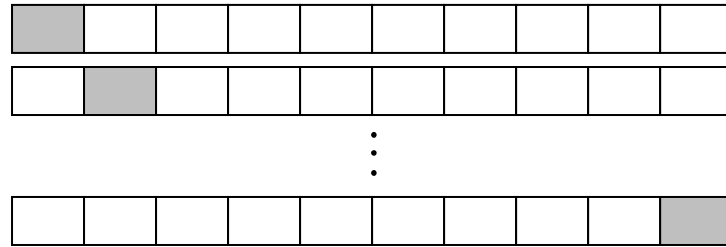


Figure 6.3 – Diagram showing selection of test (shaded cell) and training data for 10-fold CV

Thus we perform feature selection using 10-fold CV combined with forward step-wise feature selection. First we apply 10-fold CV to all possible (3136) single feature predictors. For each predictor the average prediction error is computed over the 10 different test data sets. The feature that generated the minimum average prediction error is removed from the feature pool and retained as x_i^1 (for texture T_i).

Each feature of the reduced pool (3165 features) is now tested in combination with x_i^1 .

That is the task is to determine the second feature x_i^2 from the pool that produces the best predictor $y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2$ for the dimension considered. The feature

selection process is repeated for x_i^3, x_i^4 etc. until the addition of a new feature does not cause significant change to the prediction error.

This process is repeated for each dimension i independently. This is summarised below:

Algorithm for computing prediction error using K -fold CV

For ($n=1 \dots N$) features

For ($k=1 \dots K$) partitions

Select partition k as the test partition

Merge the remaining $k-1$ partitions to create the training set

Apply multiple linear regression to the training set to estimate regression parameters

Use estimated parameters to estimate coordinate (response)

Compute the prediction error from each test set

End

Compute average prediction error for each feature

End

Select the feature with the minimum average error

REPEAT the process for additional 'meaningful' features

REPEAT the whole process for each perceptual dimension

The way in which K -fold CV has been used in this chapter makes it a cheaper option to estimate the prediction error, given that only 10 partitions are considered. Many researchers have adopted this cheap option by either using K successive partitions or K random partitions; however, this option does not guarantee that the test partitions selected would give the best prediction error. In fact, for optimal prediction using K -fold CV, all the different combinations of selecting a test partition should be explored. If we consider the 10-fold CV case and that the Tex1 dataset is being trained, then the number of ways in which a test partition could be selected is $\binom{120}{12}$, where 120 is the size of the Tex1 dataset and 12 is the size of partition. This amounts to a very large number of combinations through which the partition can be chosen, thus clearly indicating the high computational complexity of 10-fold CV.

Increasing the size of the test partition increases the computational complexity. Hence, in order to keep the complexity as low as possible without compromising the

prediction capability of the regression model used, the Leave-One-Out CV has been utilised to learn the prediction model.

Leave-One-Out Cross Validation (LOOCV)

LOOCV is an extreme case of K -fold CV with the number of partitions being equal to the size of the dataset. This implies that each texture sample in the dataset acts as the test data while the rest are used for training. The prediction error is now evaluated using the error function provided in equation (6.6).

$$R(\boldsymbol{\beta}) = \frac{1}{M} \sum_{i=1}^M \left(y_i - f(\mathbf{x}_i, \boldsymbol{\beta}^{(-i)}) \right)^2 \quad (6.6)$$

Hence the prediction error is computed M times, with each texture acting as test data and the remaining $M-1$ as the training data. $\boldsymbol{\beta}^{(-i)}$ is the set of coefficients estimated when the i^{th} sample is omitted from the training set.

Using LOOCV not only helps in preventing overfitting of the prediction model, it also helps to detect and eliminate (or ignore) outliers from the training data. However, correlated features of those already selected at successive stages of the forward selection process will still remain in the list of features to be considered and consequently increases the processing time for the next stages of feature selection.

To eliminate the influence of highly correlated features and to speed up the process, each time a feature is selected, its correlation with the remaining ones is computed. The features that correlate highly with the selected one are removed from the list of features remaining to be evaluated. The Pearson correlation coefficient is evaluated for each pair formed using the selected and remaining features. The Pearson correlation coefficient, ρ_{ij} , between any feature i and another feature j is given by

$$\rho_{ij} = \frac{E(i,j) - E(i)E(j)}{\sqrt{E(i^2) - E^2(i)}\sqrt{E(j^2) - E^2(j)}} \quad (6.7)$$

The algorithm for computing the prediction error using LOOCV and including the correlation test is summarised as follows:

Algorithm for computing prediction error for LOOCV and with removal of highly correlated features

For ($n=1 \dots N$) features

For ($i=1 \dots M$) samples

Select sample i as the test sample

Retain the remaining $M-1$ samples for training

Apply multiple linear regression to the training set to estimate regression parameters

Use estimated parameters to estimate coordinate (response)

Compute prediction error

End

Compute average prediction error for each feature

End

Select the feature with the minimum average error

Remove all the features that have a high correlation with selected feature ($R^2 > 0.7$)

REPEAT the process for additional 'meaningful' features using modified feature list

REPEAT the whole process for each perceptual dimension

The correlation constraint has also been applied to the K -Fold algorithm presented before.

6.4.4 Model Selection to estimating MFS for Tex1 and MoMA data sets

Both CV methods are used to train and validate the two datasets considered, i.e. Tex1 and MoMA. Figure 6.4 shows the average error when the first four perceptual dimensions are trained using the two CV methods. As mentioned before, only one particular setup was used for 10-Fold CV (i.e. only successive partitions). The learning process was run to select a maximum of 10 relevant features. The error for LOOCV is considerably less than that for 10-Fold, suggesting that the “best” possible subsets are not chosen for the 10-Fold test and training sets. LOOCV was chosen for the learning process on the basis that it generates lower prediction errors and does not require heavy computational capabilities.

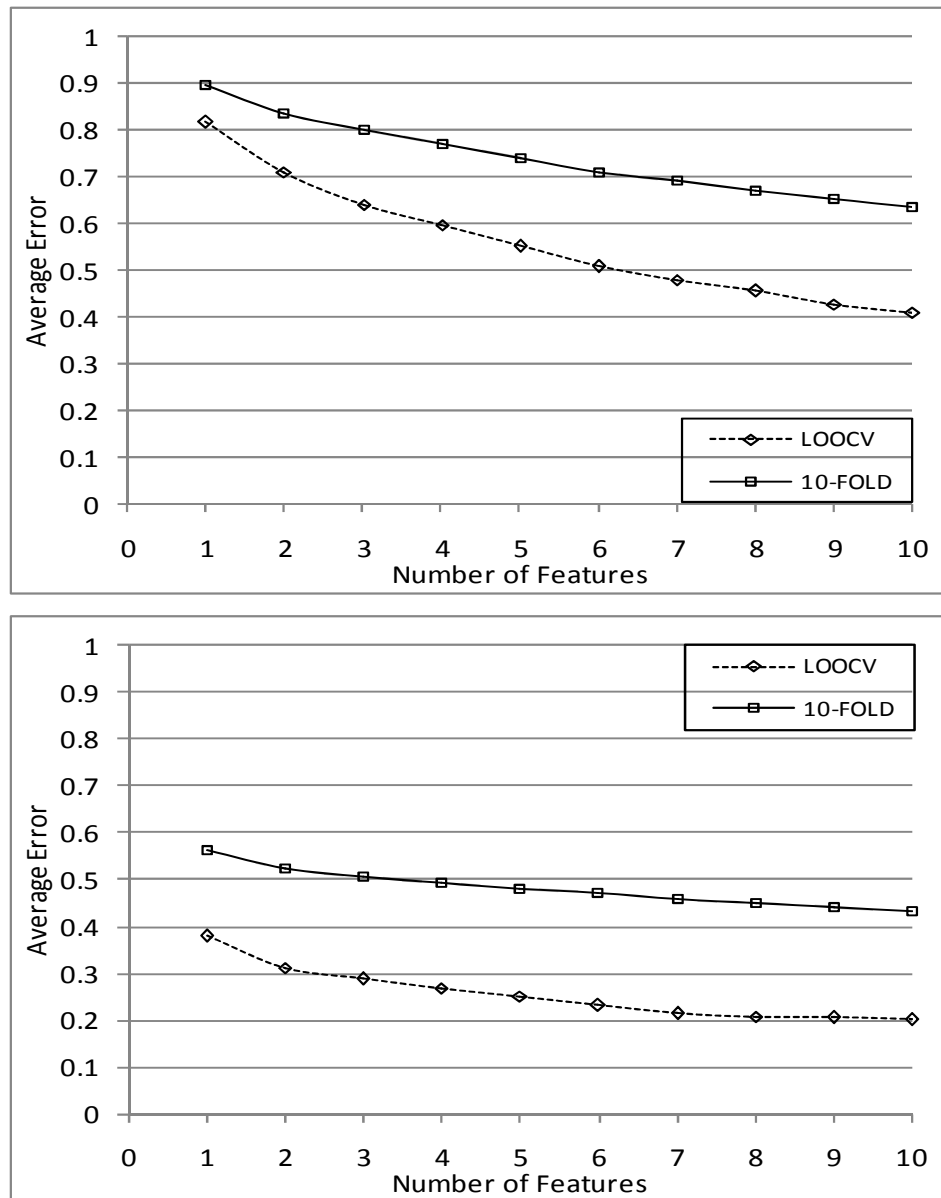


Figure 6.4- Average error for LOOCV and 10-Fold CV methods when training the Tex1 dataset (top) and MoMA dataset (bottom).

Applying the above algorithm to each perceptual dimension results in a model that estimates the RPS coordinates for each dimension. The outputs are:

- a list of relevant features (predictors); and,
- regression coefficients for each model.

Hence, assuming that a texture dataset can be represented by d perceptual dimensions, then a d dimensional MFS can be mapped to the perceived similarities using a set of d regressors (linear equations) in the following way:

$$\left. \begin{aligned} y_i^1 &= \beta_0^1 + \beta_1^1 x_i^{a0} + \beta_2^1 x_i^{a1} + \dots + \beta_{f_s}^1 x_i^{an} \\ y_i^2 &= \beta_0^2 + \beta_1^2 x_i^{a0} + \beta_2^2 x_i^{a1} + \dots + \beta_{f_s}^2 x_i^{an} \\ &\vdots \\ y_i^d &= \beta_0^d + \beta_1^d x_i^{a0} + \beta_2^d x_i^{a1} + \dots + \beta_{f_s}^d x_i^{an} \end{aligned} \right\} \quad (6.8)$$

The set of equations (6.8) thus provides d prediction models where each model is used to estimate the coordinates $y_i^1 \dots y_i^d$ of the texture i in the individual dimensions. f_s is the number of features selected to map the MFS to the perceived similarities, whereas $a0, a1, \dots, an$ are the indices of the selected features chosen from the large pool of f features extracted.

6.4.5 Determine the number of features for MFS mapping

As in the case of the RPS dimensionality, the length of the feature vector used to compute the y_i values for each MFS dimension is determined using the precision of retrievals. Figure 6.5 (a) and (b) show how precision rates vary with the number of features used to approximate a 4D RPS for the Tex1 and MoMA sets respectively.

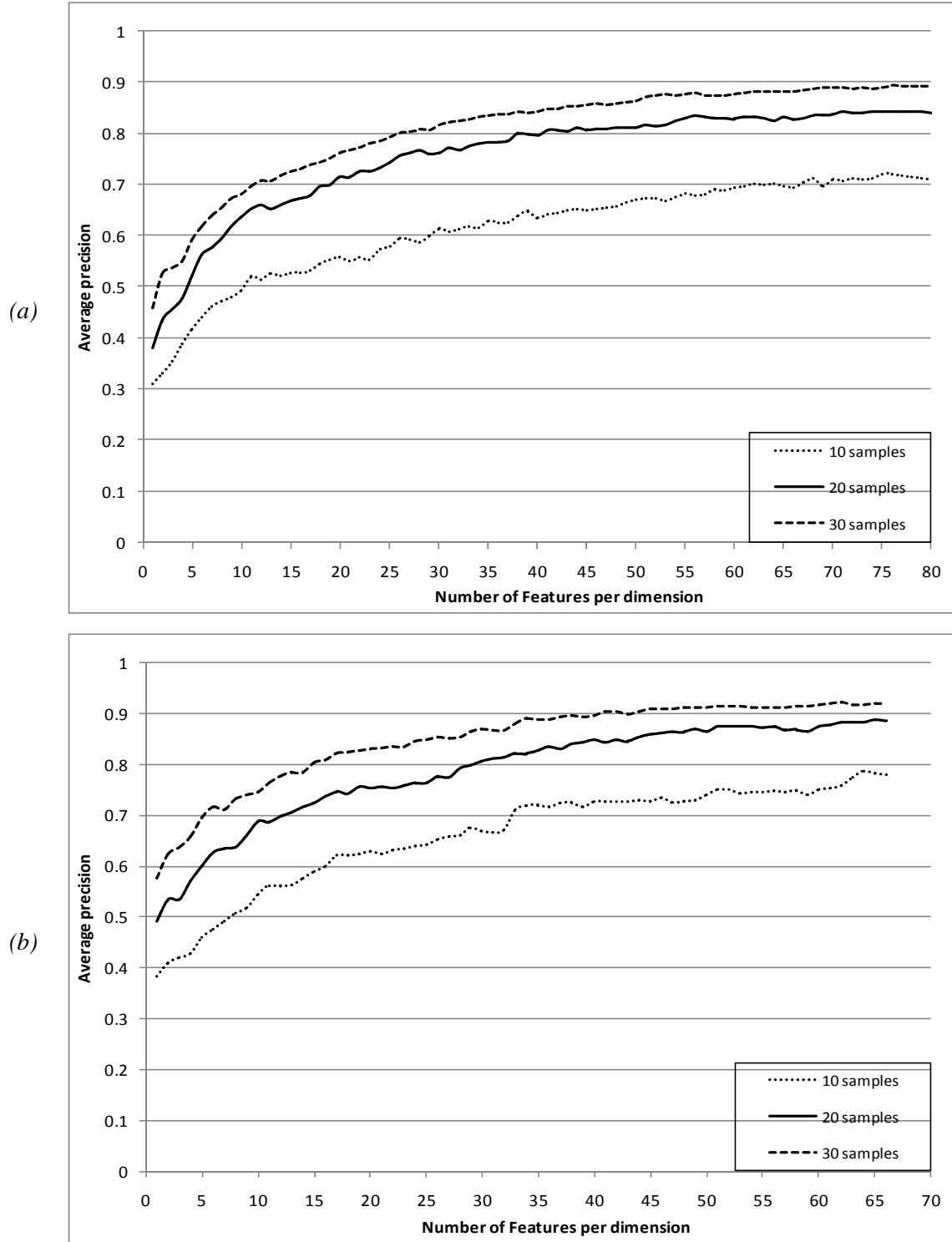


Figure 6.5- Variation of precision with increasing number of features used to approximate a 4D RPS for (a) the Tex1 dataset and (b) the MoMA dataset

We notice from the figures that the precision rate stabilises at a relatively high number of features for both datasets. This demonstrates how difficult it is to find representative sets of features for the datasets, even though a very large pool of Trace Transform features is available.

However, the high prediction rates achieved show that the prediction model presented in section 6.4 allows a good approximation of the RPS. It can be observed from Figure 6.5(a) that the precision rate starts to slowly decline when the number of features is increased beyond seventy-five, as in the case of 10 retrieved samples. Thus feature vectors containing 75 elements will be used to map the Tex1 textures from the RPS to the MFS. Similarly, Figure 6.5(b) indicates that a feature vector of length 65 provides the best precision rates for the MoMA textures.

6.4.6 Mapping results

Figure 6.6 and Figure 6.7 show scatter plots of RPS coordinates against MFS coordinates for dimensions 1 to 4 for the Tex1 and MoMA datasets respectively. Each point within the scatter plots represents a texture sample. A prediction line is provided to allow the reader to qualitatively assess the degree of fit between the RPS and the prediction model for each dimension.

It can be seen that, in the case of the Tex1 dataset, the fit for all four RPS dimensions improves as the number of features used to approximate each dimension is increased from one, to twenty-five and then to a maximum of seventy-five. This shows a high level of variability within the Tex1 texture samples that cannot be captured with only two or three main dimensions.

The same cannot be said for the MoMA textures. It can be seen that the fit for the first RPS dimension, given by row 1 from Figure 6.7, increases considerably as the number of features is increased to a maximum of sixty-five. However, the fit for the other three dimensions is not so clear due to the poor distribution of samples within the spatial configuration considered. In contrast to the Tex1 dataset, increasing the RPS dimensionality for the MoMA dataset does not capture significant variation among the textures. The distributions of the samples appear to be more compact for dimensionality two and higher. This can be shown by the length of the prediction line which decreases as the dimensionality of the MoMA RPS increases.

With considerable variation among the textures present at high RPS dimensions for the Tex1 dataset, it can be deduced that one or two dimensions are not enough to map the perceived similarities for the texture samples. For the MoMA dataset, most of the variability within the textures is captured by the main dimension indicating a high level of compactness. This observation was expected as the textures provided by MoMA are specialised and similar.

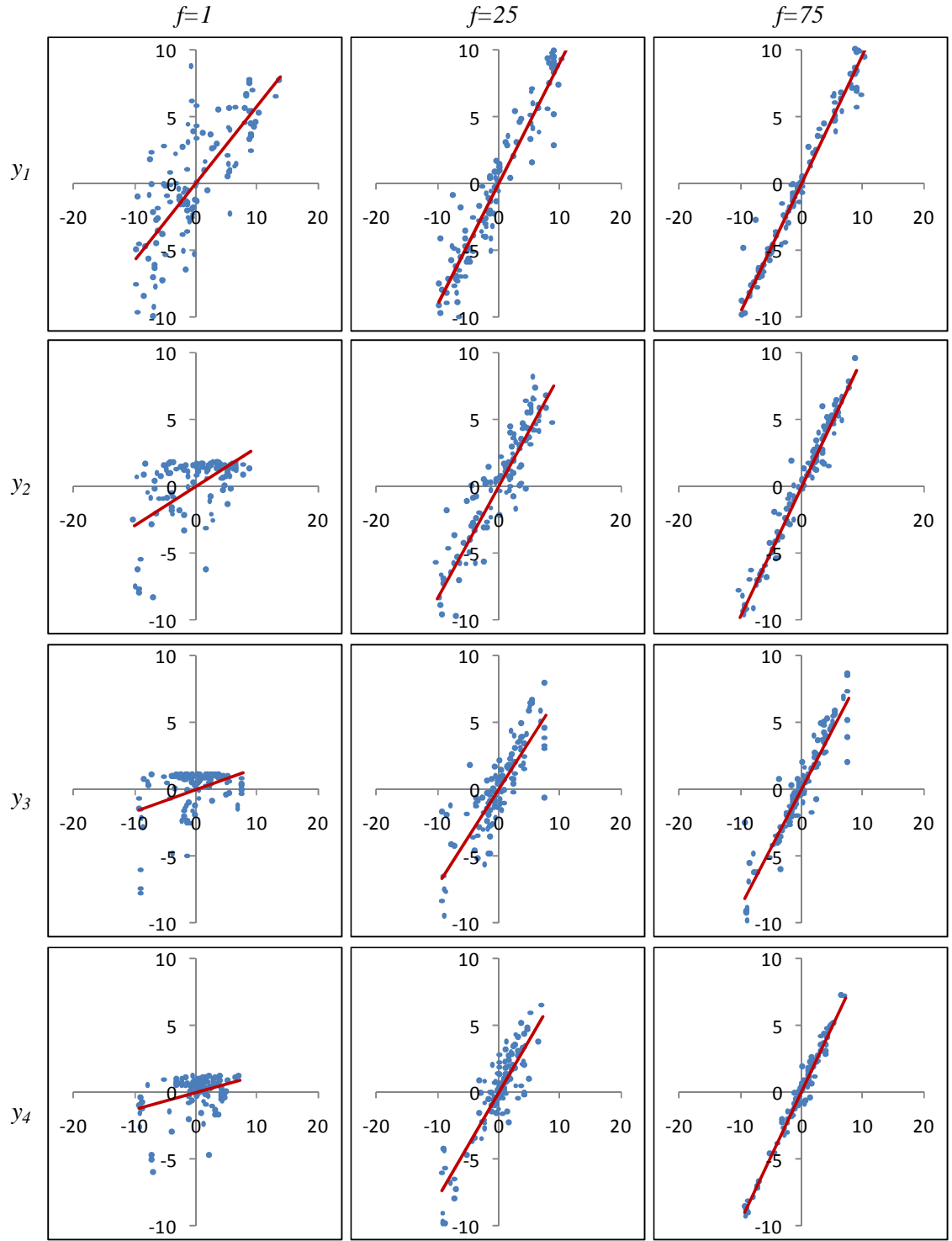


Figure 6.6 – RPS (vertical axes) vs. MFS (horizontal axes) values for *Tex1* texture samples (rows correspond to the first four dimensions, columns 1,2,&3 correspond to increasing number of features - 1, 25 and 75 respectively).

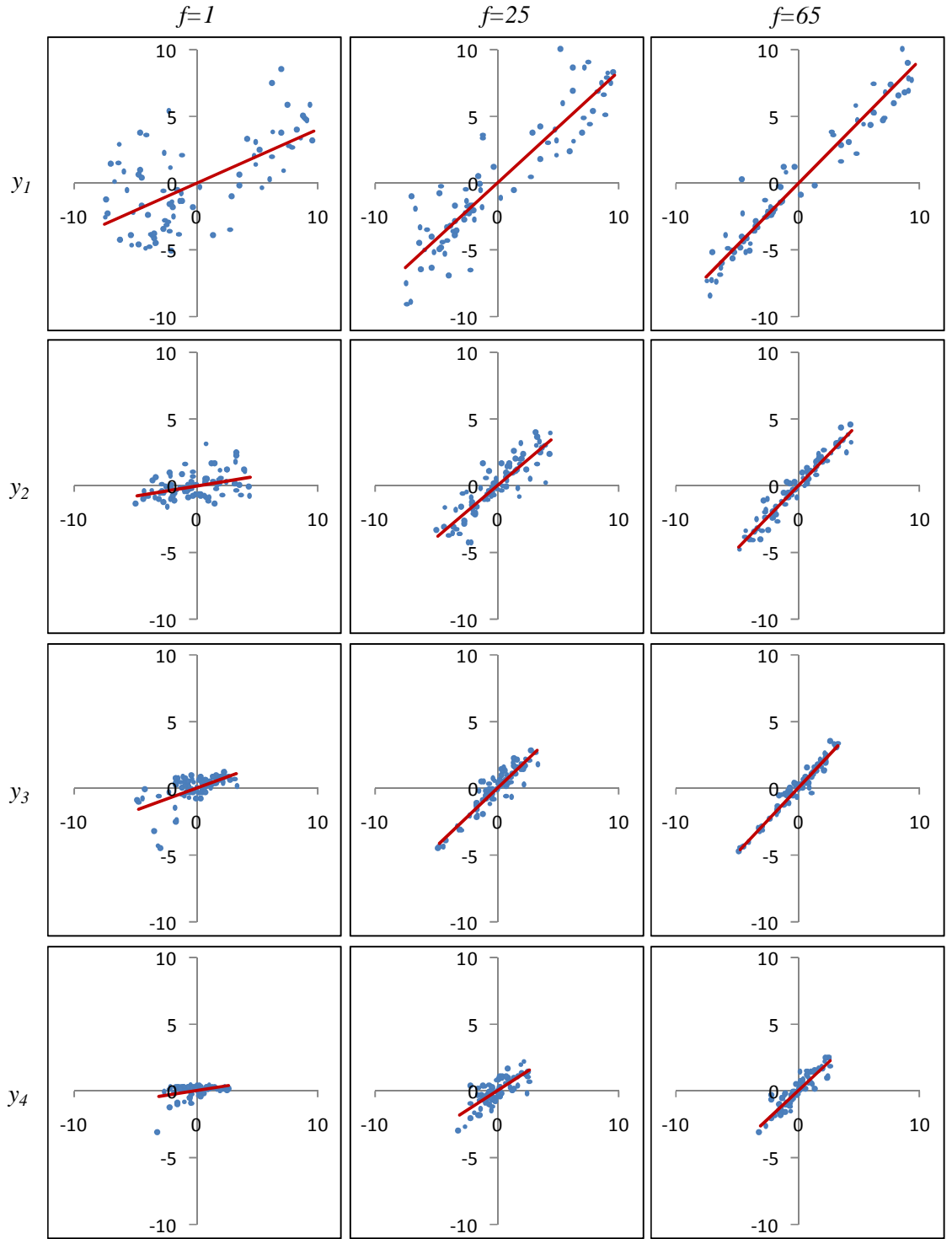


Figure 6.7 – RPS (vertical axes) vs. MFS (horizontal axes) values for MoMA texture samples (rows correspond to the first four dimensions, columns 1,2,&3 correspond to increasing number of features - 1, 25 and 65 respectively).

6.5 Stages III & IV – Query Feature Vector Calculation and Texture Retrieval

The preceding sections have described how we derived feature spaces (MFS) for both the MoMA and Tex1 data sets. Performing retrieval in response to a query now is straightforward.

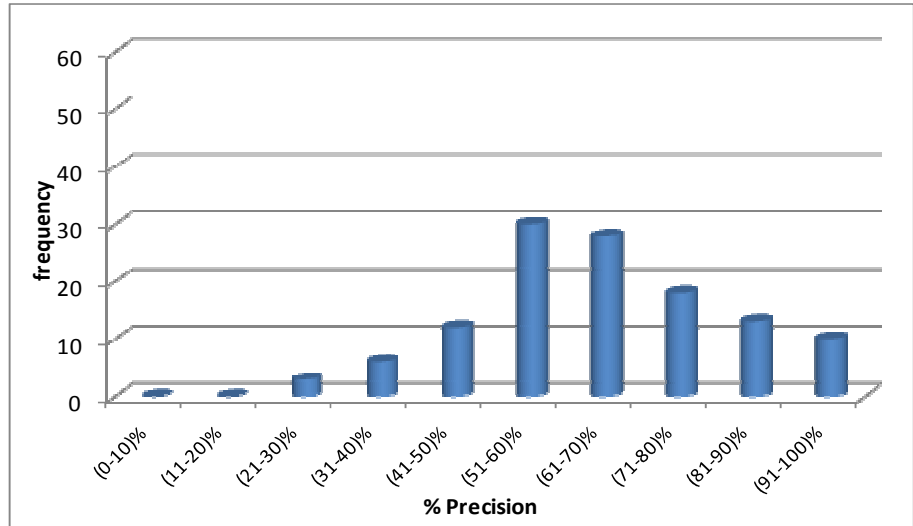
1. The required trace transform features are extracted from the query texture and normalised using the parameters applied to the whole dataset.
2. The y_i value of each the axes of the MFS is calculated using equation 6.2 to provide the query feature vector \mathbf{y}_q .
3. The n nearest textures to \mathbf{y}_q in the MFS are identified and returned using their Euclidean distances.

The retrieval model is evaluated by investigating how well it responds when presented with a query texture. No blind testing has been performed since the textures used for testing the model have also been used in the training stage.

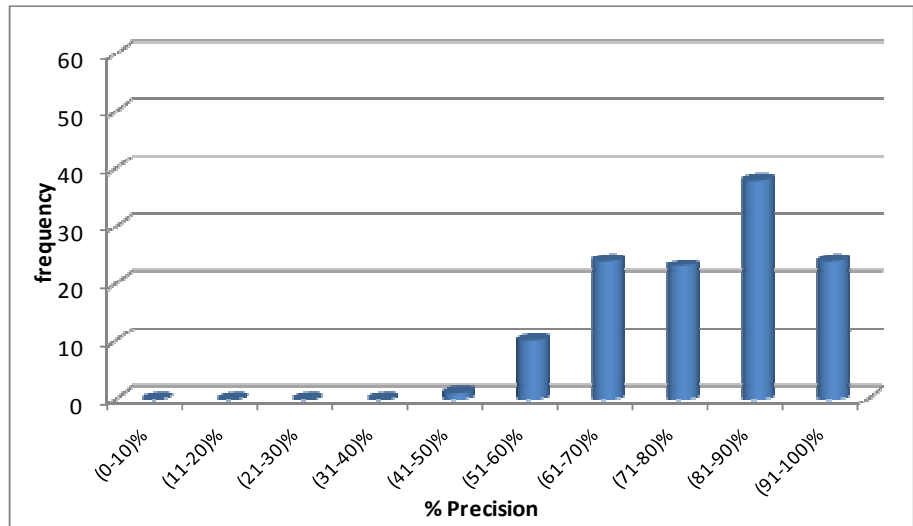
To test the efficiency of the retrieval model, the results for retrievals from the MFS are compared with the texture groupings performed by observers. In section 6.2, precision is presented as a measure to determine the dimensionality of the RPS. Using the same measure and the same procedure for its computation as in section 6.2, we investigate how well the Trace Transform features have been able to map the MFS for the purpose of retrieval. Each texture from the different datasets is used as a query and the search is performed in both the MFS and the FPS. The retrieval modes corresponding to 10, 20 and 30 textures being retrieved are outputted for comparison. Note that the order of the textures within the retrieved lists has not been given significant importance.

Dealing with a high number of dimensions involves intensive computational processing. With the considerably large number of features required to approximate the FPS of the two datasets the prediction model becomes extremely heavy computationally. For practical reasons, the MFS configurations, used to compute the average precision rates, are based on a fixed number of dimensions and varying the number of features up to a maximum of 75 for the Tex1 dataset and 65 for the MoMA dataset. A 4D MFS is used for both datasets. Figure 6.8 and Figure 6.9 illustrate the histograms of average precision values for the Tex1 and MoMA textures when the first 10, 20 and 30 textures retrieved from both the FPS and the MFS and are compared. The average precision values are binned within a 10-bins precision histogram.

10
samples



20
samples



30
samples

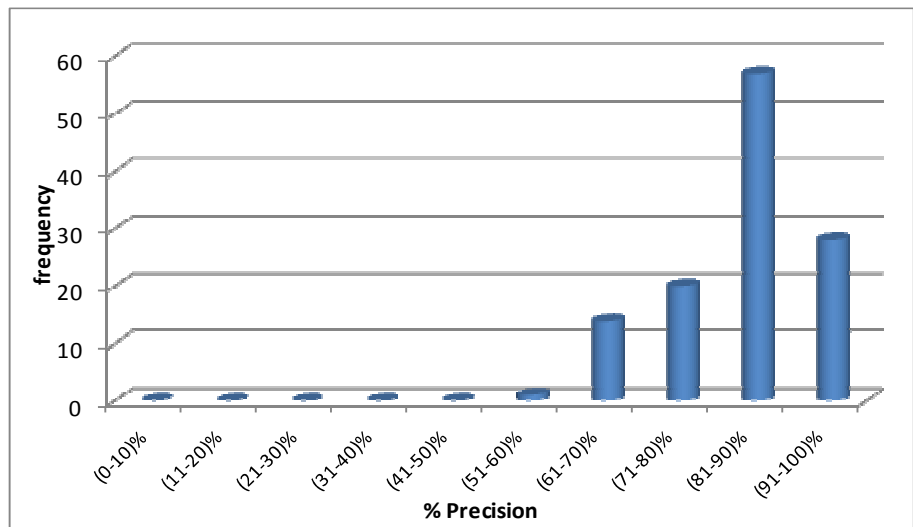
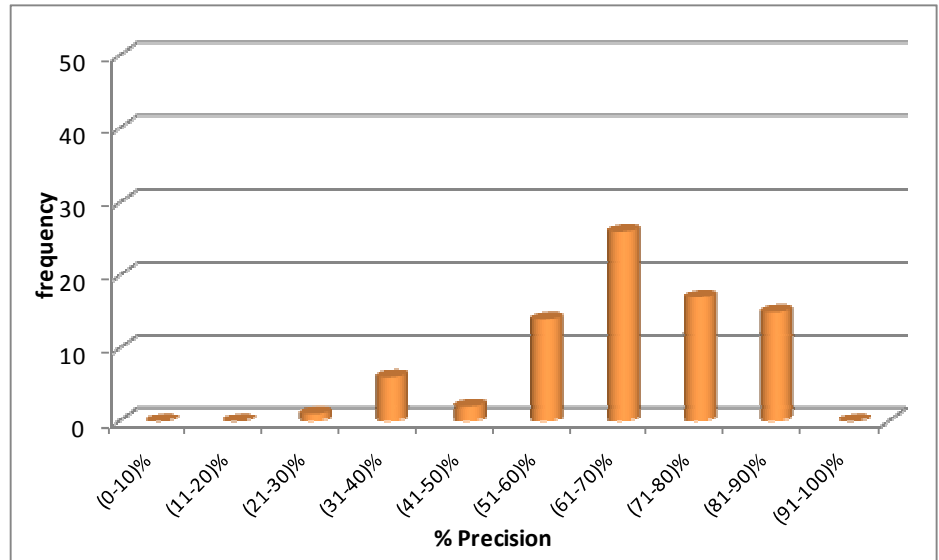
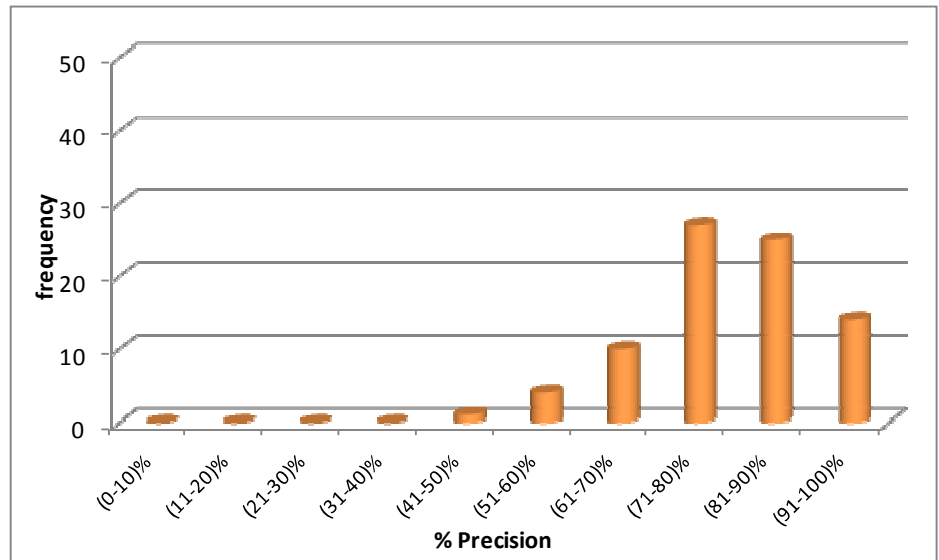


Figure 6.8 – 10-bins precision histograms for 10, 20 and 30 retrieved samples for Tex1 textures.

10
samples



20
samples



30
samples

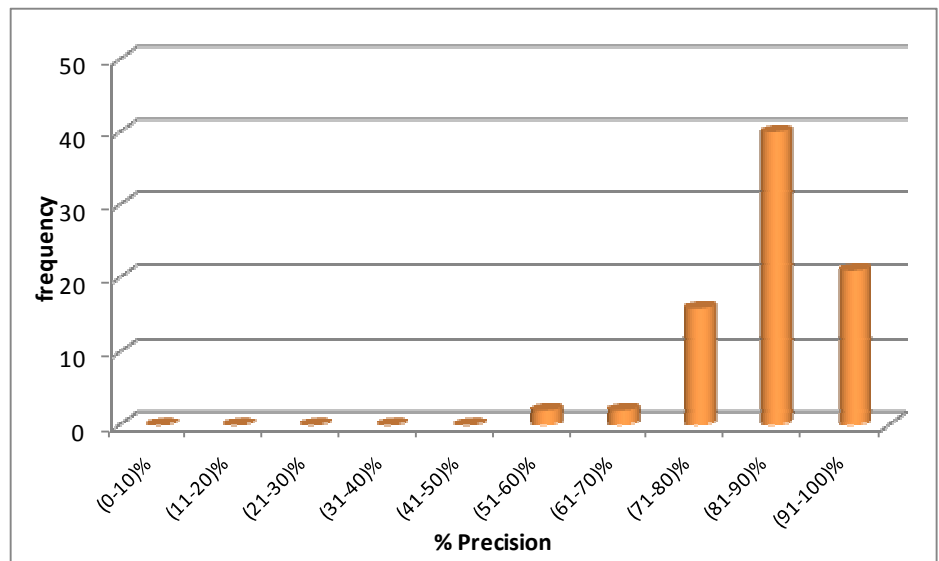


Figure 6.9 – 10-bins precision histograms for 10, 20 and 30 retrieved samples for MoMA textures.

When only 10 retrieved Tex1 textures are considered, the majority of the query textures achieve a precision rate within the range (51-60) % as indicated by the precision histogram in Figure 6.8. Increasing the number of retrieved samples skews the histograms towards higher precision ratios for most of the textures: (81-90) % in the case of 20 samples and 30 samples are retrieved. Precision histograms for the MoMA dataset, displayed in Figure 6.9, show higher levels of precisions for the MoMA query textures. This is an expected observation as in the previous section it was clearly shown that the prediction model provides better fit for the MoMA textures.

The histograms give an indication of how well the dataset on the whole performs within the retrieval framework presented in this chapter. Analysing the histograms can provide information of which textures were difficult to represent and to whether the information is consistent with the way humans have grouped the textures. The lower quartile, median and upper quartile values are computed from the precision value distributions for the different retrieval modes and are presented in Table 6.1 and Table 6.2 for the Tex1 and MoMA datasets respectively.

	Lower Quartile			Median			Upper Quartile		
No. of samples	10	20	30	10	20	30	10	20	30
Value	0.508	0.566	0.726	0.587	0.763	0.829	0.710	0.855	0.874
Textures	3, 87	55, 84	26,61	13, 81	34, 57	21, 37	27, 64	115, 116	96, 105

Table 6.1 – quantitative analysis of the precision value distributions for the Tex1 dataset

	Lower Quartile			Median			Upper Quartile		
No. of samples	10	20	30	10	20	30	10	20	30
Value	0.586	0.713	0.800	0.668	0.788	0.861	0.759	0.849	0.900
Textures	112, 1390	28, 922	250, 2233	193	2394	1439	321, 1419	1954, 2607	1413, 1765

Table 6.2 – quantitative analysis of the precision value distributions for the MoMA dataset

The lower quartile mark is a good indication of the textures that the retrieval model was not able to map properly to the MFS, hence leading to low precision values with respect to corresponding retrievals in the FPS. Similarly, the upper quartile mark represents textures whose distances with other textures within the FPS, have been preserved when mapping the MFS onto the RPS. The median value on the other hand, represents how skewed the precision histograms are. Positively skewed histograms indicate high precision rates for the majority of textures queried.

As expected, the precision histograms for the MoMA dataset are more highly skewed than those of the Tex1 dataset. The smaller difference of the interquartile range of precision values for the MoMA texture as compared to the difference for the Tex1 textures also indicate better mapping for the MoMA dataset. Even though, the upper quartile values for both datasets are very close when large numbers of retrievals are considered (30 samples for example). This implies the ordering of the similarity (or dissimilarity) values are better preserved in the case of MoMA textures leading to higher precision rates at lower number of samples.

An investigation into the average precision values of the different Tex1 groups presented in Chapter 4 shows that group 3 (patchy textures) and group 5 (circular textures) were the most difficult to encode in the 4D MFS considered. Table 6.3 shows the average precision values for the different groups.

	<i>Average precision values</i>		
Groups	10 Samples	20 Samples	30 Samples
Group 1 (regular)	0.717	0.764	0.867
Group 2(irregular)	0.699	0.740	0.863
Group 3(patchy)	0.639	0.709	0.732
Group 4(vertical)	0.772	0.821	0.851
Group 5(circular)	0.646	0.692	0.719
Group 6(horizontal)	0.751	0.790	0.804

Table 6.3 – Average precision for Tex1 groups

Even though observers had no difficulties in grouping the circular and patchy textures, the features used to create the 4D MFS were not able to encode these textures with much precision, hence resulting in the low precision rates when textures from these two groups are retrieved.

6.5.1 Retrieval results

To illustrate the nature of the retrievals that the prediction model has performed, 20 texture retrievals representing the lower quartile, median and upper quartile results are shown in Figure 6.10, Figure 6.11, Figure 6.12 and Figure 6.13.

The results displayed correspond to the nearest neighbours of the query textures within a 4D MFS generated using feature vectors containing seventy-five elements for the

Tex1 dataset, and sixty-five features for the MoMA dataset. The texture images used as queries for lower quartile, median and upper quartile results for Tex1 are textures T55, T34 and T116 respectively. Table F.1 to Table F.4 in Appendix F provide the query results for the first 30 textures for reference.

The retrievals from the MFS are presented together with the “ideal” retrievals from the FPS for comparison purposes. The textures retrieved from the MFS that match the ones retrieved from the FPS are highlighted. Those textures retrieved from the FPS and missing from the MFS retrievals have italic, bold and underlined labels. As mentioned previously, the order in which the textures appear is not considered while determining the precision of the retrievals. The observations from each set of results are as follows:

I. Query Texture T55 (lower quartile result)

Eleven out of the twenty retrieved textures match (see Figure 6.10). This represents a precision value of 0.55. The results suggest that humans grouped the textures similar to T55 based on some circular structural information available either globally or locally.

The textures retrieved from the FPS were irregular or isotropic in most cases. Texture T55 also has the appearance of a smooth or “polished” surface, hence explaining the presence of textures T81, T42, T10 or even T43 within the observers’ results. However, despite non-blind testing, it would seem that the selected features cannot detect the pertinent longer range interactions.

II. Query Texture T34 (median result)

Using texture T34 as a query texture successfully retrieves 14 matching textures. This indicates a precision value of 0.7. The higher precision is perhaps due to the bi-directional nature of these textures.

III. Query Texture T116 (upper quartile)

T116 has a precision of 0.85 (17 matches). As illustrated in Figure 6.12, the retrievals consist mainly of directional (horizontal) textures, forming one of the dominant groups that observers were quick to assemble while performing the psychophysical experiment. The high precision for this texture is likely to be due to the ease with which directional textures can be detected by texture features.

Table 6.2 shows high average precision rates for the MoMA textures as indicated by values of the lower quartile, median and upper quartile for 20 textures retrieved (0.71,

0.79 and 0.85 respectively). Even though the precision histograms are highly skewed, the upper quartile value for the MoMA textures approximates the upper quartile value for the Tex1 dataset.

For the MoMA results, the smaller interquartile range suggests that more relevant textures are available at lower retrieval modes (10 or 20 samples for example). To illustrate this point the first 20 textures retrieved from both the FPS and the MFS, when MoMA texture M1954 is used as a query, is presented in Figure 6.13. Matching the retrieved textures from both the MFS and FPS leads to a precision value of 0.95. In addition to the high precision rate, the order in which the textures are retrieved is more consistent than with the Tex1 textures.

The missing textures from the FPS give an indication of which textures were difficult to encode in the MFS. When Tex1 texture T55 is used as query, nine textures from the FPS were missed (underlined-italic-bold labels). A visual investigation of these textures shows that they all contain global information that observers were quick to pick out but difficult to be encoded using computational features. Using texture T34 as query resulted in less misses. As compared to the near-regular and coarse nature of T34, the textures that were missed appear to be finer and regular. With T116 as query, only 3 textures were missed. In this case the directionality of the query texture seems to be the dominant attribute in encoding the textures. Since T116 is a unidirectional texture, orientated in a horizontal direction, the misses from this query are basically those that lack this dominant horizontal directionality.

For the MoMA dataset, using texture M1954 as query results in only 1 miss. With less variation among the MoMA textures and very few textures containing structural information, it is not surprising to see that the set of trace transform features provided better encoding of these textures.

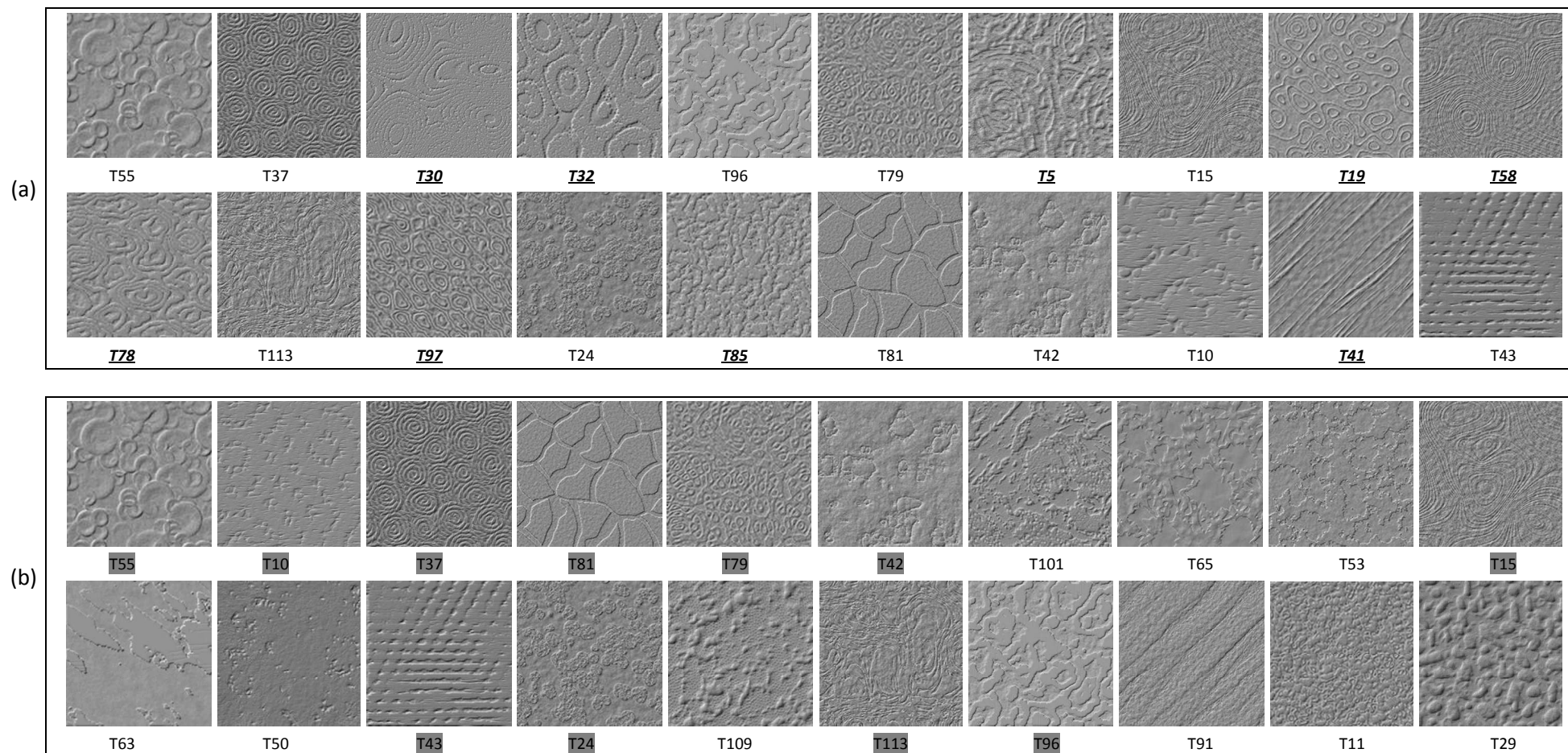


Figure 6.10- First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture T55(lower quartile) as query

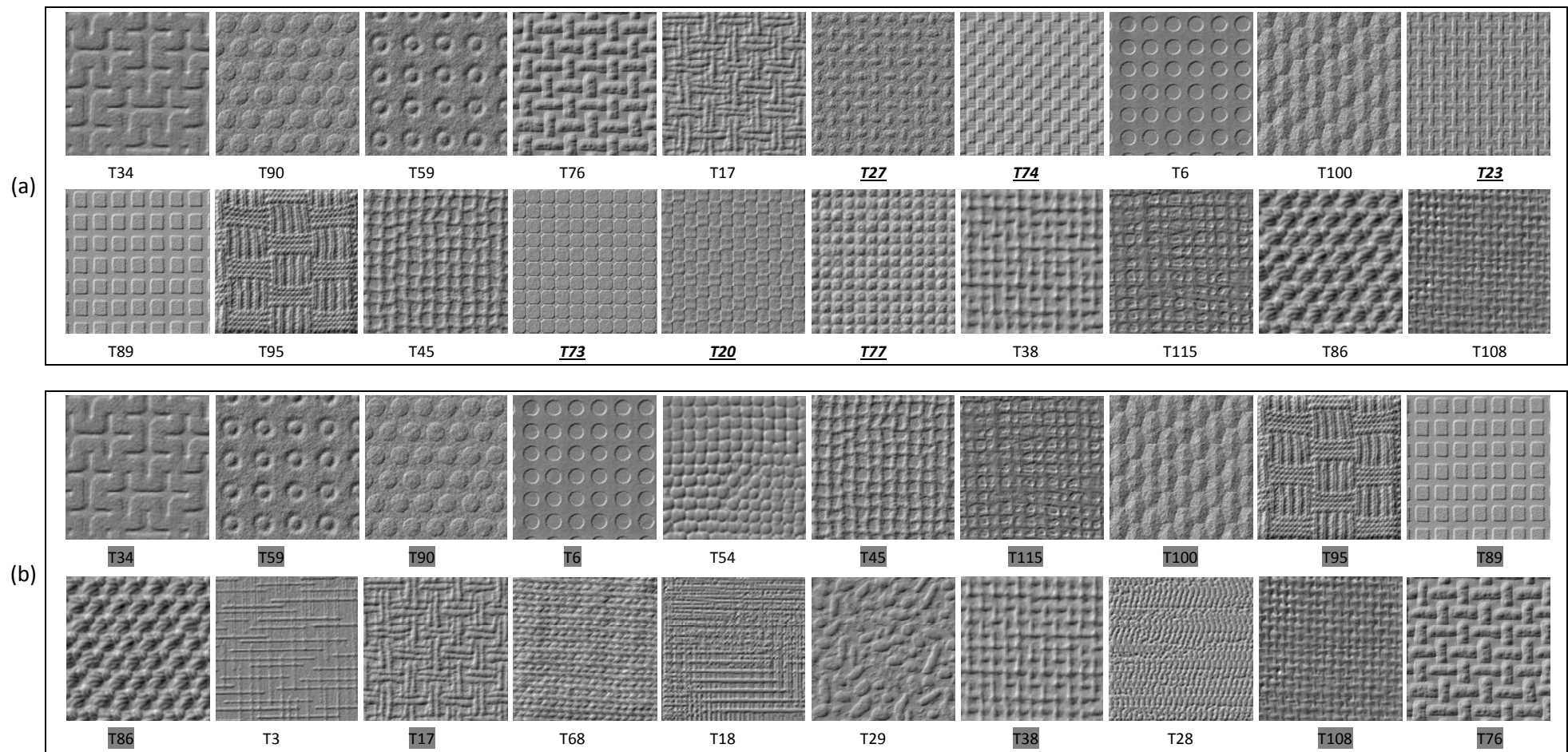


Figure 6.11-First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture T34(median) as query

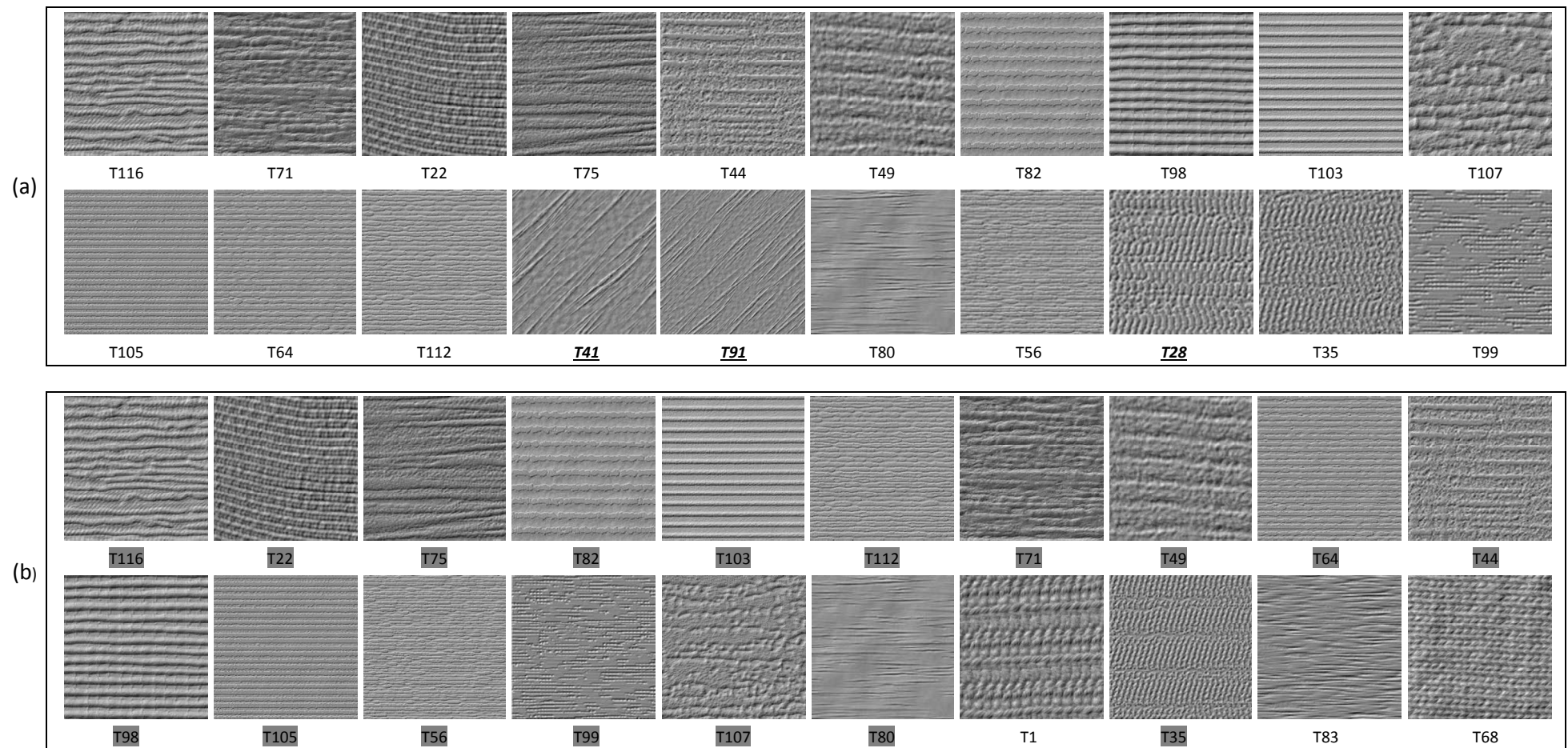


Figure 6.12- First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture T116(upper quartile) as query

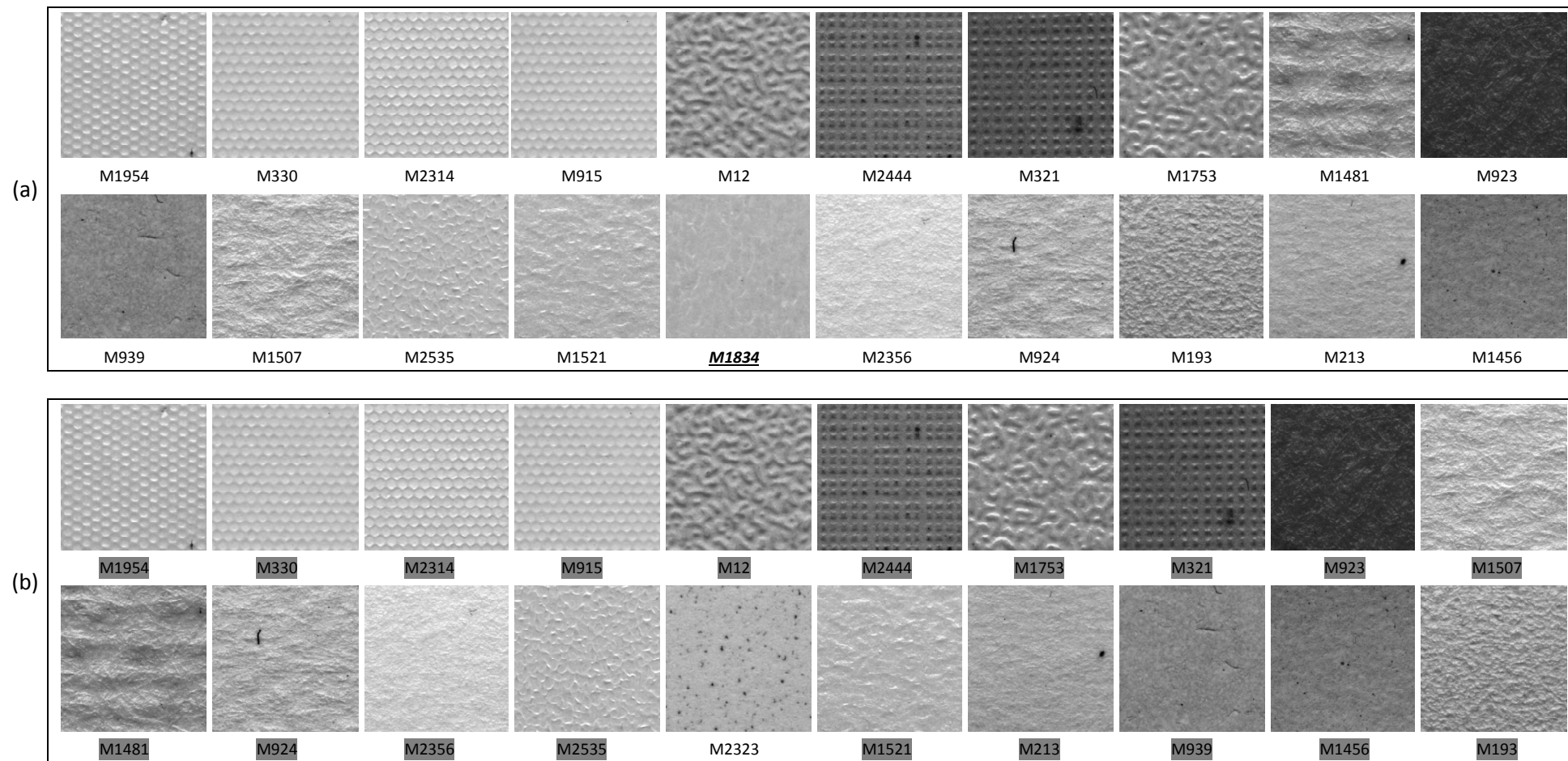


Figure 6.13- First 20 retrievals from (a) the FPS and (b) the MFS feature space using texture M1954(upper quartile) as query

6.6 Blind Testing

We have demonstrated in the previous section that when the prediction model is tested using samples already used in the training stage, the precision increases gradually with increasing number of features (refer to Figure 6.5). Precision rates of approximately 90% (in the case of 30 samples) are obtained for both the Tex1 and the MoMA datasets. However, the number of features required to fit the RPS to the MFS in order to obtain such precision rates is high: 75 features in the case of the Tex1 textures and 65 for the MoMA ones. Given that each dimension of the MFS is fitted independently, this implies that to fit a 4D MFS, the total number of features that would be required, in the worst case, is 300 for Tex1 and 260 for MoMA.

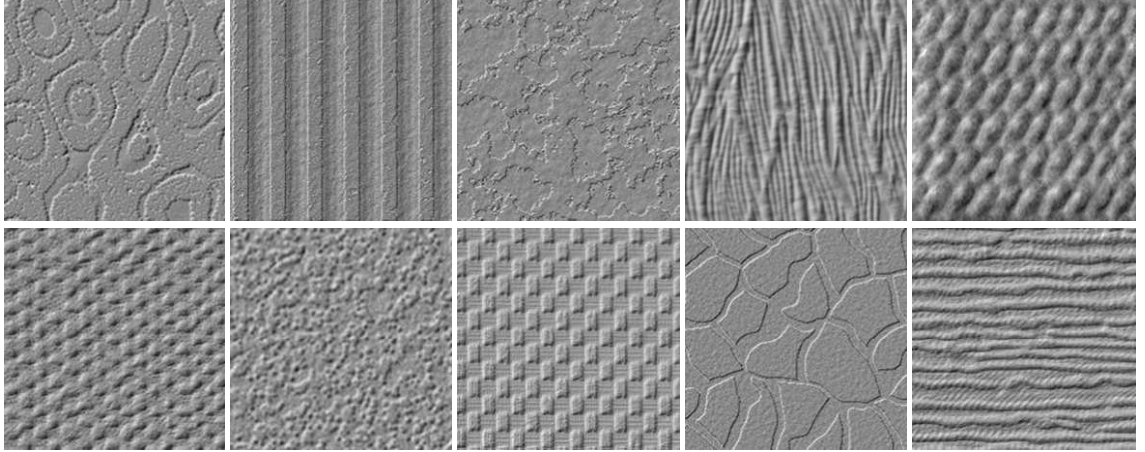
Since the sizes of the datasets are 120 and 81 respectively, we are therefore exposed to the fact that the datasets have been projected to a higher dimensional space and more than one feature is being used to fit each texture sample. This is clearly a situation of overfitting, hence explaining the excellent precision rates obtained. This section thus explores how well the retrieval model proposed in Section 6.2 fares when tested with “unknown textures”.

We therefore decided to repeat the previous experiments while withholding a proportion of the textures for use as test data.

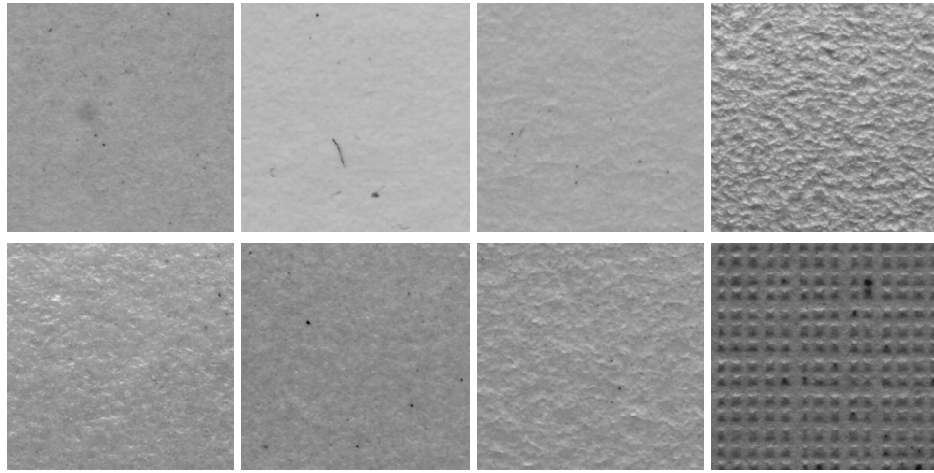
6.6.1 Test Sample Selection

The dendrograms produced in Chapter four were used to ensure that the “test” textures were reasonably distributed across the perceptual space. This was done by using the dendrograms to create n partitions and then randomly choosing one “test” sample from each partition.

The value for n was chosen approximately to be 10% of the dataset (10 textures for the Tex1 dataset and 8 for the MoMA one). The selected textures for Tex1 and MoMA are shown in Figure 6.14 (a) and (b) respectively.



(a) *Tex1* test textures



(b) *MoMA* test textures

Figure 6.14 – (a) *Tex1* and (b) *MoMA* test textures randomly selected from 10 and 8 groups obtained by applying the cluster analysis to the two datasets

6.6.2 Effect of varying the number of features (per dimension)

Figure 6.15 shows how precision varies when retrievals are performed using the test textures in Figure 6.14 for different number of features. The precision rates are significantly lower than the ones presented in Figure 6.5.

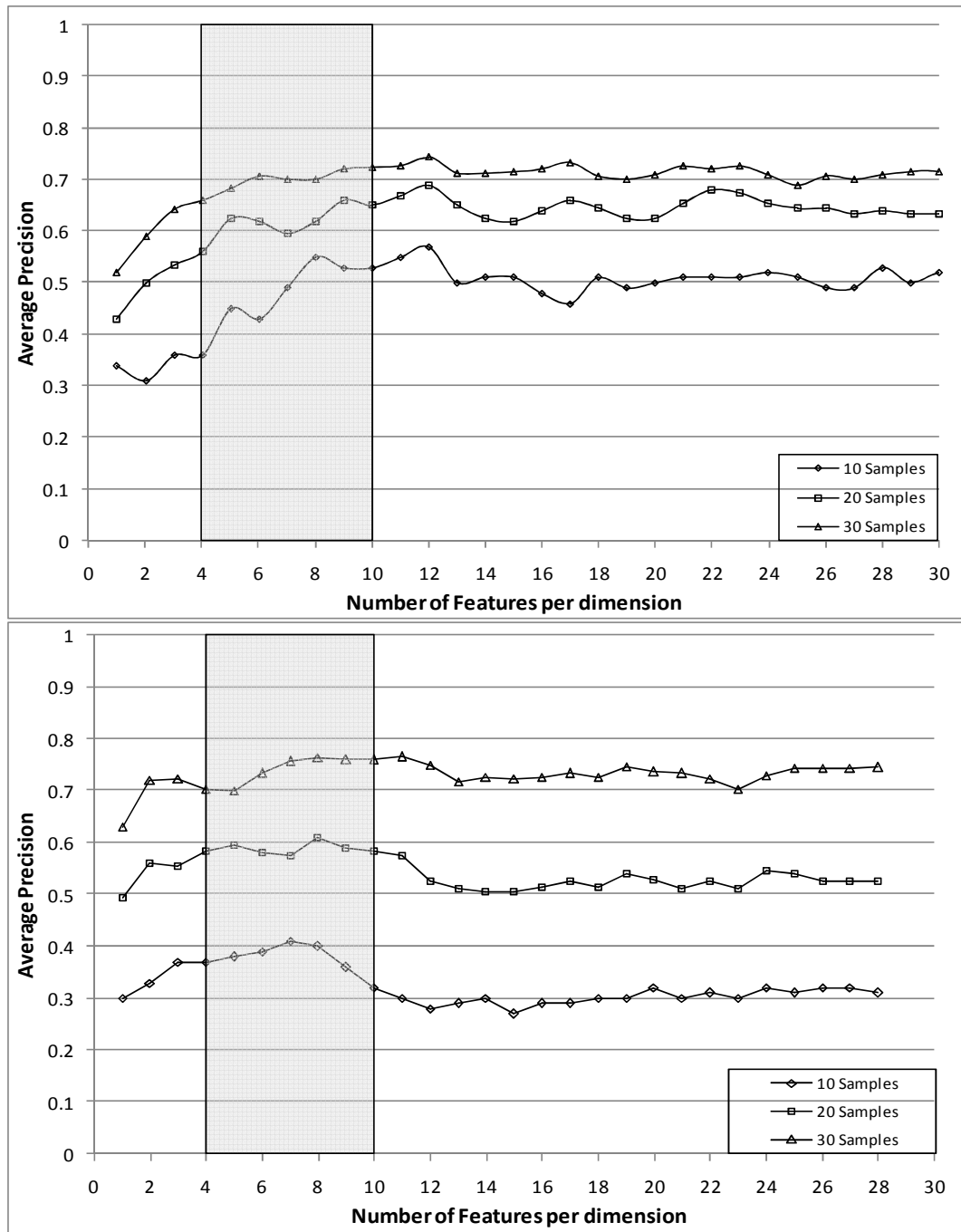


Figure 6.15 – Variation of precision with increasing number of features for (top) *Tex1* test textures and (bottom) *MoMA* textures (blind test)

The behaviour of precision indicates that the number of features to map the datasets for a 4D RPS is not obvious; however, we can observe that the use of more than 10 texture features for both the *Tex1* and *MoMA* datasets does not improve performance. The outlined areas on the two graphs of Figure 6.15 indicate the range of features that can be used to create effective application.

6.6.3 Results – further details

Table 6.4 shows how the average precision rates compare, when the 10 Tex1 test textures are applied for retrieval within the 3 different perceptual spaces: the FPS, the MFS(NB) – no blind testing, and the MFS(B) generated without using the test images (blind testing). The precision rates correspond to the first 20 samples retrieved and they are averaged for retrieval performed from 4D spaces generated using 1 to 4 features per dimension in the case of MFS(NB) and MFS(B).

Texture	FPS	MFS (NB)	MFS(B)
T83	0.850	0.388	0.325
T66	0.900	0.488	0.363
T60	0.950	0.375	0.425
T53	0.900	0.413	0.450
T62	1.000	0.475	0.450
T33	1.000	0.500	0.463
T116	0.800	0.500	0.500
T69	0.950	0.563	0.625
T32	0.950	0.700	0.688
T74	0.950	0.488	0.688

Table 6.4 – Comparative figures for average precision rates of Tex1 test textures in the 3 different perceptual spaces.

The textures in Table 6.4 are sorted according to the blind testing precision – column 3 in table. The values in bold represent the maximum precision rates for each space and the highlighted area denote the median precision for the MFS(B).

The T62 (see Figure 6.16) twenty texture retrievals for the idealised FPS case and for the automated system (the MFS case) are shown in Figure 6.17 for illustration purposes.

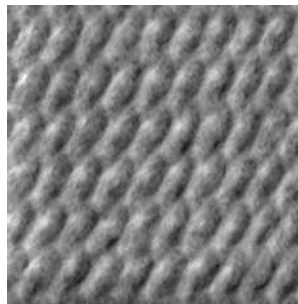


Figure 6.16 – Test texture T62

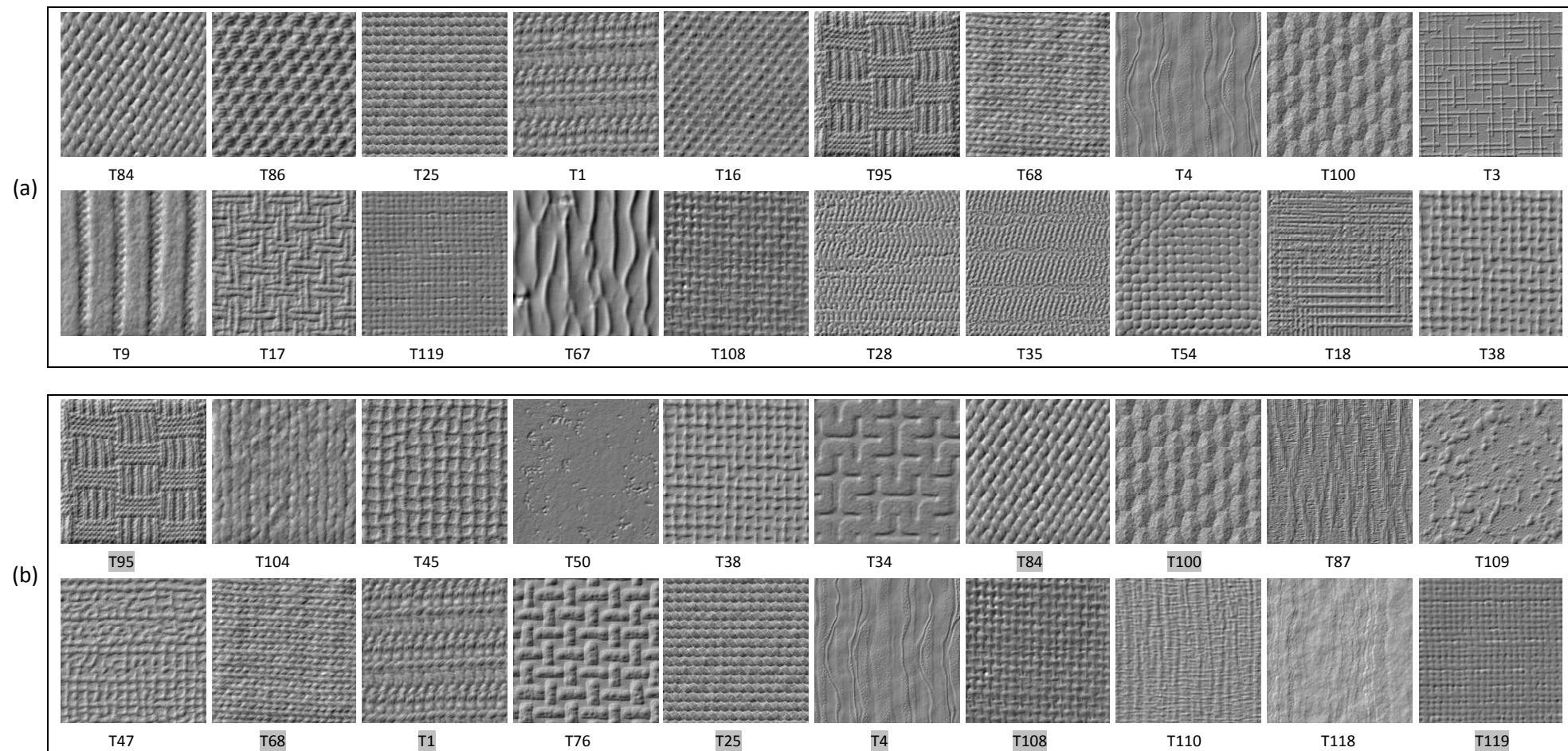


Figure 6.17- First 20 retrievals from (a) a 110D FPS and (b) the MFS feature space using texture T62 as query

6.7 Summary for Part I

The first part of this chapter has investigated how reduced perceptual spaces could be exploited for retrieval purposes. We have initially demonstrated that mapping the FPS to 4D RPS for both datasets could potentially allow retrieval systems to achieve higher than 90% success rates.

Retrievals are performed on lower dimensional feature spaces. So far this chapter has presented a retrieval model that uses a simple linear regression model to map the 4D RPS obtained through MDS to feature spaces (the MFS) of the same dimensionality. By using the whole datasets to train the retrieval model, we have shown that even by using around 75 features for the Tex1 dataset and 65 for the MoMA one, we are not able to meet the retrieval rates expected from a 4D RPS generated through MDS. This has prompted us to deduce the following.

- 1) Although a set of more than 3000 features was used, the set is not complete enough to encode all the pertinent information for the two datasets. This deduction results from the variation of the precision rates with respect to increasing number of features as shown in Figure 6.5. We observe that even when the number of features is increased to 75 per dimension (using 300 features for the 4D MFS) the average precision rate is lower than when retrieval is performed directly from the RPS as provided in Figure 6.2. With 300 features used to encode the feature space, we would expect at least one feature to represent one texture sample given that the size of the dataset is only 120.
- 2) A statistical analysis of the precision values for the different Tex1 groups (refer to Table 6.3) shows that “circular” and patchy “textures”, were more difficult to encode even though observers grouped these textures quite easily. This may well be because the longer range interactions in these textures which are so obvious to the human eye, but are difficult to encode using computationally viable texture features.

To avoid overfitting of the data by the use of too large feature sets and to test the robustness of the retrieval model when presented with unknown textures, precision rates were computed for 10 Tex1 test textures (8 for MoMA) that were removed from the learning the retrieval model.

Significantly lower retrieval rates are obtained for the blind testing, however, they still achieved higher than 70% for precisions based on 30 retrieved textures from both datasets. Additionally, the average precision rates provided in Figure 6.15 show that a minimum of 4 and a maximum of 10 features are sufficient to obtain the best performance from a 4D MFS for the training textures considered. Increasing the number of features for the mapping does not provide for better retrieval results.

PART II: Full Perceptual Space Based Texture Retrieval

6.8 Introduction

In Part I of the chapter we argued that the Full Perceptual Space obtained from the psychophysical experiments could not be used to directly produce effective retrieval systems and that we needed first to reduce the dimensionality of the problem using MDS. In this part of the chapter we briefly investigate the problems and effectiveness of pursuing such a direct approach.

6.9 Overview of modified retrieval processes

Figure 6.18 shows the processes involved in performing a retrieval using the FPS directly instead of moving to lower dimensional spaces. The shaded part shows processes from the MFS based retrieval model that have been ignored. The different stages for the Optimisation model are presented in the rest of this chapter.

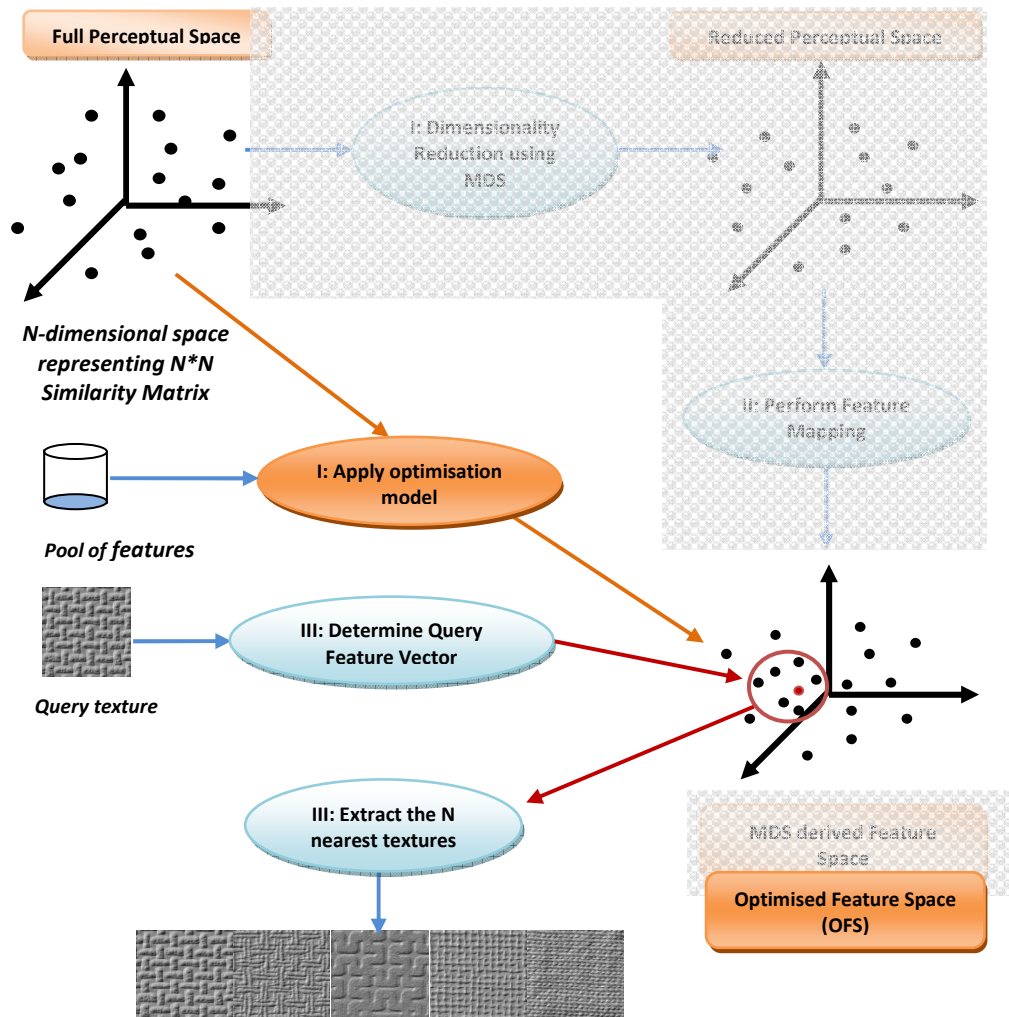


Figure 6.18 - The three main stages of the proposed Optimisation model. Shaded part show the corresponding stages from the model proposed in Section 6.2

6.10 Optimisation Model

The optimisation model proposed concerns the fitting of a feature space to the FPS. The fitting is viewed as a non-linear least squares problem to finding optimal features space that optimises the spatial arrangement of the Tex1 and MoMA textures within the relevant FPS. The resultant space is referred to as the Optimised Feature Space (OFS).

6.10.1 Statement of the problem

As in the case of the prediction model presented in Section 6.4.1, the problem can be viewed as an optimisation problem that seeks to obtain:

$$\hat{y}_i = f(x_1, x_2, \dots, x_M; \beta_1, \beta_2, \dots, \beta_F) \quad (6.9)$$

where (x_1, x_2, \dots, x_M) is a set of M texture features¹, $(\beta_1, \beta_2, \dots, \beta_F)$ is a set of F optimisation parameters and \hat{y}_i is the expected value based on the dependent dissimilarity value y_i .

As opposed to the prediction model, where the values to be estimated were the coordinates of the texture samples within the RPS, the optimisation model requires an optimum set of parameters β that estimate the dissimilarity values of the FPS as closely as possible. Using a least squares approach, an optimum set of parameters is obtained by minimising the following error function:

$$R(\beta) = \sum_{i=1}^L [y_i - \hat{y}_i]^2 \quad (6.10)$$

where L represents the number of off-diagonal elements from the dissimilarity matrix. Assuming that y_i is the dissimilarity value between any two textures T_j and T_k , then the function $f()$ in equation (6.9) computes the weighted distance that the two textures would make in the OFS. For example, for two texture features, the weighted distance is given by

$$\hat{y}_i = \beta_0 \left(\sqrt{\beta_1 (x_1^j - x_1^k)^2 + \beta_2 (x_2^j - x_2^k)^2} \right) \quad (6.11)$$

β_0 is a scaling factor for mapping the textures from the FPS to the OFS whereas β_1 and β_2 are weights that represent the involvement of each feature in estimating the dissimilarity value \hat{y}_i .

¹ Note that M feature values are required, where $M=2 \times N$ and N is the length of the feature vector for a single texture.

Generalising equation (6.11) for M features gives:

$$\hat{y}_i = \beta_0 \left[\sum_{m=1}^M \beta_m (x_m^j - x_m^k)^2 \right]^2 \quad (6.12)$$

To solve the problem posed by the error function (6.10) the Levenberg-Marquardt algorithm will be employed. The algorithm is presented in the following section.

6.10.2 Levenberg-Marquardt (LM) Algorithm

The LM algorithm is a standard technique for non-linear least-squares problems and has been widely exploited in a broad range of applications. It operates in an iterative fashion to locate the minimum of a multivariate function which is expressed in the form of the sum of squares of non-linear real-valued functions [Marquardt63].

To minimise the error function, $R(\boldsymbol{\beta})$, the LM algorithm proceeds by finding a linear approximation of a function $f(\mathbf{x}, \boldsymbol{\beta})$ in the neighbourhood of a parameter set $\boldsymbol{\beta}$. $f(\mathbf{x}, \boldsymbol{\beta})$, as used in this part of the thesis, is a function that maps an input dissimilarity value y_i to an estimated dissimilarity (or distance) value \hat{y}_i using equation (6.12). To converge to an optimum parameter set $\boldsymbol{\beta}^+$, $\boldsymbol{\beta}$ is iteratively updated using the LM update rule

$$\boldsymbol{\beta}' = \boldsymbol{\beta} - (\mathbf{H} + \lambda \text{diag}[\mathbf{H}])^{-1} \nabla R(\boldsymbol{\beta}) \quad (6.13)$$

In equation (6.13) above, \mathbf{H} represents the Hessian matrix of the function $f(\mathbf{x}, \boldsymbol{\beta})$ at search direction given by $\boldsymbol{\beta}$, $\nabla R(\boldsymbol{\beta})$ is the gradient of the error function $R(\boldsymbol{\beta})$ and λ is a dampening factor that is adjusted at each iteration to make sure a reduction in error occurs. The updated set of parameters is represented as $\boldsymbol{\beta}'$.

The algorithm initially proposed by Levenberg [Levenberg44] and modified by Marquardt [Marquardt63] works in the following way:

LM algorithm

- I. Compute the initial error from equation (6.12) using initial parameter set β^0 ,
- II. Compute the updated parameter set β' using the initial parameter set β^0 ,
- III. Determine the error from equation (6.12) using the updated parameter set β' ,
- IV. If updated error \geq previous error Then
 - Retain the previous set of parameters
 - Increase λ by a constant value (10 chosen)
 - Else
 - Retain updated set of parameters
 - Decrease λ by factor 10
 - End
- V. If error $<$ threshold or maximum number of iterations reached Then
 - STOP and return parameter set
 - Else
 - GOTO step II and perform update with new λ value and retained parameter set

The LM algorithm is first applied to a single feature problem, then two, three and more.

6.11 Texture retrieval using the optimised model

For retrieval from the OFS, the optimum set of parameters derived using the LM algorithm is applied to equation (6.12) together with selected features for the query texture, Q . Doing so positions the query texture within an M -dimensional space, where M is the number of features used. A Euclidean distance is applied to retrieve the closest textures to Q within the M -D space. Results are presented in the sections that follow.

6.11.1 Retrieval rates v/s number of features

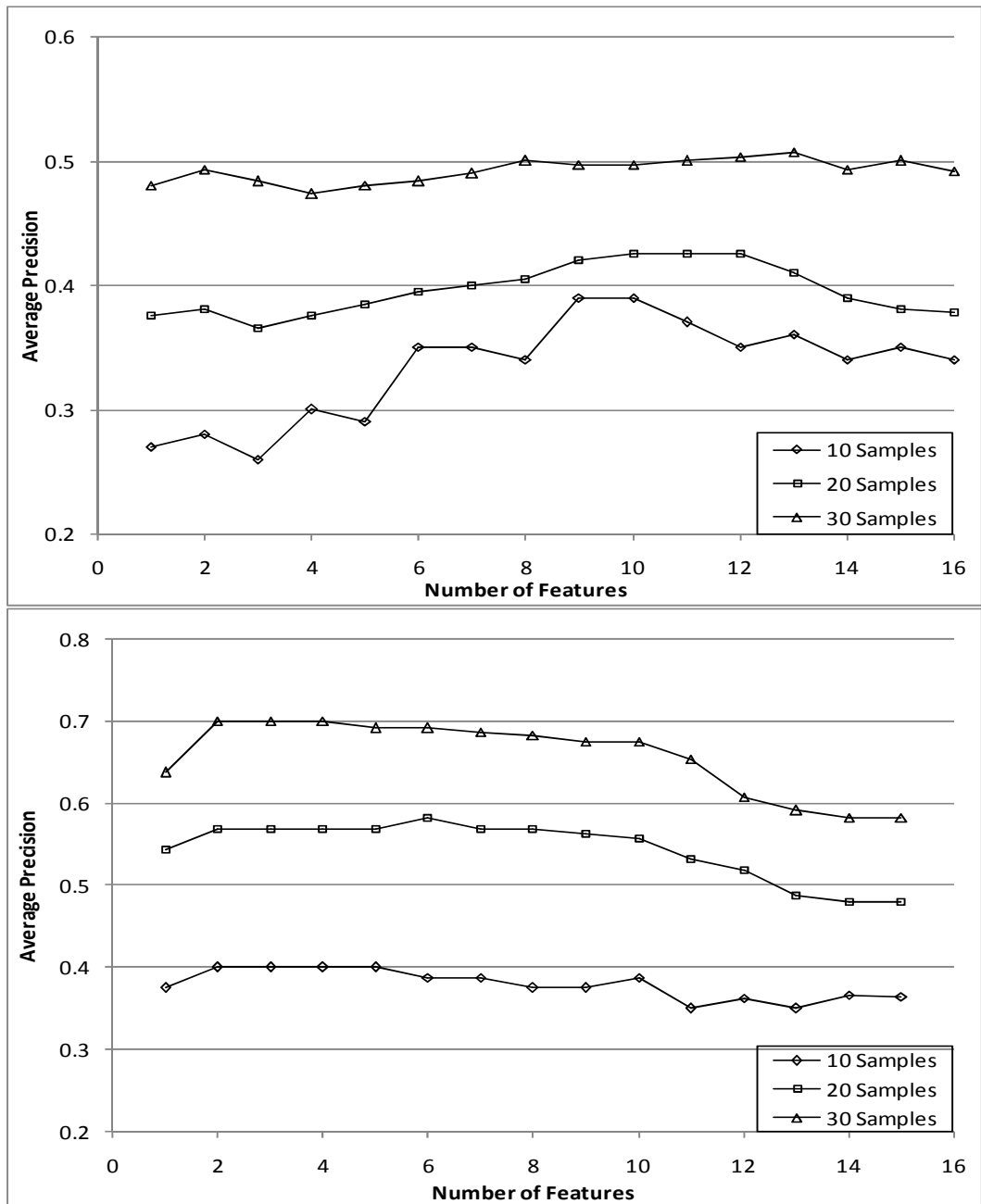


Figure 6.19 – Average precision rates for the Tex1(top) and MoMA(bottom) datasets with increasing number of features used to map the FPS to the OFS

Figure 6.19 shows the average precision rates when retrieval is performed using increasing number of features for the Tex1 and MoMA test textures. We can observe that the precision rates do not vary by large amounts for both datasets, except for the case of 10 retrievals for the Tex1 dataset where increasing the number of features appear to provide significant change in precision.

6.11.2 Blind test results using the OFS

Table 6.5 shows the average OFS precision for the ten “blind” test textures from Tex1. The results are compared with retrieval results from the MFS blind tests. As the retrieval results for the MFS were obtained using up-to four features per dimension, a maximum of 16 features was used to test for retrieval in the OFS.

Texture	MFS(B)	OFS(B)
T53	0.450	0.1962
T60	0.425	0.2692
T62	0.500	0.3308
T66	0.363	0.3615
T83	0.325	0.3769
T116	0.450	0.4115
T32	0.688	0.4346
T74	0.688	0.5462
T33	0.463	0.5808
T69	0.625	0.5808

Table 6.5 - Comparative results for average precision rates of Tex1 test textures for retrievals in MFS(B) and OFS(B)

We observe from the Table 6.5 that the precision rates for the OFS (with the exception of T83 and T33) are lower than those of the MFS. This shows that as well as being sparse, the information in the FPS is also noisy for Tex1 samples. Since the Tex1 dataset was created to cover as many texture categories as possible, it is not surprising that when the samples are projected in a 110D perceptual space, it is noisy.

Creating the OFS directly from the FPS may allow more noise information to be retained. This contrasts with the creation of the MFS, whereby using only a low-dimensional space enables much of the noise information to be “filtered” out.

Although the order of the test textures, based on retrieval from the MFS, is not retained compared with the corresponding retrieval performed in the OFS, the textures at the

higher end of the precision spectrum do tend to correspond. The median retrieval rate from the OFS for the 10 Tex1 test textures is 37.7%, corresponding to texture T83 which is shown in Figure 6.20.

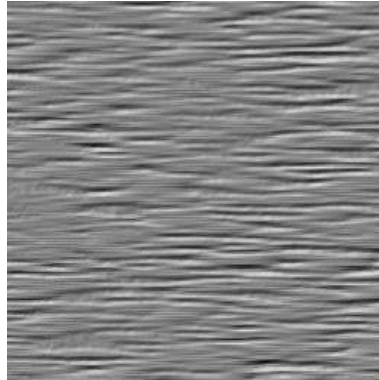


Figure 6.20 - Texture T83

Figure 6.21 shows the first 20 retrievals for T83 from both the FPS and the OFS. As the results from the FPS show, observers did not find any difficulty in perceiving the directional information and the textures that they perceived to be similar (for the Tex1 dataset) show a strong inclination towards horizontal textures.

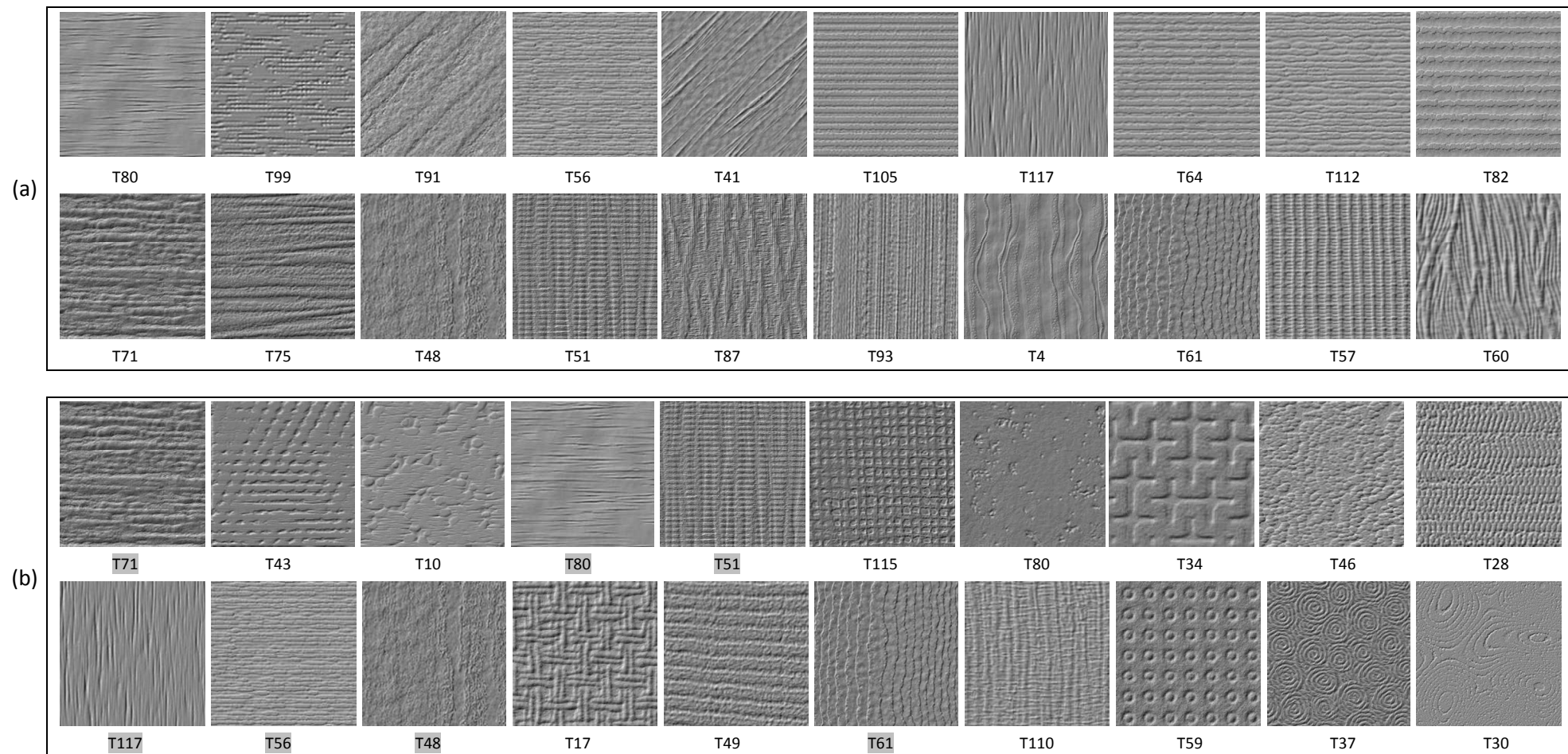


Figure 6.21- first 20 retrievals from (a) a 110D FPS and (b) the OFS feature space using texture T83 as query

6.11.3 Computation times

Part of the reason that the investigation of the FPS approach was not pursued further than the 14 texture features was because of the increasingly lengthy computation times that were required. Figure 6.22 illustrates the exponential nature of these computations for the Tex1 dataset. The processing time was recorded when executing the selection algorithm on MATLAB, hosted on a 32-bit XP machine with duo core processors (Intel Core 2 Duo 6600 operating at 2.4 GHz) and with 3 GB of RAM.

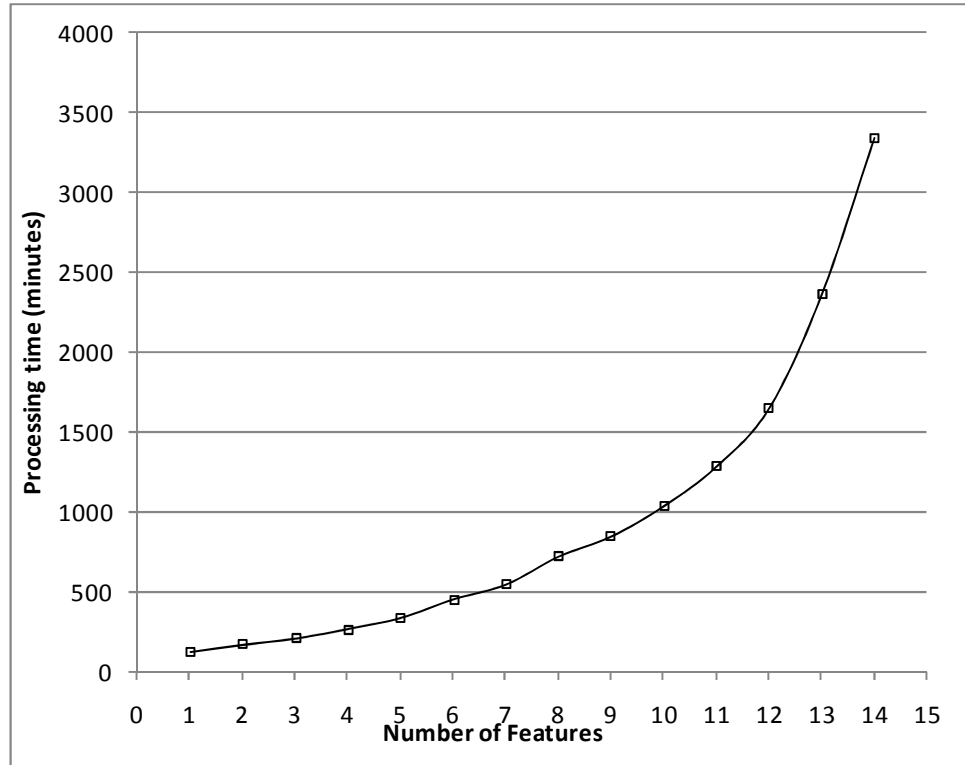


Figure 6.22- Processing time v/s number of features selected to create the OFS

6.12 Summary for Part II

Results obtained from the FPS approach show that:

- (a) it becomes prohibitively expensive to compare for higher number of features, and
- (b) that its performance is, on average, well below the reduced dimensionality approach described in Part I of this chapter.

6.13 Performance Evaluation and Discussion

Most of the texture retrieval systems encountered so far in literature have been evaluated with respect to a previous work. To do so researchers have mainly used a common dataset (Brodatz) and different feature sets for comparison such as the Tamura, Wold and Gabor features. To evaluate invariant features, researchers have applied affine transformations to existing textures that could be used as test textures. Thus, using invariant features for retrieval results in high retrieval performance, however, this performance is relative to the number of ‘identical’ textures in the test dataset rather than perceptually ‘similar’ textures. For example Payne *et al.* [Payne99] used 9 non-overlapping variants of the Brodatz textures, giving 1008 samples. They used one variant as query and based the retrieval performance on how well their system could retrieve the other 8 variants from the 1008 samples. Long *et al.* [Long00 & Long01] used cropped subimages of 60 Brodatz texture images (540 samples) and using one of the subimages as the query image, they measured how well the retrieval system would retrieve the remaining 8 textures. When “identical” textures are not considered the performance of retrieval systems drops considerably.

To improve the retrieval performance researchers have performed human studies to investigate, and learn, the way people categorise textures. Payne *et al.* [Payne99] performed retrieval based on 10 different statistical representations of textures and correlated these results with ranking of the same textures by humans. They found that only 20-25% of the retrievals matched the ranking result. Even by combining different computational methods, they could not achieve a success rate of higher than 50%. Payne *et al.* discovered that even if features invariant to scale and orientation were used for retrieval, the performance was still low, since those features are not necessarily perceptually consistent. This shows that retrieval of perceptually ‘similar’ textures rather than ‘identical’ ones is a much more difficult task to perform.

Thus the two datasets considered within this thesis, Tex1 and MoMA, do not contain any ‘identical’ textures. Moreover, we do not attempt to outperform the performance results published by Long *et al.* or Payne *et al.* with the performance of our model. In this thesis we have investigated how well the retrieval model proposed in section 6.2 performs with respect to the psychophysical results. That is how well it retrieves perceptually ‘similar’ textures.

The results provided in Part I of this chapter investigated how well a retrieval model applied to reduced perceptual spaces could be used to retrieve textures from the two datasets. Retrievals were performed from 4D MFS. We observed from Figure 6.15 that the average precision rate varies from 40% to just above 70% when the number of features per dimension considered is 10 for the Tex1 dataset and 8 for the MoMA dataset. If we consider that a retrieval engine can reasonably display 20 textures on a screen, then the model proposed does better than the expected maximum of 50% of perceptually consistent retrieval deduced by Payne *et al.* [Payne99].

We also compared our retrievals for both datasets against random chance. Figure 6.23 shows how precision rates vary with increasing numbers of features when 10 samples are retrieved from 4D MFS and the OFS. Considering 10 retrievals at a time, the random chance of retrieving 10 out of 110 samples for Tex1 and 10 out of 73 samples for MOMA are 9.1% and 13.5% respectively (these are shown as RCTex1 and RCMoMA on Figure 6.23).

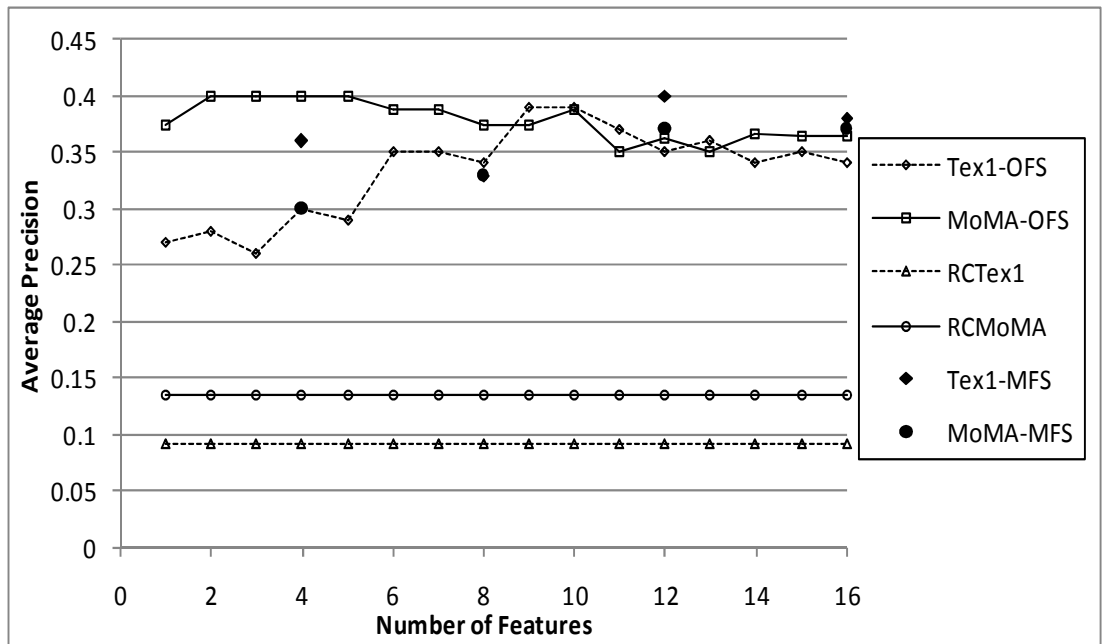


Figure 6.23- Effect of increasing texture features on retrieval success

The retrieval rates obtained from both the MFS and the OFS are significantly higher than the random chance values for the Tex1 and MoMA datasets. The results show only MFS results constructed using a maximum of 4 features per dimension and OFS results created using 16 features.

Figure 6.15 showed that performance was obtained with MFS created using 10 features per dimension. This implies a maximum of 40 features to create the OFS for a 4D

space. We did not attempt to test retrievals from OFS using higher than 16 features because

1. the results in Figure 6.19 show that the precision rates start to decrease when 14 or more features are used, and,
2. the computational cost increased significantly with larger feature numbers.

To have a better insight of how a more “conventional” retrieval engine would perform on the two datasets, we applied an LBP based retrieval method. The latter has basically been used both for retrieval and non-parametric classification of textures. LBP operators have been heavily utilised in the recent years to describe textures [Ojala96 & Ojala02] and a Chi-square based distance function has been successfully employed in both face recognition and texture retrieval [Ahonen04]. Using LBP histograms as feature vectors, the Chi-square distance between a query and a target texture is modelled as follows:

$$\chi^2(Q, T) = \sum_{b=1}^B [(Q_b - T_b)^2 / (Q_b + T_b)] \quad (6.14)$$

Q and T in equation (6.14) represent the LBP histograms for the query and target textures respectively. B is the total number of bins for the histograms. The histograms representing the query and target feature vectors are created by concatenating histograms from the following LBP operators: $LBP_{8,1}$, $LBP_{16,2}$, and $LBP_{24,3}$ (refer to Chapter 5 for LBP operator notations). Table 6.6 below shows the performance of the LBP based retrieval method as compared to the blind MFS and OFS strategies.

(a)		10 Samples	20 Samples	30 Samples	Average
	Random Chance	0.091	0.182	0.273	0.182
	LBP-ChiSquare	0.287	0.323	0.415	0.342
	MFS(B)	0.502	0.625	0.677	0.601
	OFS(B)	0.301	0.421	0.506	0.409

(b)		10 Samples	20 Samples	30 Samples	Average
	Random Chance	0.135	0.274	0.411	0.274
	LBP-ChiSquare	0.342	0.499	0.545	0.462
	MFS(B)	0.544	0.668	0.709	0.640
	OFS(B)	0.431	0.576	0.662	0.556

Table 6.6 – Precision values for different retrieval methods obtained when applied to (a) the *Tex1* dataset and (b) the *MoMA* dataset

Precision values were obtained for when retrievals are performed using the 10 test textures for the Tex1 dataset and the 8 MoMA test textures. 4 features per dimension were used for the MFS and a total of 16 features were used for the OFS. The results in Table 6.6 show that even if powerful texture features such as the LBP operators were used for retrieval, the Chi square based retrieval method performed quite poorly compared to the methods presented in this thesis. The MFS(B) method with average retrieval rates of 60.1% and 64.0% for the Tex1 and MoMA datasets is thus an efficient way to retrieve textures that match human perception.

6.14 Conclusion

In this chapter, we provided an effective retrieval model within a low-dimension perceptual space. When tested with textures already used for training, the precision rates for retrievals within a 4D MFS approximates the 90% mark for the datasets considered, however the number of features was high indicating likely overfitting. Even with a high number of features, precision rates for retrievals within a 4D MFS are below those for retrievals from a 4D RPS. This allowed us to conclude that the large pool of features used does not contain enough relevant features to represent the Tex1 and MoMA datasets and that there is likely to be correlation between the features selected for the retrieval systems. When blind testing was applied to the retrieval model proposed in Part I, the precision rates were lower but still significantly above chance. To ensure that the results obtained for the first retrieval model could not be obtained using a conceptually simpler and more direct approach, another retrieval model was proposed in Part II. The second model uses the FPS and maps the perceptual similarities directly to a feature space using an optimisation algorithm. Precision rates computed at increasing number of features for the OFS supported the case that a dimensionality reduction approach using MDS is a relatively effective and efficient approach.

Chapter 7

Summary, Conclusion and Future Works

7.1 Summary of Research

This thesis has attempted to meet a very specific objective: to develop an automatic retrieval system for surface textures by taking into account human perception of different categories of textures.

We started by presenting a survey of research work undertaken in the field of texture retrieval in Chapter 2, through which we identified that illumination conditions significantly affect the appearance of surface texture and that no work pertaining to the automatic retrieval of ‘surface’ textures has been undertaken. Additionally, the survey has allowed us to identify tools and techniques through which a perceptually relevant texture retrieval system could be built. Furthermore, it was noted that many of the retrieval systems described in the literature were tested using “identical” textures (in which multiple subimages had been obtained from a single original).

The focus of this thesis was also placed on how humans perceive different categories of surface textures. In Chapter 3 we presented the design and implementation of psychophysical experiments through which we recorded how humans group surface textures or texture images captured under the same illumination conditions. Two texture datasets were created and presented to observers for comparison: Tex1 and MoMA. The judgments from the users were aggregated in the form of similarity matrices for the Tex1 and MoMA datasets.

Chapter 4 analysed the similarity matrices and investigated the visual consistency of texture groups created from the psychophysical data through the use of dendrograms. The latter showed that the similarity matrices contained apparent structural information even if no obvious number of groups could be identified.

Using MDS as a dimensionality reduction technique, Chapter 4 also investigated whether the structural information contained in the Full Perceptual Space is preserved when the FPS is mapped to Reduced Perceptual Spaces. We found out that the textures from the Tex1 and the MoMA datasets could be well represented in any Reduced Perceptual Space with dimensionalities four to ten without significant loss in the

structural information. However, no obvious number of dimensions could be deduced when fitting the RPS to the FPS.

Chapter 5 investigated four popular texture feature sets to select one feature set that could be used to map a feature space to the perceptual space for automatic retrieval. A set of selection criteria was presented and consequently used to select the most suitable feature set for the mapping process. The Trace Transform features were found to satisfy all the criteria presented.

The different methodologies for automatic retrieval of surface textures were presented in Chapter 6. The first part described a simple approach that used linear regression to map the RPS to a corresponding feature space. Since the dimensionality of the RPS is application oriented, precision was used to determine the number of dimensions to be used for the RPS for retrieval purposes. The precision values for different retrieval modes (10, 20, 30 retrievals) showed that significant fit of the FPS is obtained when 4D or higher dimensionality RPS are considered. Thus, the TT transform features were used to map the 4D RPS for both datasets. Retrieval performances from the resulting feature space (the MFS) were below the expected performance of retrievals from the RPS. This prompted us to deduce that the TT feature set was not complete enough to encode all the textures available (mainly for the Tex1 dataset). However, with precision rates of higher than 70% for 30 samples retrieved, the retrieval model proposed proved to be effective one.

A more direct approach that uses the full perceptual space was proposed in Part II of Chapter 6. It was found that average precision for this approach was lower than the dimensionality reduction approach. Additionally the direct approach required high computation times with increasing number of features making it impractical and unattractive.

7.2 Conclusion

In this thesis we have developed retrieval models that integrate perceptual data from psychophysics to provide for perceptually relevant retrievals of textures. The texture images used in the psychophysical experiments were obtained by rendering surface textures using known illumination parameters (slant and tilt). In addition a set of texture images, captured under uniform illumination conditions have also been employed (MoMA dataset). This is the first time that surface textures have been used to capture human perception of texture and to develop automatic retrieval models where

the features used to ‘statistically’ describe the textures are not influenced by any change in illumination directions.

Two retrieval models were presented. In the first model, the Full Perceptual Space was reduced to a more manageable Reduced Perceptual Space using MDS. A large pool of features was used to create a corresponding feature space (the MFS). The performance achieved was better than that which could be obtained by pure chance when precision rates were computed for 10, 20 and 30 retrievals.

The second retrieval model was used to investigate whether a more direct approach could provide for better retrieval results. Thus an Optimised Feature Space, derived directly from the Full Perceptual Space, was exploited for retrieval. The performance of the second model was no better than that obtained from the MDS based Feature Space. Furthermore the high computational time required to select texture features for the optimised space makes it an impractical option.

	MFS	OFS	Random Chance
Tex1	60.1%	40.9%	18.2%
MoMA	64.0%	55.6%	27.4%

Table 7.1– Average performance (blind testing) for Tex1 and MoMA datasets for retrievals in the MFS and OFS. The performance by chance for 110 and 73 target textures from Tex1 and MoMA is provided for comparison.

Table 7.1 above shows the average performance when 10 and 8 test textures (blind testing) were searched from the remaining Tex1 (110) and MoMA (73) target textures respectively. The performance values represent the average precision for 10, 20 and 30 samples with retrievals performed in both the MFS and OFS. Comparative performance for pure chance retrieval is also provided.

The performance values for retrievals in the MFS are better than those obtained from the OFS for the Tex1 and MoMA datasets. This shows that the MFS based retrieval model proved to be a relatively effective and efficient retrieval methodology.

Moreover, considering the datasets contained no “identical¹” textures, a performance of above 60% for the MFS based model is very promising.

¹ ‘identical’ used in the same context as defined in Chapter 2

Analysing the results from the MFS based model, we observed that a number of textures that were easily grouped by observers, did not produce the same expected results when used as query textures. Even with the availability of a large feature set, few relevant features were able to encode the longer range structural information within the textures. However, this type of information is difficult to capture using computational features. Overall, the MFS based model is a simple, inexpensive methodology that develops efficient and effective retrieval models.

7.3 Future works

The results for perceptual texture retrieval provided in this thesis were satisfactory given (1) the small number of textures used to train the proposed retrieval model and (2) the simple mapping technique (linear regression) used to map the feature space to the perceptual space. Some obvious and immediate improvements, deduced from the conclusions provided, to obtain better performance for the perceptual retrieval are listed below:

1) Generate a large dataset (> 300) of homogeneous surface textures

120 surface textures (Tex1 dataset) was used to capture human perception of textures and to investigate perceptual dimension, however, they provide relatively sparse sampling of what is at least a four-dimensional space.

An immediate follow-up of this research would be to apply the retrieval framework proposed to a larger set of homogeneous textures. The Tex1 consisted of varied texture categories. The same categories could be used as a basis to generate an expanded surface texture dataset.

2) More detailed investigation of grouping results

This thesis focused on the use of similarity matrices to develop perceptually relevant retrieval systems. The matrices represented the frequency of occurrence of different texture pairs. Although this information was sufficient to represent the textures within a perceptual space, other information derived from the grouping results could have been used to identify perceptually relevant texture features.

Firstly, dominant groups may be identified by recording the order in which the groups are created by observers and selecting those groups that all or most observers have found very easy to create. This can lead to the identification of perceptual texture attributes that humans can distinguish more easily when comparing textures. Feedback on what criteria have been employed by observers in grouping textures can also help in understanding human perception of texture.

3) Investigate other human judgment capturing methods

In this thesis, perceptual grouping has been efficiently utilised to derive similarity matrices for the purpose of developing perceptually relevant texture retrieval systems. However the datasets used (Tex1 and MoMA) were of relatively small sizes. For larger datasets (>300) perceptual grouping may not be a very practical option and other methods would need to be investigated. For example pairwise comparison may be useful if proper false negatives rejection mechanisms are used or if reaction times are controlled.

One limitation of perceptual grouping that we encountered during the course of the thesis was that ordering information was ignored during the comparison (of retrievals from the MFS and the FPS) stage. Perceptual ordering may be considered in this case, provided it is efficiently implemented so as to reduce the comparison times from observers.

4) Investigate larger sets of independent, phase sensitive features

The performance of a texture retrieval system relies heavily on the feature sets used to encode the textures being searched. It has already been demonstrated in the literature that the phase information contains most of the structural information within an image [Oppenheim91]. However, obtaining feature sets that are sensitive to phase and insensitive to position is a difficult task. This thesis has investigated several popular feature description approaches that encode phase information and the Trace Transform (TT) features were selected to create a feature space for texture retrieval. Although a subset of the TT features was sensitive to phase, they could not encode all the variability in textures available (Tex1 and MoMA). A feature set containing a large proportion of phase sensitive features is more likely to provide for better texture representation and should be investigated.

Moreover when very large feature sets are considered, the number of correlated elements is quite high, as was the case with the TT features used in this thesis. A feature set with a large number of independent elements would contribute in having more relevant features to represent the textures.

5) More robust mapping of the feature space to the Reduced Perceptual Space

This thesis presented a retrieval model that used a linear regression model to map a feature space to a reduced perceptual space obtained through Multidimensional Scaling. The model did perform efficiently in encoding most of the textures and provided satisfactory retrieval results. We assumed that the feature data were linearly distributed across the samples; however when large feature sets are involved, this condition is difficult to achieve. Non-linear fitting models could be tested to investigate for better and more robust mapping and feature selection. Some common techniques that have been heavily utilised in the recent years are mainly the Support Vector Machines and multilayer neural networks [Long01] and could be exploited to derive more robust retrieval models.

Appendix A: Texture Datasets – Tex1 and MoMA

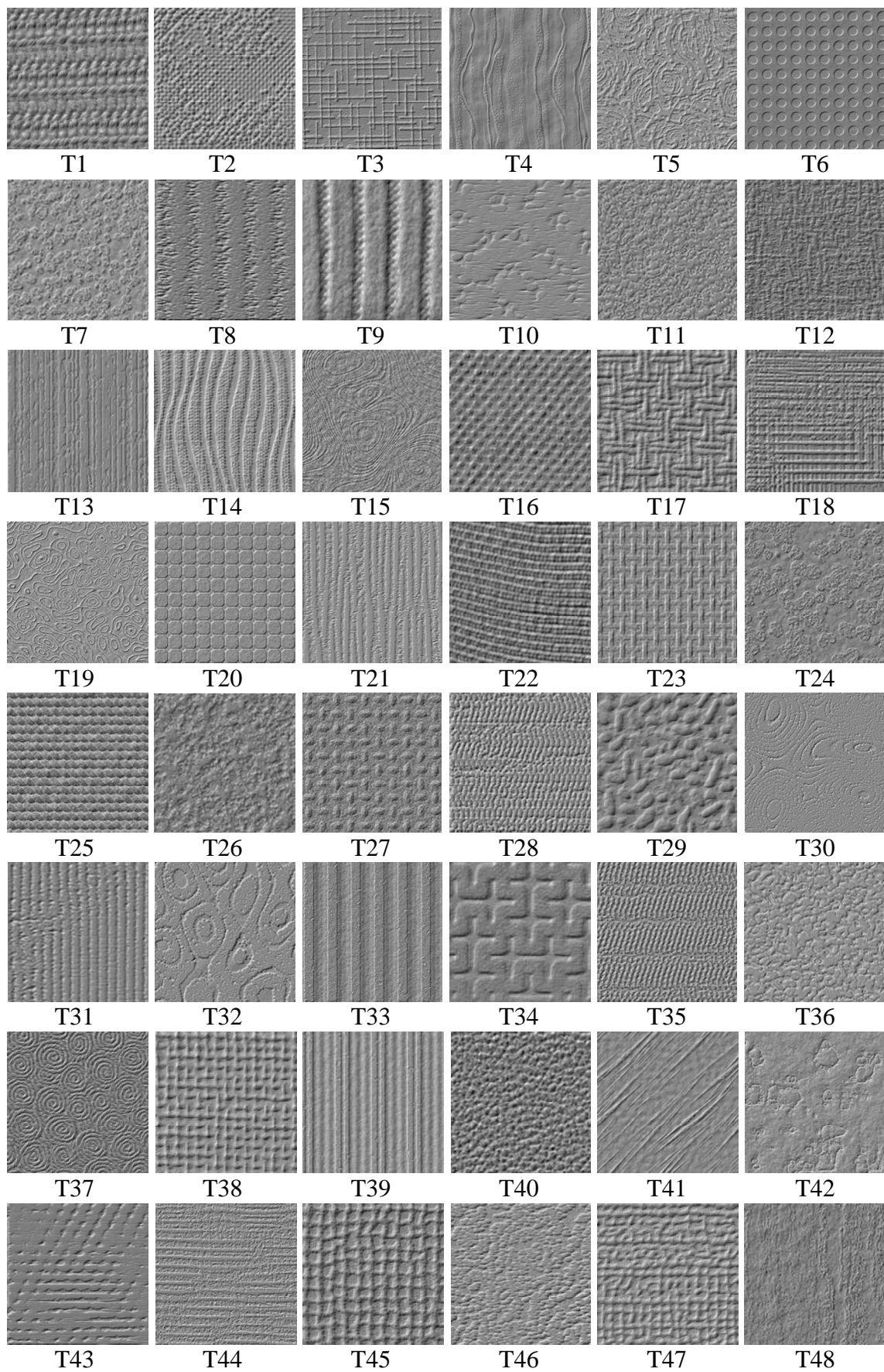


Figure A.1- Texture images for *Tex1* dataset, Part I (T1 to T48)

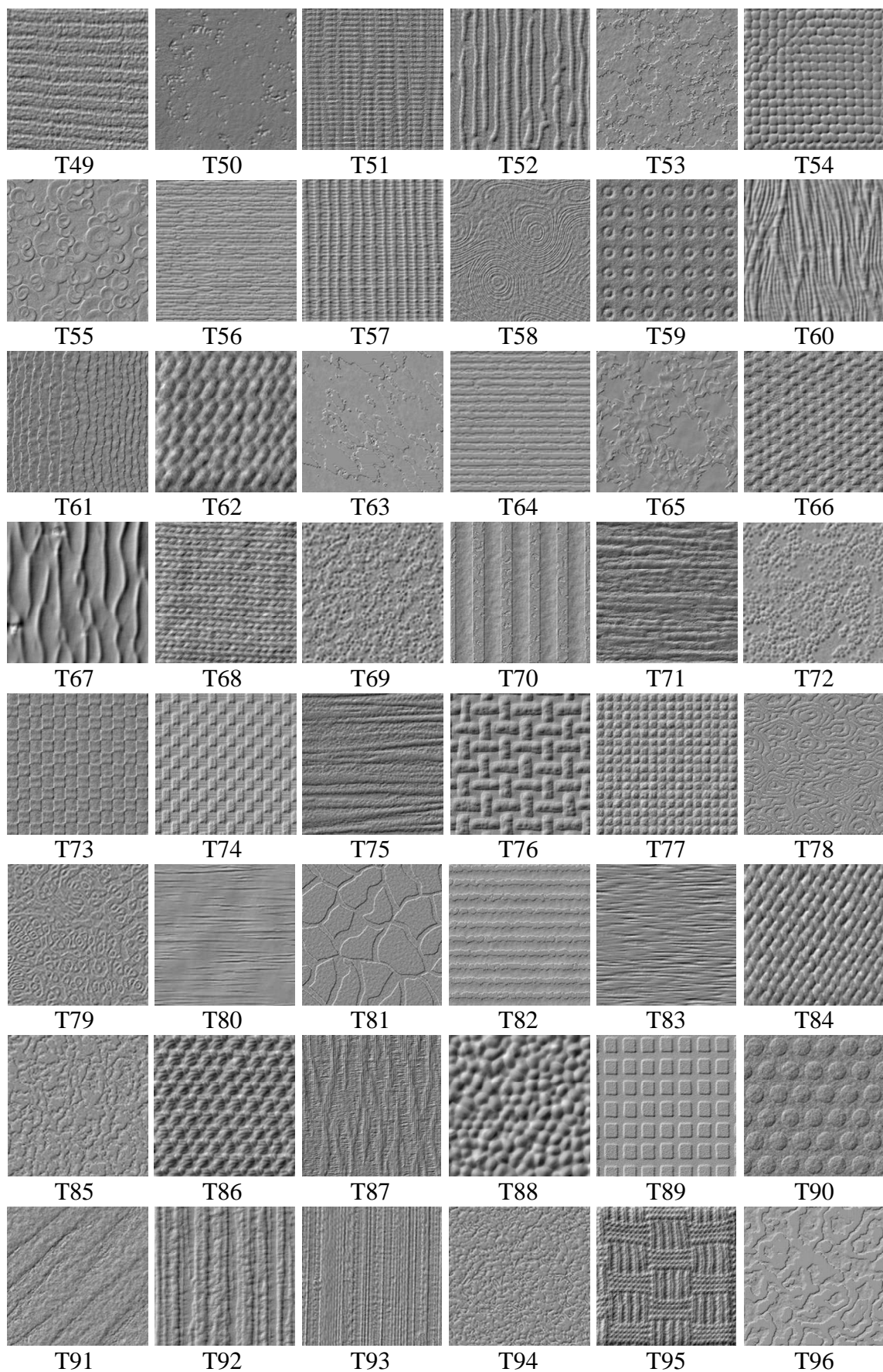


Figure A.2- Texture images for *Tex1* dataset, Part II (T49 to T96)

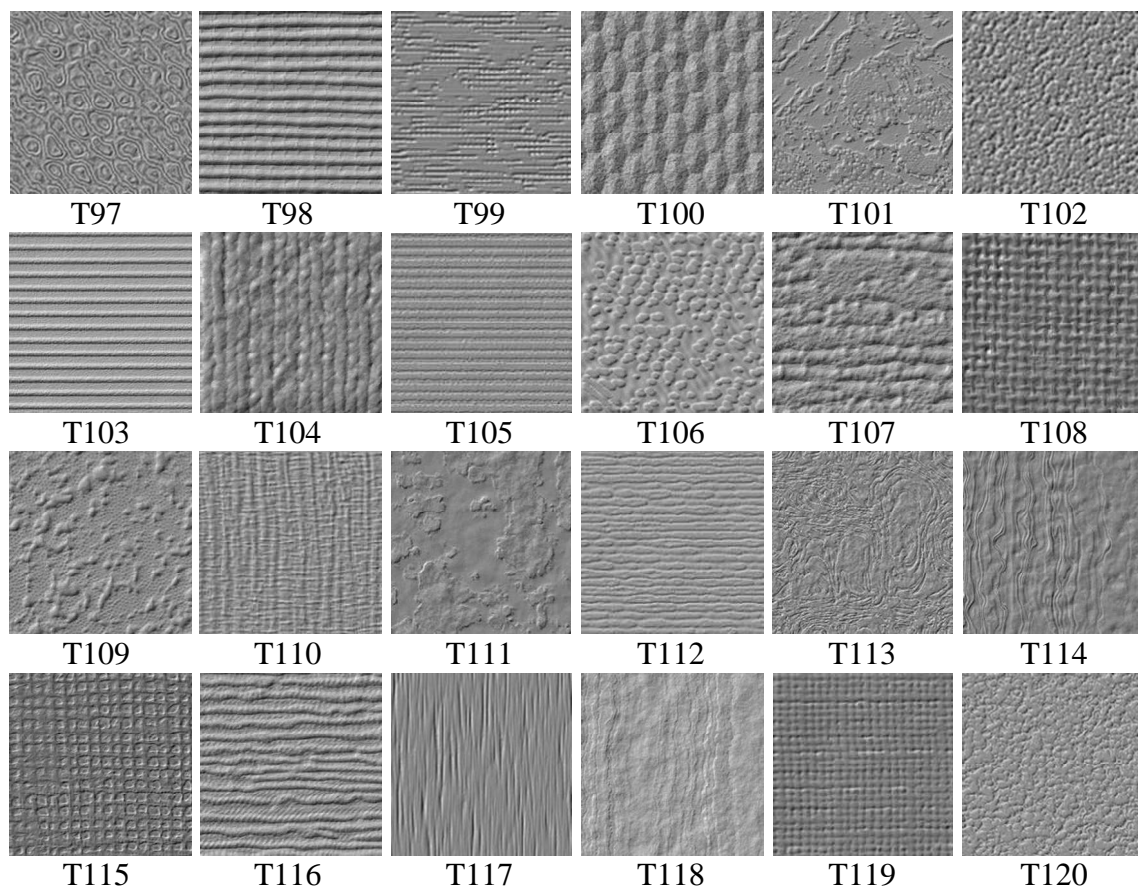


Figure A.3- Texture images for Tex1 dataset, Part III (T97 to T120)

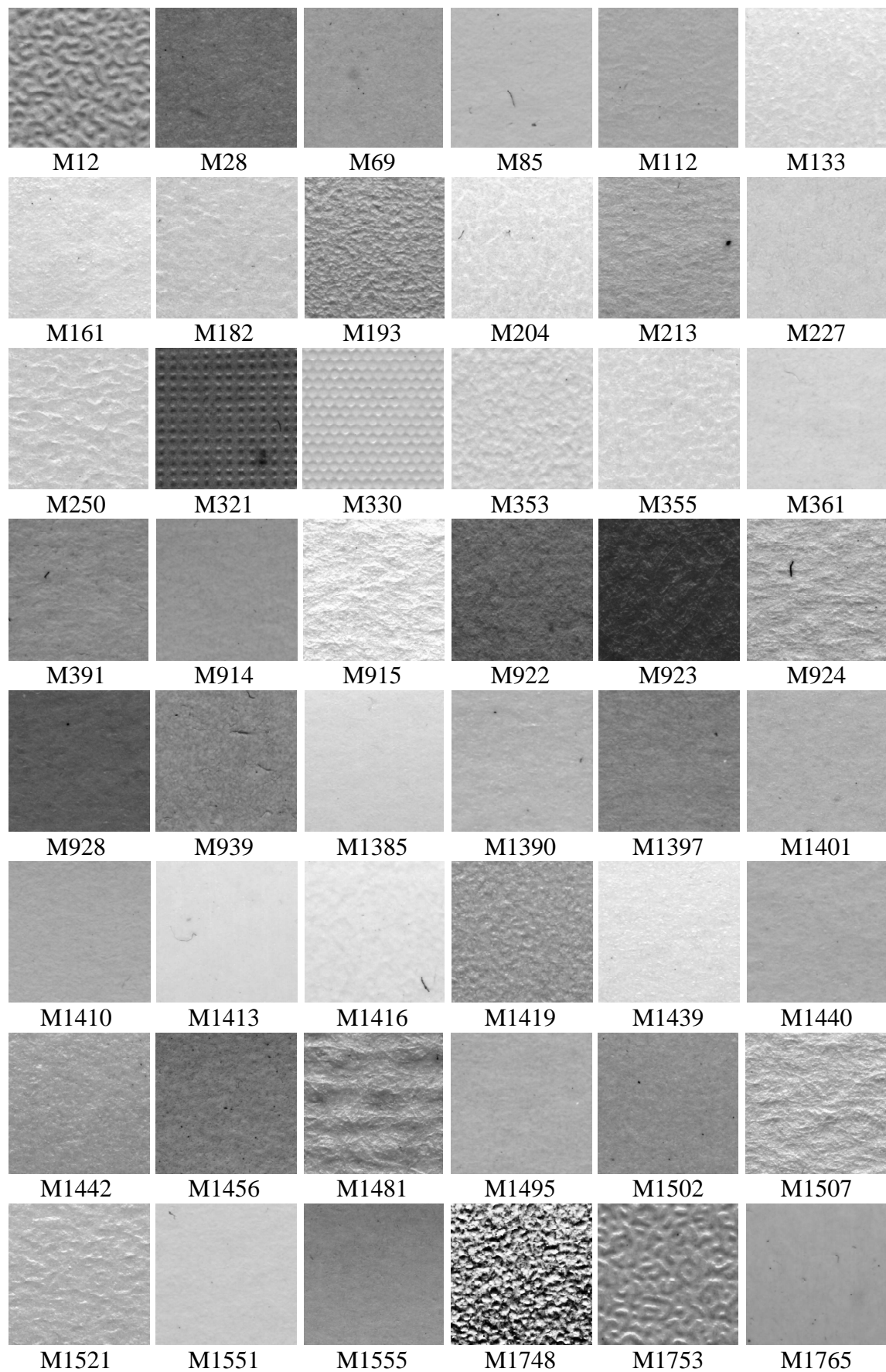


Figure A.4- Texture images for MoMA dataset, Part I (M12 to M1765)

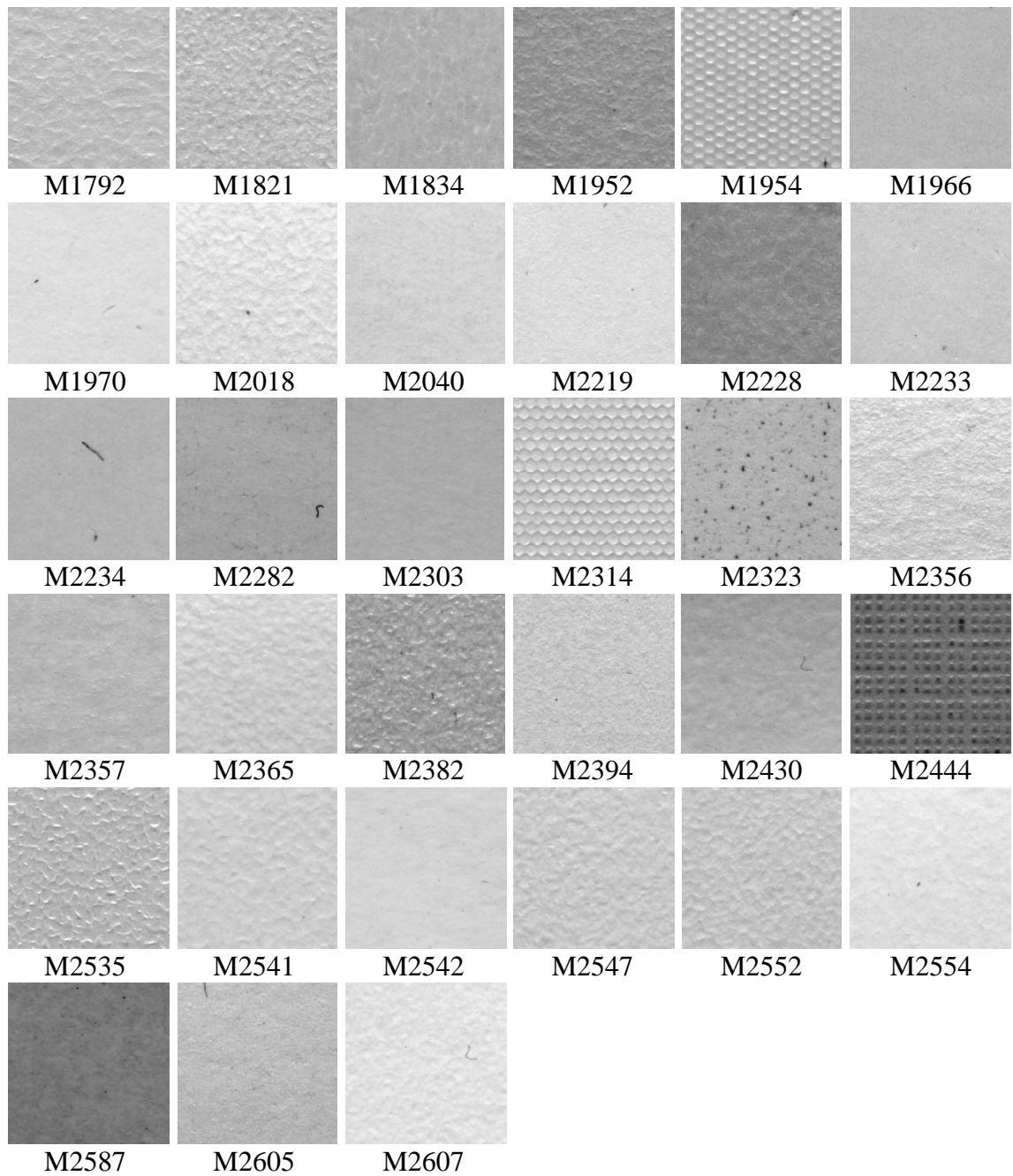


Figure A.5- Texture images for MoMA dataset, Part II (M1792 to M2607)

Appendix B: Grouping Experiment Instructions

Title: Perceptual Grouping of Textures**(I) Aim:**

The aim of this experiment is to come up with a similarity matrix that would represent the ways in which human subjects would group together specific textures from an unknown set of textures. The similarity matrix would thus represent the frequency at which a particular texture is coupled with another texture.

(II) Experiment Setup:

Hundred and twenty surface textures were chosen for this experiment. The set of surface textures is made up of both natural and synthetic textures. Lambertian illumination is used to render the surfaces. A slant of 70 degrees and tilt of 45 degrees have been used for the rendering (Top left corner). The texture images will be presented to the human subjects in form of photographs.

(III) Instructions for observers:**Precursor**

Orientation – please **DO NOT** rotate individual photos, it's important that you view them with the **PRINTED NUMBER AT THE BOTTOM**.

Similarity – when sorting the textures try not to think too consciously about the individual characteristics of the textures – rather, imagine them as real surfaces and group them according to simple gut instinct.

Procedure

- (i) Photographs would be randomly placed on a table by experimenter so that you can see all of them (remember to keep them all orientated the same way).
- (ii) Now create as many groups of textures as you feel like by moving the photographs around on the table as much as you like – the only criterion being that each group should contain “similar” textures.
- (iii) Do not feel afraid to create groups containing single textures if you feel that the texture is not sufficiently “similar” to any of the others. Above all – do not create an “oddball” group which contains textures that simply do not fit into any of the other groups.
- (iv) Once grouping completed, leave your observations on the table so that they could be registered by experimenter.

Thank you for your participation

Figure B.1– Instruction sheet presented to subjects participating in the psychophysical experiment

Appendix C: Similarity Matrix

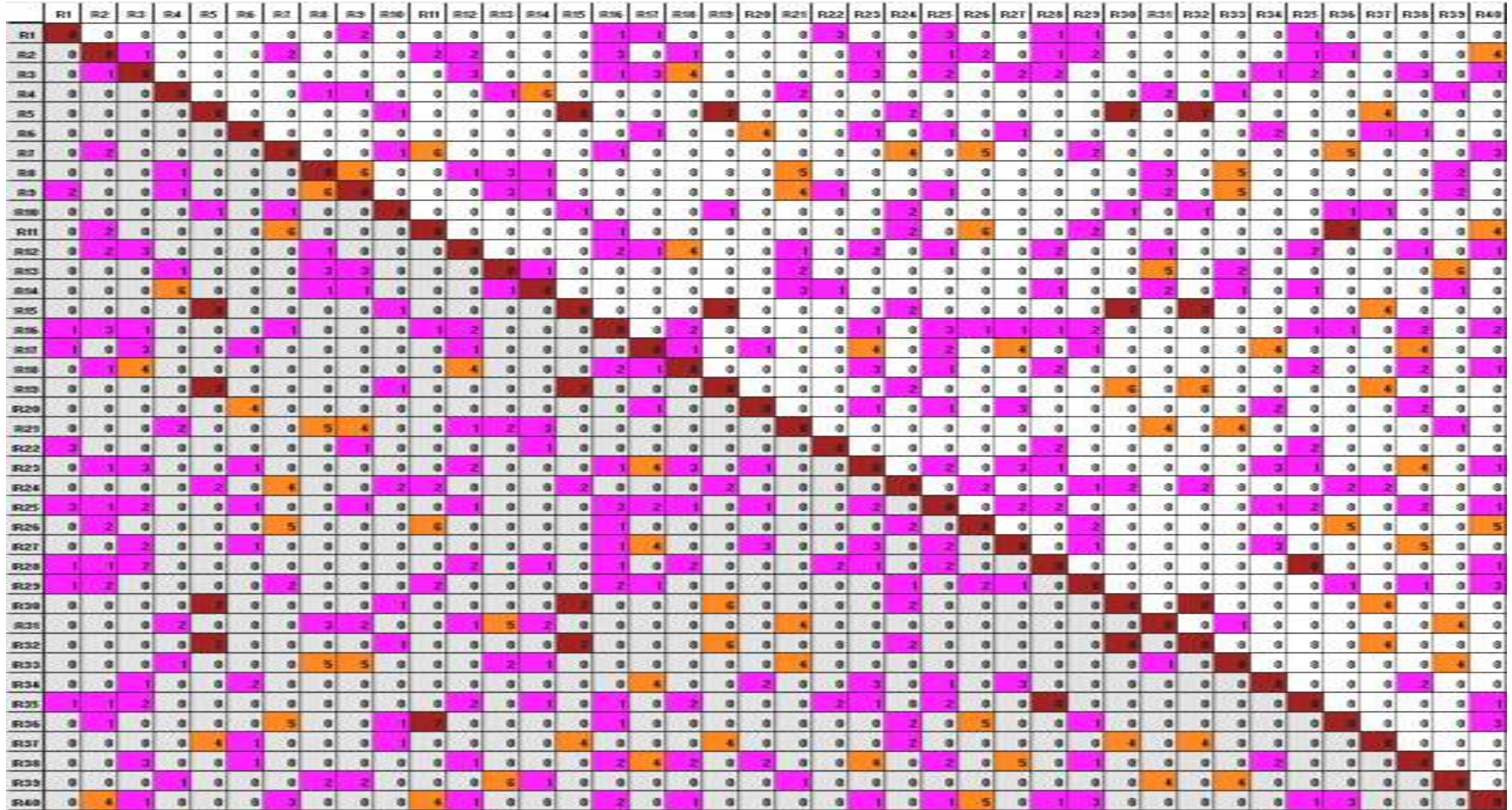


Figure C.1 Partial Similarity matrix showing pairwise occurrence of Tex1 images R1 to R40 (generated from surfaces T1 to T40). Matrix constructed from data coming from 8 subjects

Appendix D: Grouping Results

Appendix D.1: Tex1 Groups

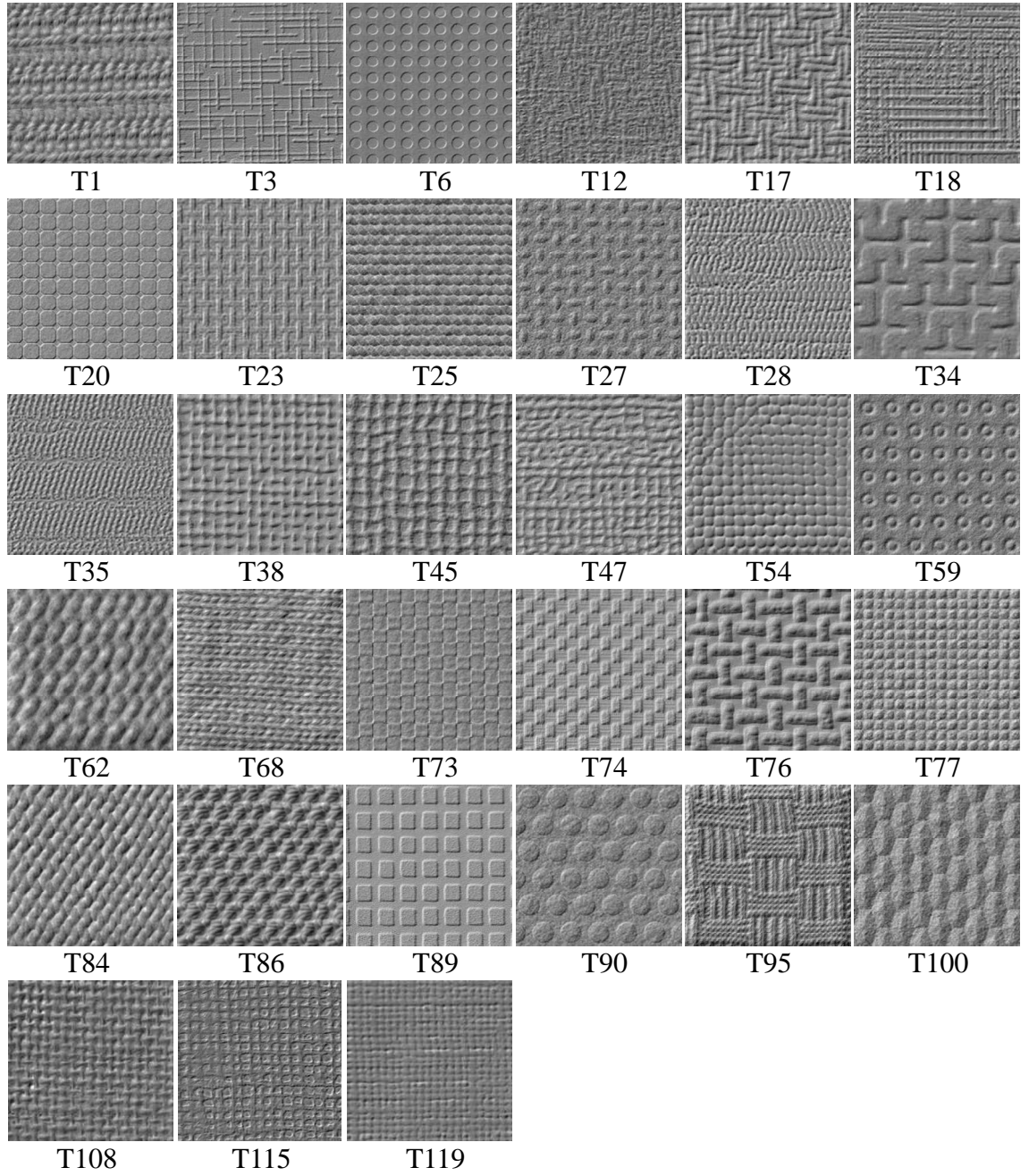


Figure D.1- Tex1: Group1 (regular textures), group size =33

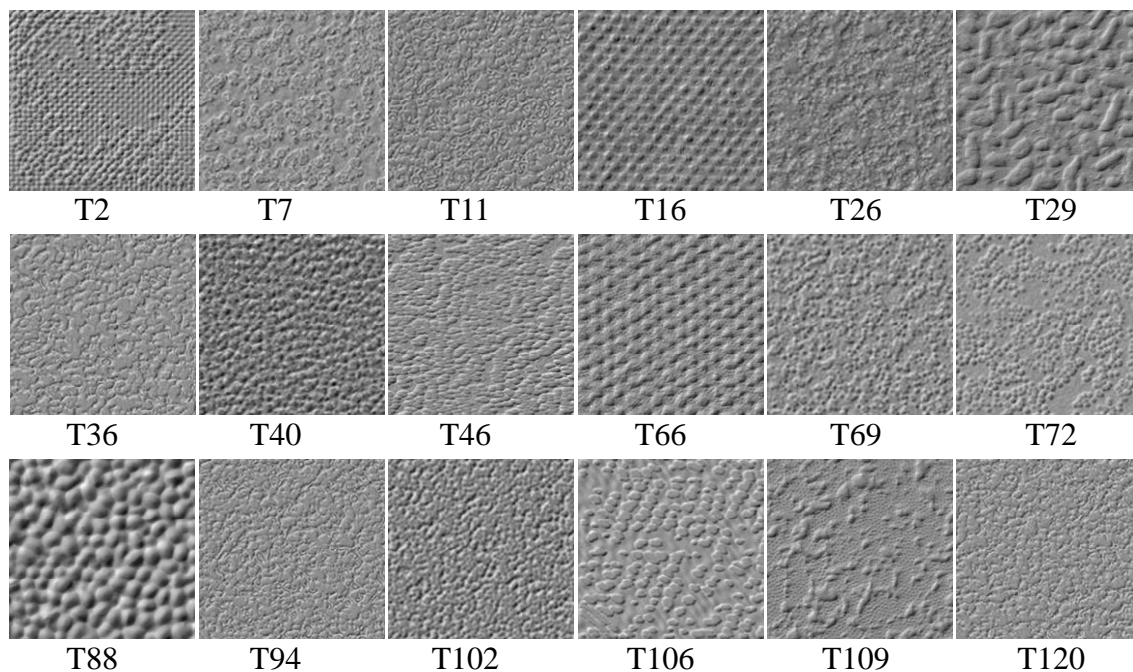


Figure D.2- Tex1: Group2 (irregular textures), group size=18

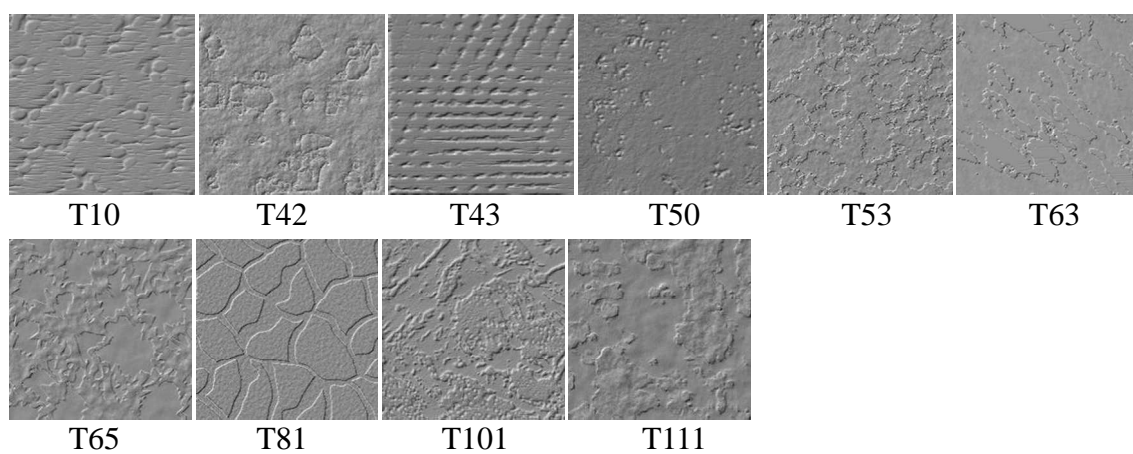


Figure D.3- Tex1: Group3 (patchy textures), group size=10

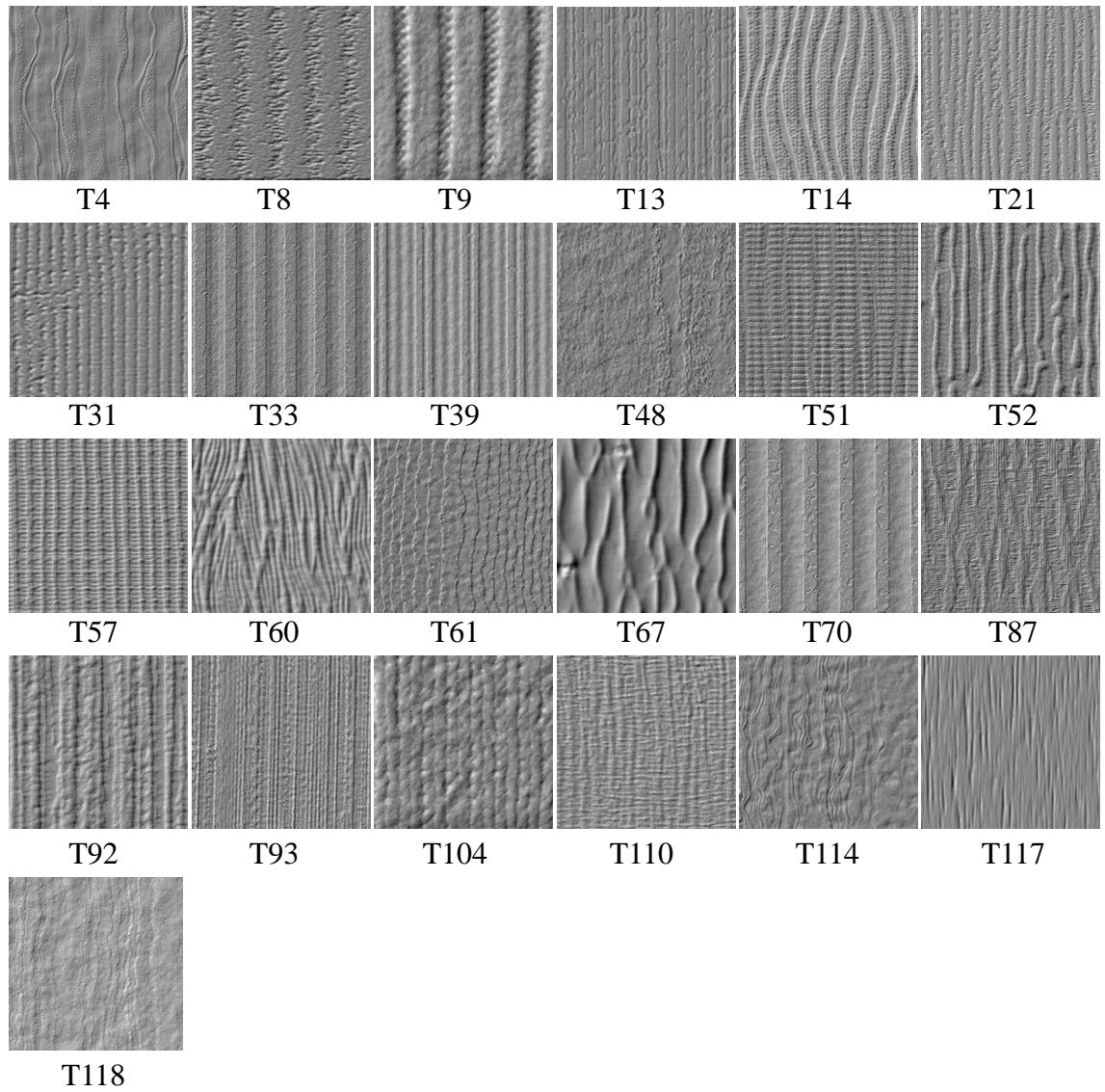


Figure D.4- Tex1: Group4 (vertical textures), group size=25

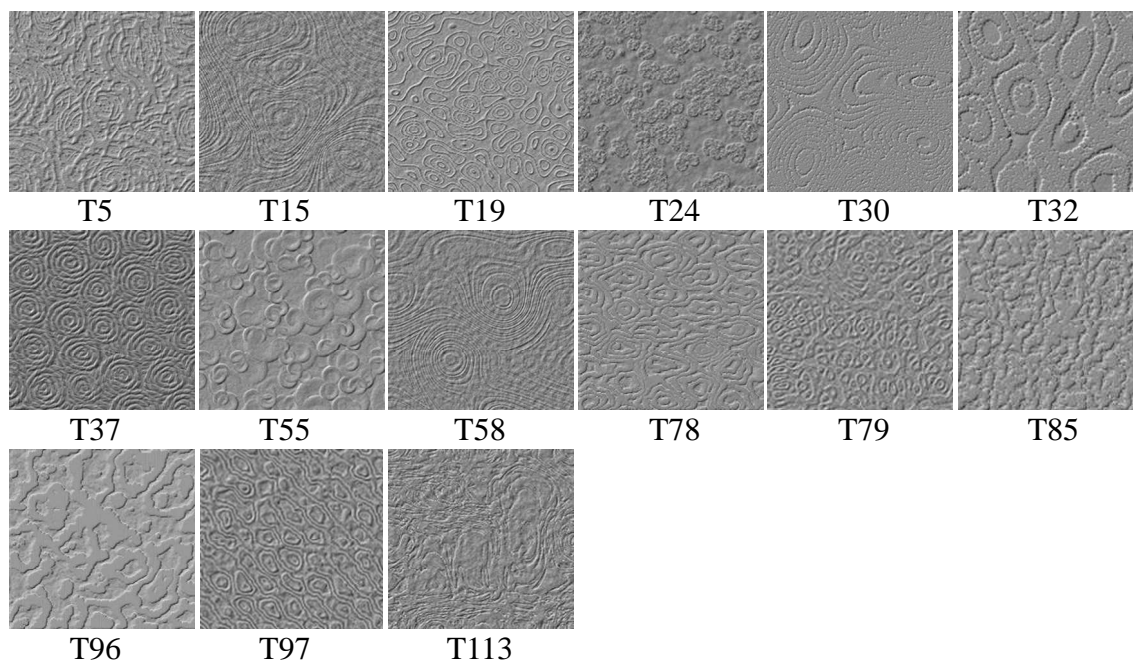


Figure D.5- Tex1: Group5 (Circular textures), group size =15

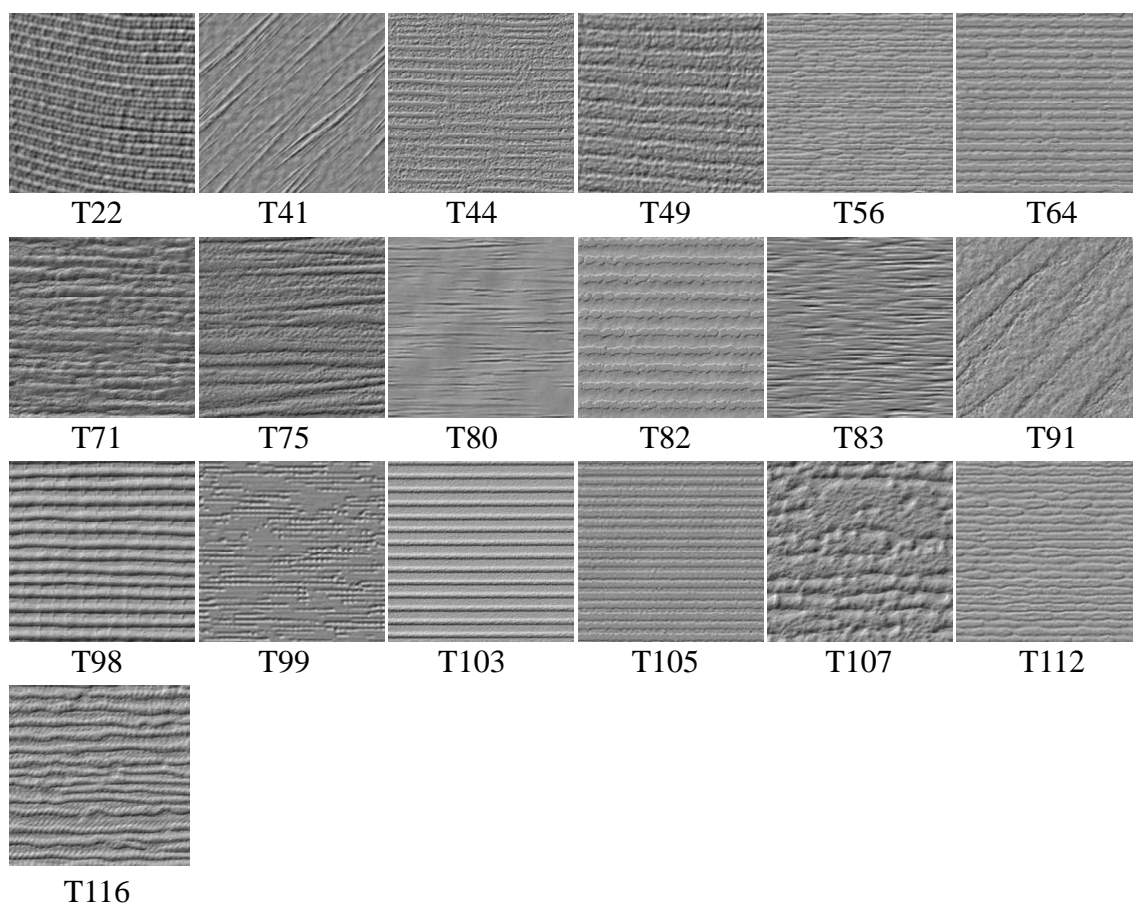


Figure D.6- Tex1: Group6 (Horizontal textures), group size= 19

Appendix D.2: MoMA Groups

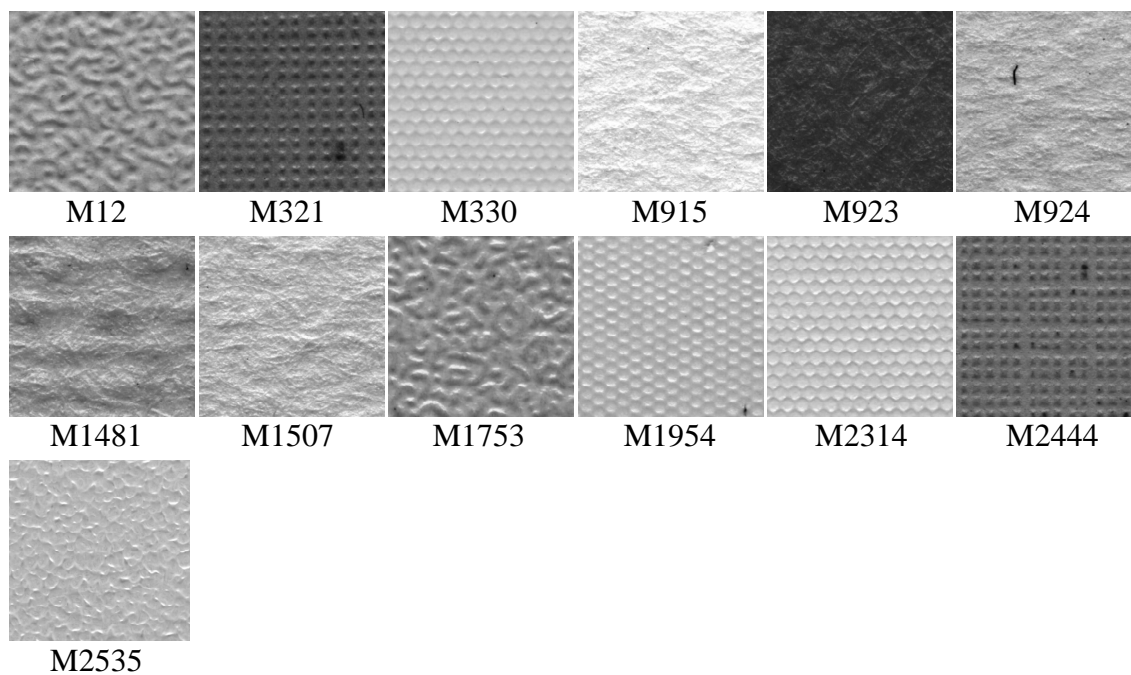


Figure D.7- MoMA: Group1, group size =13

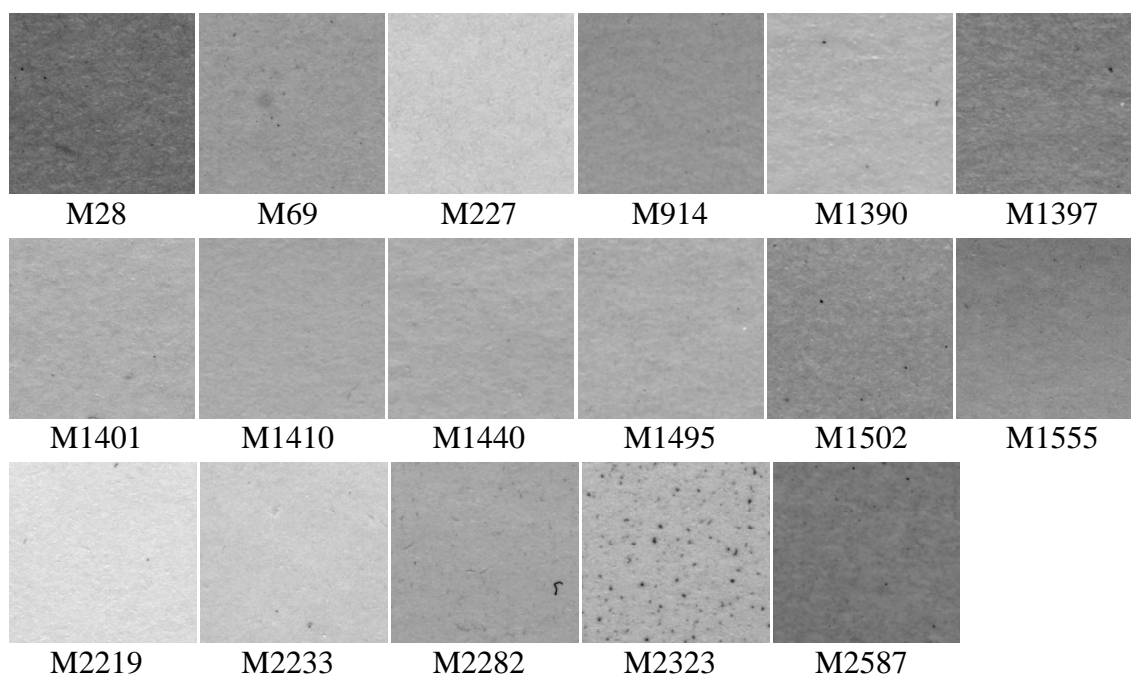


Figure D.8- MoMA: Group2, group size =17

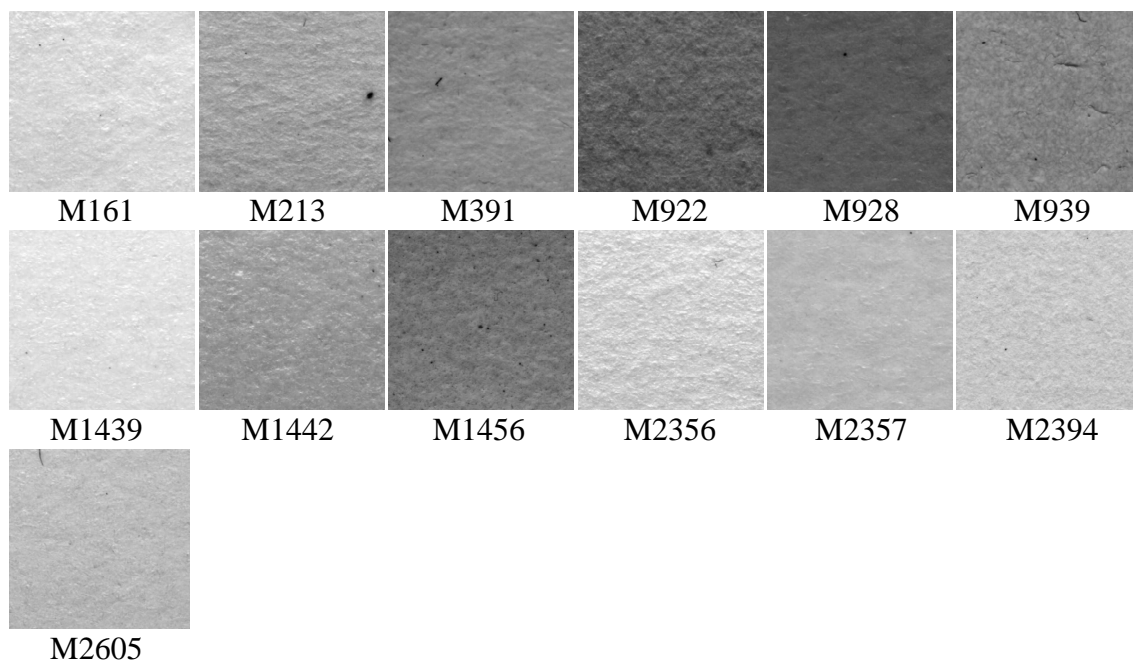


Figure D.9- MoMA: Group3, group size =13

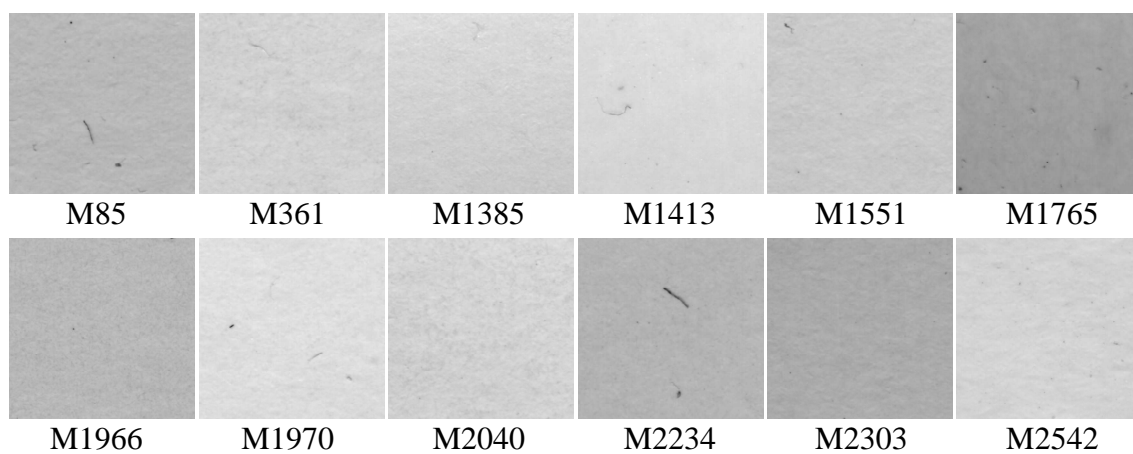


Figure D.10- MoMA: Group4, group size =12

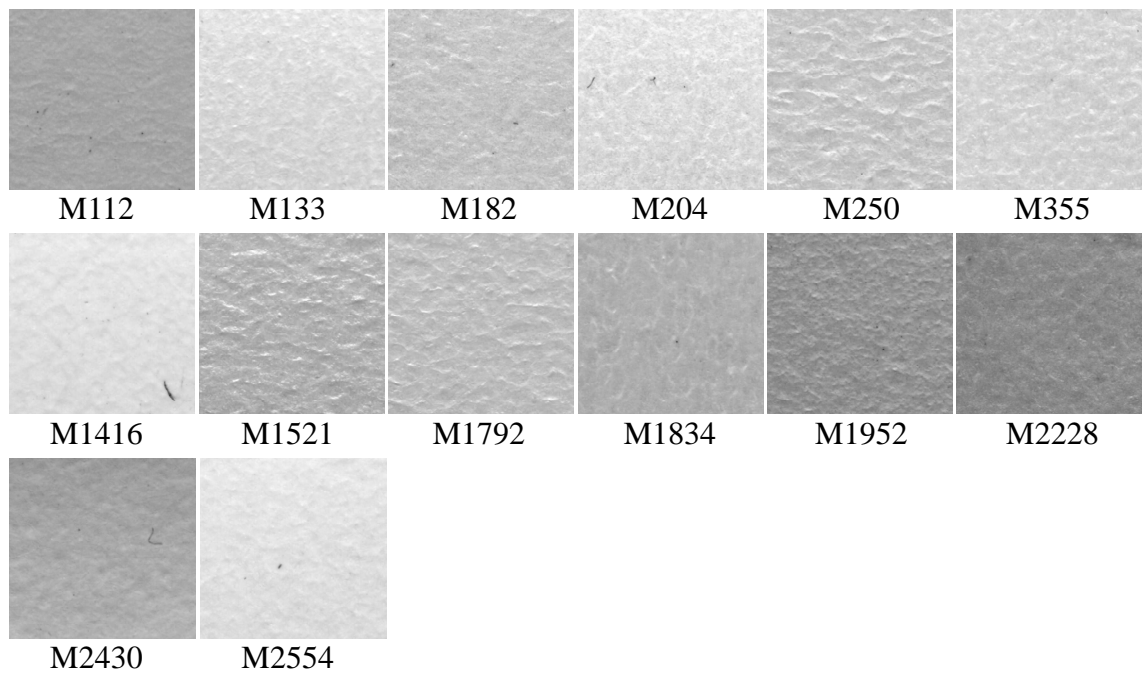


Figure D.11- MoMA: Group5, group size =14

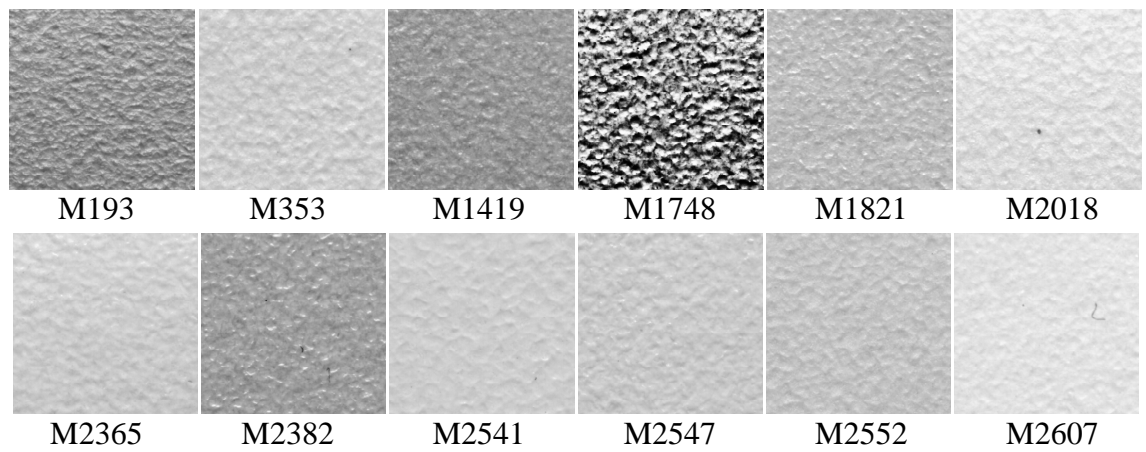


Figure D.12- MoMA: Group6, group size =12

Appendix E: Trace Transform Functionals

Index	Functionals
T1	$\sum_{i=1}^N x_i$
T2	$\sum_{i=1}^N ix_i$
T3	$\sqrt{\sum_{i=1}^N x_i^2}$
T4	$\max_{i=1}^N x_i$
T5	$\sum_{i=1}^{N-1} x_{i+1} - x_i $
T6	$\sum_{i=1}^{N-1} x_{i+1} - x_i ^2$
T7	$\sum_{i=3}^{N-2} x_{i-2} + x_{i-1} - x_{i+1} - x_{i+2} $
T8	$\sum_{i=4}^{N-3} x_{i-3} + x_{i-2} + x_{i-1} - x_{i+1} - x_{i+2} - x_{i+3} $
T9	$\sum_{i=6}^{N-5} x_{i-5} + x_{i-4} + \dots + x_{i-1} - x_{i+1} - \dots - x_{i+4} - x_{i+5} $
T10	$\sum_{i=8}^{N-7} x_{i-7} + x_{i-6} + \dots + x_{i-1} - x_{i+1} - \dots - x_{i+6} - x_{i+7} $
T11	$\sum_{i=5}^{N-4} \sum_{k=0}^4 x_{i-k} - x_{i+k} $
T12	$\sum_{i=6}^{N-5} \sum_{k=0}^5 x_{i-k} - x_{i+k} $
T13	$\sum_{i=7}^{N-6} \sum_{k=0}^6 x_{i-k} - x_{i+k} $
T14	$\sum_{i=8}^{N-7} \sum_{k=0}^7 x_{i-k} - x_{i+k} $
T15	$\sum_{i=11}^{N-10} \sum_{k=0}^{10} x_{i-k} - x_{i+k} $
T16	$\sum_{i=16}^{N-15} \sum_{k=0}^{15} x_{i-k} - x_{i+k} $
T17	$\sum_{i=21}^{N-20} \sum_{k=0}^{20} x_{i-k} - x_{i+k} $
T18	$\sum_{i=26}^{N-25} \sum_{k=0}^{25} x_{i-k} - x_{i+k} $
T19	$\sum_{i=11}^{N-10} \left(\left(1 + \sum_{k=0}^{10} x_{i-k} - x_{i+k} \right) / \left(1 + \sum_{k=-10}^9 x_{i-k} - x_{i+k} \right) \right)$
T20	$\sum_{i=11}^{N-10} \sqrt{\left(\sum_{k=0}^{10} x_{i-k} - x_{i+k} \right)^2 / \left(1 + \sum_{k=-10}^9 x_{i-k} - x_{i+k} \right)}$
T21	$\sum_{i=1}^{N-2} x_i - 2x_{i+1} + x_{i+2} $

T22	$\sum_{i=1}^{N-3} x_i - 3x_{i+1} + 3x_{i+2} - x_{i+3} $
T23	$\sum_{i=1}^{N-4} x_i - 4x_{i+1} + 6x_{i+2} - 4x_{i+3} + x_{i+4} $
T24	$\sum_{i=1}^{N-5} x_i - 5x_{i+1} + 10x_{i+2} - 10x_{i+3} + 5x_{i+4} - x_{i+5} $
T25	$\sum_{i=1}^{N-2} x_i - 2x_{i+1} + x_{i+2} x_{i+1}$
T26	$\sum_{i=1}^{N-3} x_i - 3x_{i+1} + 3x_{i+2} - x_{i+3} x_{i+1}$
T27	$\sum_{i=1}^{N-4} x_i - 4x_{i+1} + 6x_{i+2} - 4x_{i+3} + x_{i+4} x_{i+2}$
T28	$\sum_{i=1}^{N-5} x_i - 5x_{i+1} + 10x_{i+2} - 10x_{i+3} + 5x_{i+4} - x_{i+5} x_{i+2}$

Table E.1 – Trace functionals, T . N represents the number of points along trace and \mathbf{x}_i is the i^{th} sample.

P1	$\max_{i=1}^N x_i$
P2	$\min_{i=1}^N x_i$
P3	$\sqrt{\sum_{i=1}^N x_i^2}$
P4	$(\sum_{i=1}^N ix_i) / (\sum_{i=1}^N x_i)$
P5	$\sum_{i=1}^N ix_i$
P6	$\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$
P7	$\sum_{i=1}^{N-1} x_{i+1} - x_i $
P8	$\sum_{i=1}^{N-4} x_i - 4x_{i+1} + 6x_{i+2} - 4x_{i+3} + x_{i+4} $

Table E.2 – Diametric functionals, P .

$\Phi 1$	$\sum_{i=1}^{N-1} x_{i+1} - x_i ^2$
$\Phi 2$	$\sum_{i=1}^{N-1} x_{i+1} - x_i $
$\Phi 3$	$\sqrt{\sum_{i=1}^N x_i^2}$
$\Phi 4$	$\sum_{i=1}^N x_i$
$\Phi 5$	$\max_{i=1}^N x_i$
$\Phi 6$	$\max_{i=1}^N x_i - \min_{i=1}^N x_i$
$\Phi 7$	Amplitude of the first harmonic
$\Phi 8$	Phase of the first harmonic
$\Phi 9$	Amplitude of the second harmonic
$\Phi 10$	Phase of the second harmonic
$\Phi 11$	Amplitude of the third harmonic
$\Phi 12$	Phase of the third harmonic
$\Phi 13$	Amplitude of the fourth harmonic
$\Phi 14$	Phase of the fourth harmonic

Table E.3 – Circus functionals, Φ .

Appendix F: Retrieval Results

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	Lower Quartile																													
3	3	18	119	54	12	28	95	1	68	47	34	90	86	35	59	115	6	100	29	66	16	25	108	76	2	45	23	88	43	17
87	87	51	117	110	104	14	48	60	35	31	39	93	4	62	28	83	67	57	13	8	9	41	99	91	119	52	80	33	56	1
55	55	10	37	81	79	42	101	65	53	15	63	50	43	24	109	113	96	91	11	29	41	19	30	111	32	78	83	26	88	97
84	84	62	108	17	100	95	38	45	23	25	47	115	90	18	119	54	86	34	3	74	89	6	77	59	20	76	27	9	73	28
26	26	7	102	109	11	65	50	40	24	94	111	69	36	120	88	2	72	10	101	43	53	55	46	63	79	91	83	81	85	29
61	61	118	4	114	67	60	117	8	13	93	39	48	31	14	110	33	70	9	21	104	52	57	87	111	63	41	50	83	101	91
	Median																													
13	13	93	39	31	8	33	70	14	60	21	52	57	4	67	110	9	117	92	104	61	118	114	87	48	51	62	80	56	41	83
81	81	43	91	10	41	55	65	83	37	29	101	63	80	53	42	109	50	111	59	35	6	34	88	68	28	2	90	3	79	1
34	34	59	90	6	54	45	115	100	95	89	86	3	17	68	18	29	38	28	108	76	119	1	47	74	20	77	12	73	25	23
57	57	9	93	39	33	13	70	31	110	8	52	104	21	14	60	92	67	117	4	62	87	51	61	80	118	56	35	48	114	84
21	21	52	31	13	92	70	14	33	8	39	93	57	60	4	9	110	67	117	118	114	61	104	87	51	48	62	56	80	83	41
37	37	55	10	42	32	81	53	79	30	78	96	101	63	113	15	19	58	5	24	65	43	29	50	91	11	41	109	111	6	59
	Upper Quartile																													
27	27	74	73	76	77	23	38	20	86	25	115	89	45	100	17	54	90	108	34	47	59	18	6	3	95	119	12	84	68	66
64	64	112	105	116	49	82	22	71	75	103	98	44	56	99	80	107	83	35	1	41	68	91	104	51	28	87	95	43	39	81
115	115	86	100	45	76	74	89	38	54	23	25	90	73	77	20	27	17	34	108	59	47	6	95	18	3	119	68	12	1	28
116	116	22	75	82	103	112	71	49	64	44	98	105	56	99	107	80	1	35	83	68	51	41	104	91	28	95	87	39	57	33
96	96	113	19	79	15	30	42	32	78	58	37	97	5	55	10	53	101	63	24	85	50	81	65	111	11	109	26	43	91	29
105	105	64	112	49	116	71	82	98	75	103	22	56	44	99	80	107	83	35	41	1	104	91	68	51	28	87	39	33	93	57

Table F.1- Retrieval results for query textures within a 4-d MFS(75 features)for the Tex1 dataset

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>	<i>17</i>	<i>18</i>	<i>19</i>	<i>20</i>	<i>21</i>	<i>22</i>	<i>23</i>	<i>24</i>	<i>25</i>	<i>26</i>	<i>27</i>	<i>28</i>	<i>29</i>	<i>30</i>
	Lower Quartile																													
3	3	18	12	119	23	108	28	35	38	25	115	47	76	17	45	54	95	66	16	27	34	43	2	73	86	100	68	81	74	90
87	87	51	110	41	91	28	35	117	14	52	48	4	12	31	60	67	99	61	80	83	8	3	13	93	119	18	21	107	118	56
55	55	37	30	32	96	79	5	15	19	58	78	113	97	24	85	81	42	10	41	43	65	91	53	88	101	109	63	111	50	83
84	84	62	86	25	1	66	16	95	38	3	68	17	18	119	100	108	115	12	28	35	54	74	76	4	9	81	34	110	60	67
26	26	69	72	11	7	102	36	94	120	40	46	2	88	109	29	106	24	81	16	85	66	91	43	55	41	83	10	97	48	28
61	61	52	31	13	14	93	114	39	4	8	21	60	67	92	117	48	110	9	118	33	70	104	87	57	51	41	91	80	28	35
	Median																													
13	13	93	39	31	8	61	110	57	9	21	33	70	104	92	4	52	117	60	67	48	114	118	14	87	83	80	51	41	56	81
81	81	65	43	41	53	101	63	111	50	42	109	10	91	55	83	88	54	2	24	28	29	35	59	6	34	80	37	96	106	12
34	34	90	59	76	17	27	74	6	100	23	89	95	45	73	20	77	38	115	86	108	25	54	3	81	18	43	47	68	41	1
57	57	104	33	70	92	8	9	21	39	13	31	93	110	52	61	4	60	67	87	117	84	62	14	118	48	114	12	81	80	1
21	21	92	8	104	57	33	70	52	9	31	39	60	67	4	61	13	14	93	110	118	114	117	48	87	51	41	80	91	81	56
37	37	55	30	32	79	5	15	19	58	78	113	96	97	24	85	81	42	10	41	65	59	91	6	43	34	90	53	83	101	63
	Upper Quartile																													
27	27	45	38	73	115	34	100	74	17	76	20	77	23	89	54	90	95	86	59	6	108	3	25	47	18	68	43	12	81	66
64	64	112	105	56	80	99	71	82	98	103	49	83	44	75	116	22	107	41	91	81	1	68	28	35	117	51	43	93	87	39
115	115	45	27	38	73	54	20	77	108	100	47	17	34	76	89	3	86	23	59	74	95	90	6	25	18	12	43	68	119	66
116	116	71	22	75	44	49	82	98	103	107	105	64	112	41	91	80	56	28	35	99	1	68	51	83	81	87	47	117	43	14
96	96	55	42	10	37	24	85	79	30	32	53	101	5	15	19	50	58	78	113	63	111	65	97	81	109	43	41	91	83	7
105	105	64	112	56	80	82	98	103	71	49	44	75	99	116	22	83	107	41	91	81	1	68	28	35	117	93	39	51	33	43

Table F.2 - Retrieval of the 30 most similar textures to query textures (1st cell-bold) from the FPS of the Tex1 Dataset

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	Lower Quartile																													
112	112	2228	2430	1456	1834	1416	182	939	391	922	2605	1439	2323	928	133	161	2554	1390	204	1442	213	1952	1440	2394	1502	2535	2356	2357	2444	2587
1390	1390	928	1440	1502	391	2282	2357	28	1416	1397	2605	2587	2323	939	1456	2219	112	2394	2430	161	1439	1401	2233	1495	2228	227	1442	1555	69	922
28	28	1397	2282	1502	1440	1390	2219	928	2587	1401	391	1495	227	914	2357	2233	69	1555	1416	2430	939	2323	2605	112	1456	161	1385	2542	2228	2394
922	922	213	1439	1442	2228	2356	161	2605	1456	2394	939	112	1834	182	1952	204	2323	133	391	2444	1481	928	1753	12	2554	2357	321	2535	923	1954
250	250	204	1792	355	1952	133	1521	182	1834	2541	2535	193	2554	321	1481	2607	2444	1753	1419	12	2228	1456	213	2356	924	923	922	112	1507	1954
2233	2233	227	2219	1555	69	1495	1401	2282	1397	914	1502	1440	2587	1413	2542	28	1970	1551	1765	1385	1390	391	2040	928	2303	2323	1416	2357	939	1966
	Median																													
193	193	1419	2541	1748	2535	2382	1821	1792	321	133	2444	12	1753	2547	1481	353	250	204	2018	2607	355	330	1834	1954	1952	2314	1521	2365	1456	2552
2394	2394	2605	1442	1439	161	922	2357	2323	939	213	2356	1456	391	928	2228	1390	2444	2314	1954	12	1753	112	1440	330	1834	1481	321	1416	923	2535
1439	1439	1442	161	2605	922	2394	213	1456	2228	2356	939	2323	112	391	928	2357	1834	133	182	2444	1390	12	1952	1416	1753	204	1481	2535	321	1954
	Upper Quartile																													
321	321	12	1753	2444	1481	2535	330	1954	2314	193	1792	1834	923	2323	1507	204	1456	1521	939	182	133	2356	915	924	1419	250	213	922	1952	1748
1419	1419	193	2541	2382	1748	1821	2547	1792	133	353	2535	2365	321	2018	2607	2552	355	2444	12	1753	204	250	1952	1481	330	2554	1834	1954	1521	1456
1954	1954	330	2314	915	12	2444	1753	321	923	1507	1481	924	2356	2535	2323	1521	213	939	1456	193	1792	2394	1834	204	1442	922	2605	1439	182	1419
2607	2607	2541	2365	353	2547	355	2552	2018	1419	1821	193	250	133	1792	2554	2382	1952	1748	204	2535	1521	1834	321	182	2444	1753	12	1481	1456	2228
1413	1413	1765	1970	2303	2542	1551	1385	2040	1966	227	361	69	85	1555	2233	1495	2234	914	2219	1401	1397	2282	1410	2587	1502	28	1440	391	1390	2323
1765	1765	1970	2542	1551	1385	1413	2040	361	2303	69	1555	1966	914	227	1495	2234	2233	85	2219	1401	1397	2282	2587	1502	28	1440	1410	391	1416	2323

Table F.3 - Retrieval results for query textures within a 4-d MFS(65 features)

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>	<i>17</i>	<i>18</i>	<i>19</i>	<i>20</i>	<i>21</i>	<i>22</i>	<i>23</i>	<i>24</i>	<i>25</i>	<i>26</i>	<i>27</i>	<i>28</i>	<i>29</i>	<i>30</i>
	Lower Quartile																													
112	112	2554	2430	1416	2357	1456	133	2228	1952	391	1834	928	182	922	939	204	1439	213	2394	2535	1442	2356	2607	250	1792	161	2444	321	1753	12
1390	1390	1502	1401	2219	1410	1440	2282	2323	1397	2357	227	928	2605	28	1555	1495	1456	2587	2394	2233	391	1439	914	69	939	1413	161	1442	2430	1834
28	28	1397	391	928	1456	1502	2282	1401	1440	2587	1390	1495	2323	1410	2357	2219	2430	939	1555	2605	2394	227	2228	112	69	1439	1834	914	922	161
922	922	213	2356	2394	1442	1439	161	1952	1456	939	391	2357	928	1481	2535	112	2228	182	204	1834	12	2444	321	1753	2605	923	1416	1954	915	924
250	250	1792	1521	182	355	204	1952	2535	133	321	2444	1753	12	193	1481	1834	2356	2554	112	1416	213	923	924	330	1954	1419	939	1507	2541	922
2233	2233	227	1413	1765	69	85	1555	2234	2219	1966	2542	914	1551	1970	1385	1495	2040	2303	1401	361	1410	2587	1502	1440	1397	2282	2323	1390	28	2605
	Median																													
193	193	1419	2382	1821	1748	2535	1753	2547	12	2444	321	2552	1416	2541	2365	1792	1481	2607	355	250	353	1952	2018	1521	330	1954	133	182	204	1834
2394	2394	161	1439	1442	2605	922	2357	213	928	391	1456	2356	939	2228	1834	2535	1753	2323	12	2444	321	112	1481	1502	1952	2282	1954	2314	28	1390
1439	1439	161	2394	1442	2357	2605	213	922	928	1456	391	2356	1834	2228	112	939	2535	1952	1753	2554	2323	12	2444	321	2430	1481	1390	2282	28	182
	Upper Quartile																													
321	321	2444	2535	12	1753	1481	330	1954	2314	923	1521	1507	939	915	924	193	2356	1834	250	1792	1456	213	204	182	922	2323	1419	2394	2382	1952
1419	1419	193	1821	2382	1748	2535	2547	2552	2541	2607	2365	353	1416	2018	355	2444	321	12	1753	1792	1481	133	250	1952	1521	182	330	1954	204	1456
1954	1954	330	2314	915	12	2444	321	1753	1481	923	939	1507	2535	1521	1834	2356	924	193	213	1456	922	2323	182	250	1792	204	2394	1442	2228	2357
2607	2607	2541	2365	2547	353	2552	2018	1416	1419	1821	355	193	2382	2535	133	2554	112	1748	1952	1792	250	2444	182	1834	321	12	204	1753	2430	1481
1413	1413	1765	2234	2233	85	227	1966	1970	2542	1555	69	1385	1551	361	2040	2303	2219	914	2587	1495	2282	1502	2323	1397	1401	1410	1390	1440	28	2535
1765	1765	2234	1413	1970	1966	85	2233	69	2542	1385	1551	361	2040	2303	914	227	1555	2587	1495	1397	1401	2219	2282	2323	1502	1410	28	1440	1390	12

Table F.4 - Retrieval of the 30 most similar textures to query textures (*1st* cell-bold) from the FPS of the MoMA Dataset

References

- [Abbadeni05] N. Abbadeni, "Perceptual Image Retrieval", in *8th International Conference, VISUAL 2005*, pp. 259-268, 2005.
- [Ahonen04] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Recognition with Local Binary Patterns", *Lecture Notes in Computer Science: Computer Vision - ECCV 2004*, pp. 469-481, 2004.
- [Amadasun89] M. Amadasun and R. King, "Textural features corresponding to textural properties", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19(5), pp. 1264-1274, 1989.
- [Aksoy01] S. Aksoy and R. M. Haralick, "Feature Normalization and Likelihood-based Similarity Measures for Image Retrieval", *Pattern Recognition Letters*, Vol. 22(5), pp. 563-582, 2001.
- [Baker75] F. B. Baker and L. J. Hubert, "Measuring the Power of Hierarchical Cluster Analysis", *Journal of the American Statistical Association*, Vol. 70(349), pp. 31-38, 1975.
- [Beck87] J. Beck, A. Sutter, and R. Ivry, "Spatial Frequency Channels and Perceptual Grouping in Texture Segregation", *Computer Vision, Graphics, and Image Processing*, Vol. 37, pp. 299-325, 1987.
- [Bergen88] J. R. Bergen and E. H. Adelson, "Early vision and texture perception", *Nature*, Vol. 333(6171), pp. 363-364, 1988.
- [Bhushan97] N. Bhushan, A. R. Rao, and G. L. Lohse, "The Texture Lexicon: Understanding the Categorization of Visual Texture Terms and Their Relationship to Texture Images", *Cognitive Science*, Vol. 21(2), pp. 219-246, 1997.
- [Bigun94] J. Bigun and J. M. Du Buf, "N- folded symmetries by complex moments in the Gabor space and their application to unsupervised texture segmentation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16(1), 1994.
- [Blum97] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning", *Artificial Intelligence*, Vol. 97, Issue 1-2, pp. 245-271, 1997.

- [Bovik90] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel Filtering Analysis Using Localized Spatial Filters", *IEEE Trans. Pattern and Machine Intelligence*, Vol. 12(1), pp. 55-73, January 1990.
- [Brodatz66] P. Brodatz, "Textures - a photographic album for artists and designers", Dover, New York, 1966.
- [Caelli78] T. Caelli and B. Julesz, "On perceptual analyzers underlying visual texture discrimination", *Biological Cybernetics*, Vol. 28, pp. 167-175, 1978.
- [Chantler94] M. J. Chantler, "Why illuminant direction is fundamental to texture analysis", *IEE Proc. on Visual Image and Signal Processing*, Vol. 142(4), pp. 199-206, 1994.
- [Chantler05] M. J. Chantler, M. Petrou, A. Penirschke, M. Schmidt, and G. McGunnigle, "Classifying Surface Texture While Simultaneously Estimating Illumination", *Int'l Journal of Computer Vision (VISI)*, Vol. 62(1-2), pp. 83-96, 2005.
- [Clausi00] A. Clausi, and M. E. Jernigan, "Designing Gabor filters for optimal texture separability" *Pattern Recognition*, Vol. 33(11), pp. 1771-1933, November 2000.
- [Coggins85] J. M. Coggins and A. K. Jain, "A spatial filtering approach to texture analysis", *Pattern Recognition Letters*, Vol. 3(3), pp. 195-203, 1985.
- [Connors79] R. W. Connors, "Towards a Set of Statistical Features which Measure Visually Perceivable Qualities of Textures", in *Proceedings, IEEE Conference on Pattern Recognition and Image Processing*, pp. 382-390, 1979.
- [Cox00] T. F. Cox and M. A.A. Cox, "Multidimensional Scaling". Chapman & Hall/CRC, 2000.
- [Dai04] Y. Dai and D. Cai, "Visual perception-based structure analysis of images for digital collection retrieval, in *IEEE Int. Conf. on Systems, Man and Cybernetics*, Vol. 1, pp. 1104- 1111, 2004.
- [Davis79] L. S. Davis, S. Johns and J. K. Aggarwal, "Texture Analysis Using Generalized Co-Occurrence Matrices", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 1(3), pp.251-259, 1979.
- [Datta08] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image Retrieval: Ideas,

Influences, and Trends of the New Age”, *ACM Computing Surveys*, 2008. (to appear)

- [DeBonet97] J. S. De Bonet, “Multiresolution sampling procedure for analysis and synthesis of texture images”, In *SIGGRAPH 97*, pp. 361-368, 1997.
- [Derin87] H. Derin and H. Elliott, “Modelling and segmentation of noisy and textured images using Gibbs random fields”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9(1), pp. 39-55, 1987.
- [Dewangan05] D. Dewangan, V. J. Samar, R. Rao, and P. Paul, “Factors influencing psychophysically valid taxonomies of image texture”, in *IEEE International Conference on Image Processing, ICIP 2005*, Vol. 3, pp. 1196-1199, 2005.
- [Ding02] C. Ding and X. He, “Cluster merging and splitting in hierarchical clustering algorithms”, in *Proceedings IEEE Int. Conf. on Data Mining*, pp. 139-146, 2002.
- [Dong03] J. Dong, “Three-dimensional Surface Texture Synthesis”, Ph.D. Thesis, Heriot-Watt University, 2003.
- [Dong05] J. Dong and M. J. Chantler, “Capture and Synthesis of 3D Surface Texture”, *International Journal of Computer Vision (VISI)*, 62(1-2), pp. 177-194, 2005.
- [DuBuf90] J. M. H. Du Buf, “Gabor phase in texture discrimination”, *Signal Processing*, Vol. 21(3), pp. 221-240, November 1990.
- [DuBuf91] J. M. H. Du Buf and P. Heitkamper, “Texture feature based on Gabor phase”, *Signal Processing*, Vol. 23, pp. 227-244, 1991.
- [Efros99] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling”, in *Proc. 7th IEEE International Conference on Computer Vision (ICCV)*, Vol. 2, pp.1033-1038, 1999.
- [Efros01] A. A. Efros and W T. Freeman, “Image Quilting for Texture Synthesis and Transfer”, in *Proc. of SIGGRAPH '01, Los Angeles, California*, pp. 341-346, August 2001.
- [Fraley98] C. Fraley and A. E. Raferty, “How many clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis?”, *The Computer Journal*, Vol. 41 (8), pp. 578-588, 1998.

- [Franco04] A. Franco, A. Lumini, D. Maio, "A new approach for relevance feedback through positive and negative samples", in *Proceedings, Int'l Conf. on Pattern Recognition*, Vol. 4, pp. 905 – 908, 2004.
- [Frankot88] R. T. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10(4), pp. 439-451, 1988.
- [Gagalowicz85] A. Gagalowicz and S.D. Ma, "Sequential synthesis of natural textures", *Computer Vision, Graphics and Image Processing*, Vol. 30(3), 1985.
- [Galloway75] M. M. Galloway, "Texture analysis using gray level run lengths", *Computer vision, graphics, and image processing*, Vol. 4, pp. 172-179, 1975.
- [Gibson50] J. J. Gibson, "The Perception of Visual Surfaces", *The American Journal of Psychology*, Vol. 63(3), pp. 367-384, 1950.
- [Gluckman05] J. Gluckman, "Visually distinct patterns with matching subband statistics", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27(2), pp. 252-264, Feb. 2005.
- [Gordon87] A. D. Gordon, "A Review of Hierarchical Classification", *Journal of Royal Statistical Society*, Part 2, pp. 119-137, 1987.
- [Graham92] N. Graham, J. Beck, and A. Sutter, "Nonlinear processes in spatial-frequency channel models of perceived texture segregation: effects of sign and amount of contrast", *Vision Research*, Vol. 32(4), pp. 719-743, 1992.
- [Gullón03] C. Gullón, "Height Recovery of Rough Surfaces from Intensity Images", Ph.D. Thesis, Heriot-Watt University, 2003.
- [Gurnsey01] R. Gurnsey and D. J. Fleet, "Texture space", *Vision Research*, Vol. 41, pp. 745–757, 2001.
- [Haralick73] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification", *IEEE Trans. Systems, Man and Cybernetics*, Vol. SMC-3, pp. 610-662. Nov. 1973.
- [Haralick79] R. M. Haralick, "Statistical and Structural Approaches to Texture", in *Proceedings of the IEEE*, Vol. 67(5), May 1979.

- [Harvey81] L. O. Harvey, Jr., and M. J. Gervais, "Internal Representation of Visual Texture as the Basis for the Judgment of Similarity", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 7(4), 741-753, 1981.
- [Haynes04] J. D. Haynes, R. B. Lotto, and G. Rees, "Responses of human visual cortex to uniform surfaces", in *Proceedings of the National Academy of Sciences of the United States of America*, Vol.101(12), pp.4286 – 4291, 2004.
- [Heaps99] C. Heaps and S. Handel, "Similarity and features of natural textures". *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 25, pp. 299–320, 1999.
- [Heeger95] D. J. Heeger and J. R. Bergen, "Pyramid-Based texture analysis/synthesis", In *SIGGRAPH 95 Conference Proceedings*, pp. 229–238, 1995
- [Ho08] Y. Ho, M. S. Landy, and L. T. Maloney, "Conjoint Measurement of Gloss and Surface Texture", *Psychological Science*, Vol. 19(2), pp. 196–204, 2008.
- [Hubel68] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex", *Journal of Physiology*, Vol.195, pp. 215-243, 1968.
- [Iqbal99] Q. Iqbal and J. K. Aggarwal, "Applying perceptual grouping to content-based image retrieval: Building images", in *Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, pp. 42-48, 1999.
- [Jain91] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters", *Pattern Recognition*, Vol. 24(12), pp. 1167-1186, 1991.
- [Jain99] A.K Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, Vol. 31(3), pp.264–323, 1999.
- [Johnson04] A. P. Johnson and C. L. Baker, Jr., "First- and second-order information in natural images: a filter-based approach to image statistics", *Journal of the Optical Society of America A*, Vol. 21(6), 2004.
- [Julesz62] B. Julesz, "Visual Pattern Discrimination," *IRE Transaction on Information Theory*, IT-8, pp. 84-92, 1962.

- [Julesz75] B. Julesz, "Experiments in the visual perception of texture," *Scientific American*, 232, pp. 34-43, 1975.
- [Julesz81] B. Julesz, "Textons, the Elements of Texture Perception, and Their Interactions," *Nature*, 290, pp. 91-97, 1981.
- [Kadyrov01] A. Kadyrov and M. Petrou, "The Trace Transform and Its Applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23(8), pp. 811-828, 2001.
- [Kadyrov02] A. Kadyrov, A. Talebpour and M. Petrou, "Texture classification with thousands of features", in *Proceedings of the British Machine Vision Conference*, 2002.
- [Kherfi03] M. L. Kherfi, D. Ziou and A. Bernardi, "Combining positive and negative examples in relevance feedback for content-based image retrieval", *Journal of Visual Communication and Image Representation*, Vol. 14(4), pp. 428-457, 2003.
- [Kingsbury99] N.G. Kingsbury, "Image Processing with Complex Wavelets", *Philosophical Transactions of the Royal Society London*, pp. 210–223, 1999.
- [Kohavi95] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", in *Proc. of the Fourteenth Int'l Joint Conference on Artificial Intelligence, IJCAI 95*, pp. 1137-1145, 1995.
- [Kourtzi06] Z. Kourtzi, "Textures of Natural Images in the Human Brain. Focus on 'Orientation-Selective Adaptation to First- and Second-Order Patterns in Human Visual Cortex'", *Journal of Neurophysiology*, Vol. 95, pp. 591-592, 2006.
- [Kruskal64a] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", *Psychometrika* (29), pp. 1-27, 1964.
- [Kruskal64b] J. B. Kruskal, "Nonmetric Multidimensional scaling: A numerical method", *Psychometrika* (29), pp. 115-129, 1964.
- [Kube88] P. Kube and A.P Pentland, "On the Imaging of Fractal Surfaces" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10(5), pp. 704-707, 1988.
- [Kwatra03] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick, "Graphcut

textures: Image and video synthesis using graph cuts”, In *ACM SIGGRAPH*, pp. 277-286, 2003.

- [Landy91] M. S. Landy and J. R. Bergen, “Texture segregation and orientation gradient”, *Vision Research*, Vol. 31(4), pp.679-691, 1991.
- [Landy04] M. S. Landy and N. Graham, “Visual perception of texture”, *The Visual Neurosciences*, pp. 1106-1118, Cambridge, MA: MIT Press, 2004.
- [Laws80] K. I. Laws, “Rapid texture identification”, In *Proc. of the SPIE Conference on Image Processing for Missile Guidance*, pp. 376-380, 1980.
- [Levenberg44] K. Levenberg, “A method for the solution of certain non-linear problems in least squares”, *The Quarterly of Applied Mathematics*, Vol. 2, pp. 164-168, 1944.
- [Lew06] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, “Content-based multimedia information retrieval: State of the art and challenges”, *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 2(1), pp. 1 – 19, 2006.
- [Liang01] L. Liang, C. Liu, Y. Xu, B. Guo, and H.Y. Shum, “Real-time texture synthesis by patch-based sampling”, *ACM Transactions on Graphics*, Vol. 20(3), pp. 127–150, 2001.
- [Liu05] H. Liu, E. R. Dougherty, J. G. Dy, K. Torkkola, E. Tuv, H. Peng, C. Ding, F. Long, M. Berens, L. Parsons, Z. Zhao, L. Yu, and G. Forman, "Evolving Feature Selection," *IEEE Intelligent Systems*, Vol. 20(6), pp. 64-76, 2005.
- [Liu07] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, “A survey of content-based image retrieval with high-level semantics” *Pattern Recognition*, Vol. 40, pp. 262 – 282, 2007.
- [Long00] H. Long, W. K. Leow, and F. K. Chua, “Perceptual texture space for content-based image retrieval”, in *Proc. Int. Conf. on Multimedia Modelling*, pp. 167-180, 2000.
- [Long01] H. Long, C. W. Tan, and W. K. Leow, “Invariant and perceptually consistent texture mapping for content-based image retrieval”, in *Proceedings, Int’l Conf. on Image Processing*, Vol. 2, pp. 117-120, 2001.

- [Lowe85] D. G. Lowe, "Perceptual organization and visual recognition", *Kluwer Academic publishers*, 1985.
- [Mallat89] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11(7), pp. 674-693, 1989.
- [Malik90] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms" *Journal of the Optical Society of America A*, Vol. 7(5), pp. 923-932, 1990.
- [Marquardt63] D.W. Marquardt, "An Algorithm for the Least-Squares Estimation of Nonlinear Parameters", *SIAM Journal of Applied Mathematics*, Vol. 11(2), pp.431-441, 1963.
- [Matsuyama83] T. Matsuyama, S. I. Miura, and M. Nagao. "Structural analysis of natural textures by Fourier transformation", *Computer Vision, Graphics and Image Processing*, Vol. 24(3), pp. 347-362, 1983.
- [McGunnigle01] G. McGunnigle and M. J. Chantler, "Evaluating Kube and Pentland's fractal imaging model", *IEEE Trans. on Image Processing*, Vol. 10(4), pp. 534-542, 2001.
- [Mingolla86] E. Mingolla and J. T. Todd, "Perception of solid shape from shading" *Biological Cybernetics*, Vol. 53, pp. 137-151, 1986.
- [Molinaro05] A. M. Molinaro, R. Simon and R. M. Pfeiffer, "Prediction error estimation: a comparison of resampling methods", *Bioinformatics*, Vol. 21(15), pp. 3301-3307, 2005.
- [Ojala96] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions". *Pattern Recognition*, Vol. 29(1), pp. 51-59, January 1996.
- [Ojala02] T. Ojala, M. Pietikainen, and M. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 971-987, 2002.
- [Oppenheim91] A. V. Oppenheim and J. S. Lim, "The Importance of Phase in Signals", in *Proceedings of the IEEE*, Vol. 69 (5), May 1991.
- [Paget96] R. Paget and I. D. Longstaff, "A nonparametric multiscale Markov random field model for synthesising natural textures", in *Fourth*

International Symposium on Signal Processing and its Applications (ISSPA), Vol. 2, pp. 744-747, 1996.

- [Paget98] R. Paget and I. D. Longstaff, "Texture synthesis via a noncausal nonparametric multiscale Markov random field", *IEEE Transactions on Image Processing*, Vol. 7(6), pp. 925-931, 1998.
- [Payne99] J. S. Payne, L. Hepplewhite, and T. J. Stonham, "Perceptually Based Metrics for the Evaluation of Textural Image Retrieval Methods", in *IEEE Int'l Conf. on Multimedia Computing and Systems, ICMCS99*, Vol. 2, pp. 793-797, 1999.
- [Payne05] J. S. Payne and T. J. Stonham, "Mapping Perceptual Texture Similarity for Image Retrieval", in *Scandinavian Conf. in Image Analysis, SCIA2005*, pp. 960-969, 2005.
- [Pentland84] A. P. Pentland, "Fractal-based description of natural scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 661-674, 1984
- [Petrou04] M. Petrou and A. Kadyrov, "Affine Invariant Features from the Trace Transform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26(1), pp. 30-44, 2004.
- [Petrou07] M. Petrou, A. Talebpour, and A. Kadyrov, "Reverse engineering the way humans rank textures", *Pattern Analysis and Applications*, Vol. 10(2), pp. 101-114, 2007.
- [Porat89] M. Porat, and Y. Zeevi, "Localized Texture Processing in Vision: Analysis and Synthesis in the Gaborian Space", *IEEE Trans. On Biomedical Engineering*. Vol. 36(1), January 1989.
- [Portilla00] J. Portilla and E. P. Simoncelli, "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients", *IJCV*, 40(1), pp. 49-71, 2000.
- [Rao93a] A. R. Rao and G. L. Lohse, "Identifying high level features for Texture Perception", *CVGIP: Graphical Models and Image Processing*, Vol. 55(3), pp. 218-233, 1993.
- [Rao93b] A. R. Rao and G. L. Lohse, "Towards a Texture Naming System: Identifying Relevant Dimensions of Texture", in *IEEE Conference on Visualization*, pp. 220-227, October 1993.
- [Randen99] T. Randen and H. J. Husoy, "Filtering for Texture Classification: A

Comparative Study”, *IEEE Trans. Pattern and Machine Intelligence*, vol. 21(4), pp. 291-310, April 1999.

- [Rogowitz98] B. E. Rogowitz, T. Frese, J. Smith, C. A. Bouman, and E. Kalin, “Perceptual Image Similarity Experiments” in *Proc. of the SPIE, 3299, Conference on Human Vision and Electronic Imaging III, San Jose, California*, pp.576-590, January 1998.
- [Rui98] Yong Rui, T. S. Huang, and S. Mehrotra, “Relevance Feedback Techniques in Interactive Content-Based Image Retrieval” *Storage and Retrieval for Image and Video Databases (SPIE)*, pp. 25-36, 1998.
- [Rui99] Y. Rui, T. S. Huang and S. F. Chang, “Image Retrieval: Current Techniques, Promising Directions and Open Issues”, *Journal of Visual Communication and Image Representation*, Vol. 40(4), pp. 39-62, 1999.
- [Salton68] G. Salton and M. E. Lesk, “Relevance Assessments and Retrieval System Evaluation”, *Information Storage Retrieval*, Vol. 4(4), pp.343--359, 1968.
- [Sclaroff99] S. Sclaroff, M. La Cascia, and S. Sethi, “Unifying Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web”, *Computer Vision and Image Understanding*, Vol. 75, Nos. 1/2, pp. 86-98, 1999.
- [Shepard62] R. N. Shepard, “The analysis of proximities: Multidimensional scaling with an unknown distance function”, *Psychometrika* (27), pp. 219-246, 1962.
- [Simoncelli95] E. P. Simoncelli and W. T. Freeman, “The steerable pyramid: A flexible architecture for multi-scale derivative computation”, in *2nd IEEE Int’l Conf. on Image Processing, Washington DC*, Vol. III, pp. 444-447, Oct.1995.
- [Smeulders00] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-Based Image Retrieval at the End of the Early Years”, *IEEE Trans. on PAMI*, Vol. 22(12), pp. 1349-1380, 2000.
- [Tamura78] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” *IEEE Transactions Systems, Man and Cybernetics*, Vol. 8(6), pp. 460-473, 1978.
- [Todd97] J. T Todd, J. F. Norman, J. J. Koenderink, and A. M. Kappers,

- “Effects of texture, illumination, and surface reflectance on stereoscopic shape perception”, *Perception*, Vol. 26, pp. 807-822, 1997.
- [Tuceryan90] M. Tuceryan and A. K. Jain, “Texture Segmentation Using Voronoi Polygons,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12(2), pp. 211-216, 1990.
- [Tuceryan98] M. Tuceryan and A. K. Jain, “Texture Analysis”, *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, pp. 207-248, 1998.
- [Unser86] M. Unser, “Sum and difference histograms for texture classification”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, pp.336-357, 1986.
- [Veltkamp02] R.C. Veltkamp and M. Tanase,”Content-Based Image Retrieval Systems: A Survey”, Technical Report UU-CS, 2002.
- [VisTex95] “Vision Texture dataset”, Media Laboratory, MIT, <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>.
- [Voorhees87] Voorhees, H. and T. Poggio, “Detecting textons and texture boundaries in natural images,” In *Proc. of the First Int’l Conference on Computer Vision*, pp. 250-258, 1987.
- [Wang04] J. Wang and K. J. Dana, “Hybrid Textons: Modeling Surfaces with Reflectance and Geometry”, in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR’04)*, Vol. 1, pp. 372-378, 2004.
- [Wei00] L. Wei and M. Levoy, “Fast Texture Synthesis using Tree-structured Vector Quantization”, in *Proc. of SIGGRAPH 2000*, pp. 479-488, 2000.
- [Wenger97] R. Wenger, “Visual Art, Archaeology and Gestalt”, *Leonardo*, Vol. 30(1), pp. 35-46, 1997.
- [Woodham80] R. J. Woodham, “Photometric Method for Determining Surface Orientation from Multiple Images”, *Optical Engineering*, Vol. 19(1), pp.139-144, 1980.
- [Wu03] J. Wu, “Rotation Invariant Classification of 3D Surface Texture Using Photometric Stereo”, Ph.D. Thesis, Heriot-Watt University, 2003.

- [Yu03] L. Yu and H. Liu, "Efficiently Handling Feature Redundancy in High Dimensional Data", In *Proceedings of the Ninth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 685-690, 2003.
- [Yu04] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy", *Journal of Machine Learning Research*, Vol. 5, pp. 1205-1224, 2004.
- [Zhu98] S.C. Zhu, Y.N. Wu, and D. B. Mumford, "FRAME: Filters, Random fields And Maximum Entropy-towards a unified theory for texture modeling", *Int'l Journal of Computer Vision*, Vol. 27(2), pp. 1-20, 1998.
- [Zhu05] S. Zhu, C. Guo, Y. Wang and Z. Xu, "What are Textons?", *Int'l Journal of Computer Vision*, Vol. 62(1/2), pp. 121-143, 2005.
- [Zucker76] S. W. Zucker, "Toward a model of Texture", *Computer Graphics and Image Processing*, Vol. 5(2), pp. 190-202, 1976.
- [Zucker80] S. W. Zucker and D. S. Terzopoulos, "Finding Structure in Co-Occurrence Matrices for Texture Analysis" *Computer Graphics and Image Processing*, Vol. 12(3), pp. 286-308, 1980.