

Perceptually Relevant Browsing Environments for Large Texture Databases

Fraser Halley

Submitted for the degree of Doctor of Philosophy



Heriot Watt University

School of Mathematical and Computer Sciences

April 2012

The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

Abstract

This thesis describes the development of a large database of texture stimuli, the production of a similarity matrix reflecting human judgements of similarity about the database, and the development of three browsing models that exploit structure in the perceptual information for navigation. Rigorous psychophysical comparison experiments are carried out and the SOM (Self Organising Map) found to be the fastest of the three browsing models under examination. We investigate scalable methods of augmenting a similarity matrix using the SOM browsing environment to introduce previously unknown textures. Further psychophysical experiments reveal our method produces a data organisation that is as fast to navigate as that derived from the perceptual grouping experiments.

Acknowledgements

I owe a debt of gratitude to my esteemed supervisor, Mike Chantler, for his insight and encouragement throughout my PhD.

My research involved over 270 experimental sessions and I would like record my thanks to the observers who gave so generously of their time and effort for such meagre reward.

It would not have been possible for me to pursue a PhD without the financial support of the Engineering and Physical Sciences Research Council.

Finally, I would like to thank my colleagues in the Texture Lab, Heriot Watt University, for their support and camaraderie.

Contents

List of Figures	vi
List of Tables	viii
Table of Terms	1
1 Introduction	1
1.1 Motivation	1
1.2 Mission and Goals	2
1.3 Scope	3
1.4 Novelties & Contribution	4
1.5 Thesis Organisation	4
2 Survey	6
2.1 Introduction	6
2.2 Browsing Environments	6
2.3 Existing Datasets	15
2.4 Identifying Structure in Texture Databases	17
2.4.1 Capturing Human Judgements	18
2.4.2 Identifying Structure in Psychophysical Data	19
2.4.3 Image Collection Annotation	22
2.5 Mapping Computational Features to Perceptual Space	23
2.6 Crowdsourcing	24
2.7 Discussion	25
2.8 Conclusions	27
3 Dataset Development & Capturing Human Judgements	30
3.1 Introduction	30
3.2 Dataset Development	31
3.2.1 Specifications	31
3.2.2 Acquisition of Stimuli	35

3.3	Experimental Design	37
3.3.1	Perceptual Grouping Experiment	37
3.3.2	Scalable Grouping Experiment	42
3.4	Conclusions	45
4	Preliminary Analysis of Results	46
4.1	Introduction	46
4.2	Hierarchical Analysis	48
4.2.1	Dendrogram	48
4.3	Dimensional Analysis	53
4.3.1	Stress	53
4.3.2	Meaningful Dimensions	56
4.4	Neural Network Analysis	56
4.5	Conclusions	61
5	Design Browsing Environments	63
5.1	Introduction	63
5.2	Rapid-Fire Browsing Environment (RF_G)	64
5.3	MDS Browsing Environment (MDS_G)	67
5.4	SOM Browsing Environment (SOM_G)	69
5.5	Conclusions	71
6	SOM_G, MDS_G & RF_G Browsing Environments Comparison	73
6.1	Introduction	73
6.2	Experimental Design	73
6.2.1	Task Specification	73
6.2.2	Interface Presentation	74
6.2.3	Stimuli Selection	74
6.2.4	Sample Size	74
6.2.5	Performance Measurement	75
6.2.6	Observer Selection	75
6.2.7	Instructions to Observers	75
6.3	Analysis of Result	77
6.3.1	Mean Task Time	77
6.3.2	Accuracy	79
6.4	Discussion	81
6.5	Conclusions	82
7	Pilot Dataset Augmentation	83
7.1	Introduction	83

7.2	Relationship Between New Textures and Existing Textures	84
7.3	Inter-Relationship Between New Textures	86
7.4	Design Data Augmentation Experiment	87
7.4.1	Stimuli Selection	87
7.4.2	Design SOM _R Browsing Environment	88
7.4.3	Task Specification	88
7.4.4	Sample Size	88
7.4.5	Observer Selection	88
7.4.6	Instructions to Observers	88
7.5	Augmentation Experiment Data Analysis	89
7.5.1	Votes	89
7.5.2	Build SOM _A Browsing Environment	90
7.6	Conclusions	90
8	Identifying Features for Data Organisation	91
8.1	Introduction	91
8.2	Identifying Features for Texture Description	92
8.2.1	Feature Selection Criteria	92
8.2.2	Feature Extraction Methods	93
8.2.3	Feature Set Selection	94
8.2.4	Feature Normalisation	94
8.3	Producing a Feature-Based Dataset Description	94
8.3.1	Similarity Matrix Dimensionality Reduction	94
8.3.2	Feature Selection Training Set	95
8.3.3	Optimise Number of Features	95
8.3.4	Build SOM _F Browsing Environment	96
8.4	Conclusions	97
9	SOM_G, SOM_A & SOM_F Browsing Environments Comparison	100
9.1	Introduction	100
9.2	Experimental Design	101
9.2.1	Task Specification	101
9.2.2	Interface Presentation	101
9.2.3	Stimuli Selection	101
9.2.4	Sample Size	101
9.2.5	Performance Measurement	102
9.2.6	Observer Selection	102
9.2.7	Instructions to Observers	102
9.3	Analysis of Result	103

9.3.1	Mean Task Time	103
9.3.2	Statistical Significance and Experimental Effect	103
9.4	Conclusions	105
10	Scalable Dataset Augmentation	106
10.1	Introduction	106
10.2	New Grouping Experiment	107
10.3	Design Data Augmentation Experiment	110
10.3.1	Stimuli Selection	110
10.3.2	Design SOM _{R2} Browsing Environment	110
10.3.3	Task Specification	110
10.3.4	Sample Size	111
10.3.5	Payment Criteria	111
10.3.6	Observer Selection	113
10.3.7	Instructions to Observers	113
10.3.8	Completion Time	115
10.4	Augmentation Experiment Data Analysis	115
10.4.1	Votes	115
10.4.2	Build SOM _{A2} Browsing Environment	115
10.5	Conclusions	117
11	SOM_G & SOM_{A2} Browsing Environments Comparison	118
11.1	Introduction	118
11.2	Experimental Design	119
11.2.1	Task Specification	119
11.2.2	Interface Presentation	119
11.2.3	Stimuli Selection	119
11.2.4	Sample Size	119
11.2.5	Performance Measurement	120
11.2.6	Observer Selection	120
11.2.7	Instructions to Observers	120
11.3	Analysis of Result	121
11.3.1	Mean Task Time	121
11.3.2	Statistical Significance and Experimental Effect	121
11.4	Conclusions	123
12	Summary and Future Work	124
12.1	Summary of Thesis Argument	124
12.2	Future Work	128

A	Experimental Stimuli	131
A.1	<i>Tex500[024]</i>	131
A.2	<i>Tex500[100]</i>	132
A.3	<i>Tex500</i>	135
B	Detailed Experimental Results	150
B.1	SOM _G , MDS _G & RF _G Browsing Environments Comparison	150
B.1.1	Mean Task Time	150
B.1.2	Accuracy	151
B.2	Pilot Dataset Augmentation	151
B.3	Neurons Containing <i>Tex500[024]</i> Stimuli in SOM _F Browsing Environ- ment	167
B.4	SOM _G , SOM _A & SOM _F Browsing Environments Comparison	175
B.5	Crowdsourcing Dataset Augmentation	175
B.5.1	Acceptable Response Examples	175
B.5.2	Rejected Response Example	186
B.6	SOM _G & SOM _{A2} Browsing Environments Comparison	190
	Bibliography	191

List of Figures

1.1	Thesis Organisation	5
3.1	Example Textures from <i>Tex500</i> Dataset	32
3.2	Example Directional Textures from <i>Tex500</i> Dataset	33
3.3	Photometric Capture Set Up	37
3.4	Photometric stereo images of stimulus 067 illuminated at constant slant angle 60° and tilt angles 0° (left), 90° (bottom), 180° (right), 270° (top)	38
3.5	Height Map and Rendered Surface (Lambertian Slant 45° and Tilt 135°) for Stimulus 067	38
3.6	Printed Stimulus 067 (Front)	39
3.7	Printed Stimulus 067 (Reverse)	39
3.8	Grouping Experiment Setting	40
3.9	Pilot Experiment Similarity Matrix (Excerpt)	42
3.10	Opportunity Matrix (Excerpt)	44
3.11	Augmented Similarity Matrix (Excerpt)	44
4.1	Dendrogram	49
4.2	Partial Dendrogram (rightmost 50 textures from Figure 4.1)	50
4.3	Partial Dendrogram with Images (rightmost 9 textures from Figure 4.2)	52
4.4	Stress Plot	55
4.5	Updating the Best Matching Unit (BMU)	57
4.6	Local Lattice Structures	57
4.7	Map Shapes	58
4.8	Neural Network Produced by SOM Analysis	58
4.9	Top Left Numbered Neurons from Figure 4.8	59
4.10	Top Right Numbered Neurons from Figure 4.8	60
5.1	Rapid-Fire Tree Structure	65
5.2	RF _G Browsing Environment Top Level Screen Shot	66
5.3	RF _G Browsing Environment Next Level Screen Shot	67
5.4	RF _G Browsing Environment Detail Level Screen Shot	67

5.5	MDS _G Browsing Environment Top Level Screen Shot	68
5.6	MDS _G Browsing Environment Detail Level Screen Shot	69
5.7	SOM _G Browsing Environment Top Level Screen Shot	70
5.8	SOM _G Browsing Environment Detail Level Screen Shot	70
6.1	Mean Task Time (with Standard Error Bars) from SOM _G , MDS _G & RF _G Browsing Environments Comparison Experiment	78
6.2	Non-Exact Observer Selections in SOM _G , MDS _G & RF _G Browsing Environments Comparison Experiment	80
6.3	Accuracy (with Standard Error Bars) from SOM _G , MDS _G & RF _G Browsing Environments Comparison Experiment	81
7.1	Mapping of Measured (m), Derived (d) and Double Derived (dd) Similarity Values in the Augmented Similarity Matrix	87
8.1	Alignment Error Analysis Plot	96
8.2	SOM _F Browsing Environment Top Level Screen Shot	97
8.3	Neuron from SOM _F with Similar Members and Intuitive Exemplar . .	98
8.4	Neuron from SOM _F with Dissimilar Members and Unintuitive Exemplar	98
9.1	Mean Task Time (with Standard Error Bars) from SOM _G , SOM _A & SOM _F Browsing Environments Comparison Experiment	104
10.1	Dendrogram from Grouping Experiment on <i>Tex500[100]</i>	108
10.2	Dendrogram sub-sampled from <i>Tex500</i> Grouping Experiment	109
10.3	SOM _{R2} Browsing Environment Screen Capture	111
10.4	Crowdsourcing Augmented <i>Tex500[100]</i> Dendrogram	116
11.1	Mean Task Time (with Standard Error Bars) from SOM _G & SOM _{A2} Browsing Environments Comparison Experiment	122
12.1	Summary of Thesis Argument	125

List of Tables

2.1	Eligibility of Existing Datasets by Criterion	17
3.1	<i>Tex500</i> Dataset Requirements	35
4.1	Range of Possible Similarity Coefficients by Opportunity Incidence . .	47
4.2	Convergence times and stress values from MDS	54
4.3	Stress Value Goodness of Fit Interpretation	55
4.4	SOM Data Processing Result	59
6.1	Statistical Analysis of SOM _G , MDS _G & RF _G Mean Task Time	79
6.2	Experimental Effect Sizes	79
6.3	Observer Preference from SOM _G , MDS _G & RF _G Browsing Environ- ments Comparison Experiment	79
6.4	Statistical Analysis of SOM _G , MDS _G & RF _G Accuracy	80
8.1	Eligibility of Feature Extraction Methods by Criterion	94
8.2	Stress Values for Final Data from Figure 4.4	95
9.1	Summary of SOM _G , SOM _A & SOM _F Underlying Data Organisations	100
9.2	Statistical Analysis of SOM _G , SOM _A & SOM _F Mean Task Time . . .	104
9.3	Experimental Effect Sizes	105
10.1	Comparison of <i>Tex500[100]</i> and sub-sampled <i>Tex500</i> Dendrograms .	110
11.1	Summary of SOM _G & SOM _{A2} Underlying Data Organisations	118
11.2	Statistical Analysis of SOM _G & SOM _{A2} Mean Task Time	122
11.3	Experimental Effect Sizes	122

Chapter 1

Introduction

The overarching aim of this research is to produce an efficient¹ texture browser whose performance can be verified with robust psychophysical experiments. In the context of this work texture is defined as 3D digital representations of real, homogeneous surface textures with variation only due to surface relief, which may be rendered as 2D images of such surfaces using consistent illumination conditions and appearance properties (for further detail see sections 2.3 and 3.2).

1.1 Motivation

Over the past thirty years there has been much activity in texture research in the areas of digital capture, synthesis, segmentation, classification, perceptual dimension analysis, computational features, search and retrieval. A very active area of texture research in recent years concerns the development of Content Based Image Retrieval (CBIR) systems [18, 37]. Such projects have focussed on building systems that allow users to retrieve textures that are similar to a query texture [29]. Although these have enjoyed various degrees of success, the approach is not without its limitations [46, 65].

CBIR: Retrieve images in response to a query image

- Users may not have access to:
 - a query image similar to those they wish to retrieve, or
 - the tools or skills to create one.
- User simply may not know ahead of time what they are looking for

¹With respect to the user's time.

- Repetitive queries can become trapped among small groups of undesirable images

Browsing: Search by a user without an image query

- User browses freely until they recognise the images they want

Applications of texture browsing include (but are not limited to) browsing and selection of:

- floor, wall and ceiling coverings by interior designers,
- building render, roofing and paving finishes by architects,
- synthesised leather grains and wood finishes for vehicle cockpit interiors, and
- wallpapers and textiles by consumers.

Such browsing environments may be designed and populated by product manufacturers, suppliers or retailers, and distributed to architects, designers, purchasers and consumers on digital media (e.g. optical discs or flash drives) or through publication on the internet or via electronic mail.

The motivation of this thesis is to investigate efficient browsing environments that can assist users in quickly browsing a large texture database without being in possession of a query texture or necessarily having previous knowledge about the type of texture one wishes to find. By efficient, we mean the instrument must be accurate and quick to use.

1.2 Mission and Goals

In order to establish a framework for assessing the success of this research project we must establish a set of specific, measurable and outcome-targeted objectives. The mission of this research project is to investigate perceptually relevant browsing environments for large texture databases. As this can have a broad interpretation in terms of intent, it is the purpose of this section to provide a sense of direction and purpose by summarising the project goals.

Literature Survey: Review literature on browsing environments and related texture research with a view to discovering the current state of the art and areas where new contributions could be made.

Large Texture Database: Develop a set of criteria to assess the eligibility of existing texture databases to this project and if suitable database is not

available, develop a large surface texture database that meets our criteria.

Human Judgements: Capture human judgements describing the similarity between members of our texture database and investigate if these can be used to organise the database within browsing environments for the purpose of efficiently navigating texture.

Develop and Evaluate Browsing Environments: Develop a variety of browsing environments for navigating texture and subject these to psychophysical experiments in an attempt to discover which, if any, is superior in terms of minimising mean task time in a texture browsing task.

Investigate Scalable Methods: Investigating methods of performing perceptual grouping experiments with prohibitively large datasets of images and developing and testing methods of augmenting existing datasets with large numbers of previously unknown textures.

1.3 Scope

It would be a large task indeed for this thesis to look at browsing of the whole world of texture. As the research of the Texture Lab has focussed on *surface texture*, i.e. variation only due to surface relief (ignoring albedo and reflectance function), we shall restrict our stimuli database to images of this type. This has the significant benefit of allowing us to render the textures using consistent viewpoint and illumination conditions, variability in either of which are known to effect human perception of texture. By capturing textures using photometric stereo and encoding them as height maps we can also efficiently compute computational features for each texture (see chapter 8).

It is entirely possible that the browsing environments we investigate may be applicable to browsing data of other types but we make no attempt to verify this.

The focus of this research is solely on browsing environments and is not intended to enhance the state of the art in CBIR or any other search-by-query approach.

Part of this thesis evaluates the suitability of using computational features to identify structure in texture databases that can be exploited for navigation. We rely heavily on the work of Emrith [25, ch. 5 & 6], whose techniques we use in a *black box* fashion. No new contribution is offered in the areas of computational features or feature selection.

1.4 Novelties & Contribution

The products from this research project that may be regarded as novel and contributing to the state of the art can be summarised as follows.

Database Development: Development of a database of five hundred surface textures with accompanying perceptual similarity data that can be used for future research projects in the Texture Lab or by external researchers. The dataset was later utilised in a project by the author with Clarke et al [14] *Perceptual Similarity: A Texture Challenge* in 2011.

Rigorous Browser Comparison: Use of several best practice psychophysical experiments and robust results analysis to evaluate the performance of browsing models with a view to identifying whether, with statistical significance, any could be deemed superior in terms of mean task time.

Scalable Dataset Augmentation: We developed and tested a scalable method of augmenting a dataset with a large number of additional textures for use in browsing environments by capturing human judgements on texture similarity from members of a crowdsourcing community.

1.5 Thesis Organisation

Figure 1.1 has been provided to guide the reader through the thesis structure. Strong associations between chapter topics are contained within a dotted outline and the arrows between chapters reflect the chapter dependencies. In chapter 12 a more detailed flowchart (figure 12.1) can be found which additionally provides the reader with a summary of the thesis argument.

Conventionally, chapter 2 provides a survey of the current literature relating to the topics of the thesis while chapters 3 and 4 deal with dataset development, capturing human similarity judgements and preliminary inspection of the dataset for structure.

In chapter 5 we describe the development of three browsing environments that exploit the structure found in chapter 4 for organisation and navigation. These are evaluated in chapter 6 and the SOM_G found to have the best performance. From that point onwards we restrict future browsing environments to use the SOM browsing model for brevity.

Chapters 6, 7 and 8 produce three browsing environments with differing methods of structuring the data. In chapter 9 we evaluate these browsing environments and

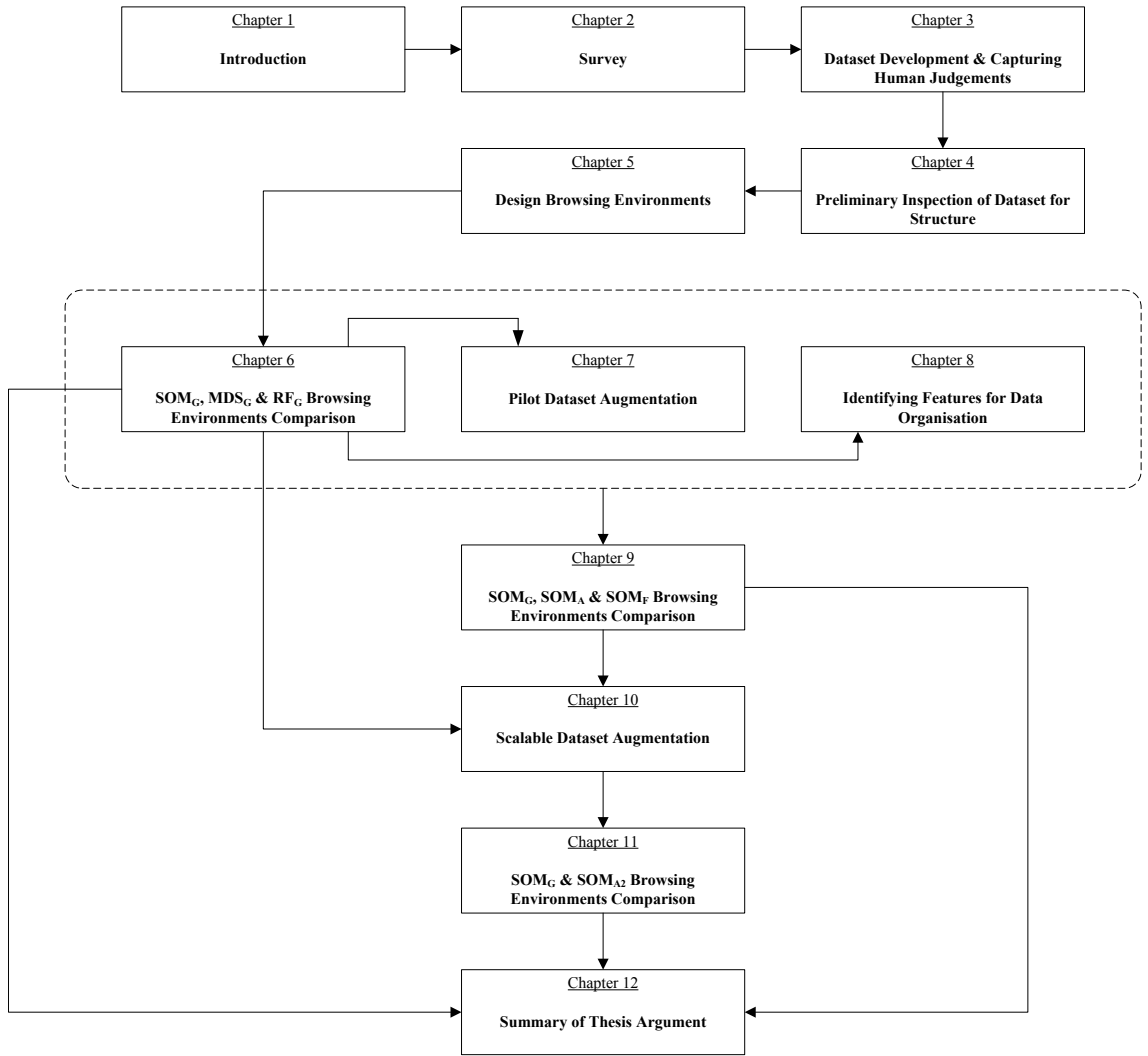


Figure 1.1: Thesis Organisation

find that our pilot augmentation approach (SOM_A) has no significant degradation in performance when compared with SOM_G . However, the feature based organisation (SOM_F) has significantly reduced efficiency compared to the other two.

In chapter 10 we produce a browsing environment based on a scalable version of our augmentation approach using crowdsourcing. A comparison in chapter 11 between this (SOM_{A2}) and the SOM_G browsing environment reveals no significant degradation in performance.

Finally, in chapter 12, we summarise the thesis argument textually and diagrammatically.

Chapter 2

Survey

2.1 Introduction

A study of perceptually relevant browsing environments for large texture databases must draw from the *state of the art* in a number of disciplines in computer science and cognitive psychology. In this survey we provide a comprehensive review of the relevant literature which informs our selection of techniques for capturing human judgements, analysing psychophysical data and mapping computational features to perceptual space. We also evaluate the currently available texture databases, review the limited research in the area of browsing environments, and identify a range of candidate graphical user interface (GUI) components that may suggest suitable browsing models for navigation.

As a large proportion of texture research has been in the area of content-based image retrieval (CBIR) we review and contrast these approaches against perceptually relevant browsing to give a flavour of how our research differs from *search-by-query*. For completeness, a short section on image collection annotation is included and we conclude our survey with a comprehensive treatment of the benefits and risks of using crowdsourcing platforms for research experiments.

2.2 Browsing Environments

Browsing environments offer an alternative to conventional search-by-query, but have received much less attention [11]. In general, a browsing environment seeks to logically and predictably organise the database so that users can find the images that they need. If users are provided with a spatial interface in which content similarity

between images can be intuitively conveyed by their spatial proximity, then such interfaces may help users to benefit more from a given image database [9].

The Collins English Dictionary gives a basic definition of the term browse,

to look through in a casual leisurely manner.

The Penguin Concise Dictionary of Computing offers the definition of a browser as,

a program that is used to view the contents of a large collection of files or other software objects. It will typically present the user with a list of such objects, selecting one of which will open it for inspection.

Combs et al [16] offers the following distinction between *Image Retrieval (IR) Systems* and *Image Browsers*:

1. An *Image Retrieval (IR) System* is an application that returns one or more images given some descriptive information. This information can be in the form of:
 - (a) an image,
 - (b) keywords or phrases, or
 - (c) natural language.
2. An *Image Browser* is an application that allows users to select one or more images from multiple images. This browser has to:
 - (a) be able to display multiple images at one time (possibly reduced resolution versions), and
 - (b) support inspection of original full resolution versions of an image.

For databases with large numbers of images (N), it is not feasible to browse linearly through the images in the database. A desirable characteristic is to let the user navigate through the database in a structured manner [56].

Navigation would normally be considered to be the activity of finding one's way through an environment. In architecture, the term 'wayfinding' is preferred (and used synonymously with the term 'navigation') [6]. From an architectural perspective, Passini [78, p. 154] defines wayfinding as "a person's ability, both cognitive and behavioural, to reach spatial destinations". This conception is based on Downs et al [22] who see wayfinding as composed of four steps:

1. orienting oneself in the environment,
2. choosing the correct route,

3. monitoring this route, and
4. recognising that the destination has been reached.

Rodden [91] recognised, with the popularity and affordability of digital cameras and visual capture devices, that digital image collections would increase in size. Even the personal collections of amateur photographers would need some organisational approach to support finding items at a later date. Her early work asked the question “how do people organise their photographs?” She discovered the importance of browsing over conventional querying in this respect. Further work by the same author and Sinclair [93] asked, “Does organisation by similarity assist image browsing?” They found that arranging a set of thumbnail images according to their similarity does indeed seem to be useful to designers. They also concluded that although labels attached to media may be very helpful to finding particular images, that the process was time consuming and prone to inconsistencies between those carrying out the labelling task.

Chen et al [10] proposed a technique of browsing using similarity pyramids. The similarity pyramid organises large image databases into a three dimensional pyramid structure. Each level of the similarity pyramid contains clusters of similar images organised on a 2D grid. As users move down the pyramid, the clusters become smaller, with the bottom level of the pyramid containing individual images. Users can also pan across a level to see images or clusters that are similar. Like much work in this area, they used a similarity measure based on a distance function incorporating colour, edge and texture features.

Pang [77] observed that colour pickers were convenient means of selecting colours in a range of image editing software but that no similar selection tool existed for texture. He proposed a texture picker for selecting binary textures for designing manga artwork that included both a browsing component for choosing from a range of textures in a palette and a query engine where a user could draw a texture and select to see a range of similar or dissimilar (contrasting) textures. Texture similarity in his application employed Gabor wavelets to quantify the texture characteristics. Multidimensional scaling (MDS) was used to reduce the dimensionality of the feature space and to allow the textures to be projected on the manifold required.

Holmquist et al [42] developed a hierarchical browser for browsing collections of digital images including photographs, scanned document pages, drawings, renderings, etc. Images are grouped into folders which can be brought in and out of focus using a focus and context method called flip zooming, developed by Holmquist [41]. The main image is displayed in the centre of the screen, surrounded by the category containers. The hierarchical containers aspect of the browsing model may be

applicable to our project but here the organisation of the data is based on meta data – assigning a category and sub-category to each container rather than a perceptual similarity model. The method was proposed as a useful alternative to traditional image-browsing and no formal usability testing of the method was offered.

Chiu et al [12] presented the MediaMetro system for browsing document collections in response to the rapid advances in technologies for content creation and wider availability of digital media. MediaMetro employs a city metaphor where the user navigates by flying around a 3D model of a city in a helicopter where the facade of buildings show media storyboards or Video Manga [111] and the roofs display a single representative keyframe of that building’s media content. By zooming and then zeroing in on interesting buildings, the user can select the content to be played. Although designed primarily for video media, this approach benefits from making good use of screen real estate and could be adapted so that buildings are textured with image thumbnails of similar images. Buildings would therefore represent perceptual clusters of images. Problems include occlusion of content by nearer buildings and loss of detail due to scaling. These problems could be addressed by approaches for displaying large images on multi-projector display walls proposed by Jaing et al [45], but not without considerable hardware costs and excluding internet applicability of the solution. No formal measurement of efficiency was offered by the authors.

Martinez et al [66] proposed the use of Galois’ (or concept) Lattices to access databases of images for browsing. A directed acyclic graph is created where the nodes represent a set of descriptions. Sets of images can be produced which share *exactly* the same description and *at least* the same description, for the purpose of navigation. Like many browsing projects this endeavour is largely based on the assignment of linguistic variables to the dataset, in other words meta data or labels. However, this technique could also be used to model other measurable components such as colour histograms, and therefore may be able to model perceptual similarity data. The implementation is a browsing / querying hybrid and further investigation would have to be made to assess the applicability of this approach in pure play browsing and navigation.

FABRIC is a project based at Dundee University that seeks to navigate image sets in meaningful ways to provide design inspiration to people in the fashion industry. Ward et al [116] have been working on mapping image collections into 2 dimensional or 3 dimensional space such as to preserve meaningful inter-image similarities. They make the distinction that the project aim is browsing and navigation, although they conceded that the process must be initiated with a query of some kind. They use content-based image retrieval (CBIR) techniques such as colour histograms to represent dimensions. They have failed to address problems of utilising screen real

estate effectively, and highlight problems in displaying images in low dimensional space, such as one image occluding another or others. They also do not address perceptual relevance of the measures employed.

Strong et al [107] identified the difficulties of browsing through thousands of unorganised photos and proposed an approach that generates a feature vector for each image in the collection which are then used to train a Self Organising Map (SOM). The features they used were the global colour distribution and colour correlogram. Similarity between images was calculated using the Euclidean distance between their feature vectors. For smaller datasets the images are plotted centred at their best matching unit (BMU) co-ordinate. As the collection gets larger, and images begin to overlap, a dynamic collage interface is used to group clusters of images together and produce a representative collage image in their place. The manifold on which the images or collages are projected can be explored by scrolling and zooming. The use of the Self Organising Map (SOM) has also been successful in other similar projects, but concerns about occlusion have often remained unsolved. Quantized SOMs may be a possible solution - where data is fitted to a grid or lattice structure. The use of computational features to derive similarity between images has met with varying degrees of success.

Plant et al [85] review and contextualise existing browsing approaches applicable to image databases. They make the distinction between horizontal browsing (all images are available to the viewer, albeit some may be off-screen), vertical browsing (images are clustered hierarchically), graph-based browsing (global view which can be zoomed in on, such as Pathfinder by Chen et al [9]) and time-based browsing (images clustered by time stamp). Horizontal browsing is achieved by panning, zooming, magnification or scaling and can be applied to any single cluster of the other models. No new work is offered but in their conclusions they highlight a lack of efficiency and scalability testing of existing approaches.

A detailed comparison between the merits of using MDS (Multidimensional Scaling) and PCA (Principal Components Analysis) for the purpose of creating low dimensional information for browsing images was made by Keller et al [51]. They applied these dimensionality reduction approaches to MPEG-7 descriptors and concluded that the performance of PCA was powerful and sufficient for the compression of high dimensional feature spaces to only three dimensions. However, their work was not texture specific and the dataset was defined in computational feature space as opposed to perceptual similarity space.

A Microsoft Research project, PhotoTOC (Photo Table Of Contents) is a system that helps users browse photographs in their own collection of photographs. The

authors, Platt et al [86], attempt to cluster collections into events based on two clustering approaches. One is time-based clustering, where the creation time of the digital image is used to cluster the photographs. Sometime this information is absent or incorrect due to the digital capture device's time being set incorrectly, or if the image is scanned from an analogue source after the event. In this case, content-based clustering is employed, clustering being performed using colour information in the image. The project is focussed entirely on clustering by event, and these clustering approaches are not applicable to our needs.

Rodden et al [92] made an evaluation of a visualisation of image similarity as a tool for image browsing. A set of computational features giving a balance between global image properties and local region based properties provided an image similarity metric. MDS (Multidimensional Scaling) was used to approximate those features into a low dimensional output configuration (2 dimensions) to allow thumbnails of the images to be plotted at their MDS co-ordinates. 48 sets of 80 images were subjected to this treatment, and a corresponding number of random arrangements were generated. 16 subjects were tasked with a search task on each of the MDS / random pairs and ANOVA applied to the result to discover statistical significance of the difference. They found that the visualisation was more efficient than the random case.

An investigation of visual structures for image browsing was made by Torres et al [106]. Their approach was in the context of displaying results from a Content Based Image Retrieval (CBIR) query, and so the browsing structure included the query image along with the results of the query spacially organised with proximity by image ranking. The two structures investigated are concentric rings and a spiral. With the concentric rings visualisation, the results are displayed on concentric rings around the query texture, in reducing order of ranking as the radius of the rings increases. With the spiral approach, the query texture is positioned at the origin of the spiral and the result textures are mapped on the spiral in reverse order of similarity. In both cases images are thumbnail representations of the data and their size reflects the similarity or ranking of the result. As this is not pure browsing and navigation, but rather a means of displaying a query result, there is little to commend the use of the approach in our project.

An article by Fan et al [26] with Microsoft Research Asia on image browsing on mobile devices was examined and found to be concerned with browsing image contents rather than browsing collections of images as the title may have suggested.

Work towards the effective use of limited display space for integrated multimedia navigation was carried out by Walter et al [115]. They proposed using *hyperbolic*

space, a non-Euclidean space with negative curvature, as the manifold onto which a browsing model can be projected. Their approach relies on the recent development of Hyperbolic MDS (HMDS) by Walter et al [114]. The resulting graphical user interface allows the usual mouse interactions for visualisation control (navigation, zooming, image scaling) as well as image set selection.

A *next generation* browsing environment was proposed by Schaefer [98] and tested on a database of about 4500 images. His approach involved mapping thumbnails of the dataset onto the surface of a sphere according to coordinates produced by an angular hue feature. Images were organised into a grid structure to prevent negative browsing effects of images being overlapped and therefore occluded by other images. To ensure scalability, a hierarchical approach was utilised where zooming operations reveal previously hidden images on a deeper level of the underlying browser tree structure. Although this was an image (not texture) database, and the browsing model was based on colour features rather than perceptual similarity, there are some useful aspects of this project which could be of use to our project, in particular, the spherical manifold, grid layout and hierarchical tree structure for scalability.

An article by Wu et al [120], Efficient Retrieval for Browsing Large Image Databases, gave the titular impression that browsing may have been a key theme. The paper actually describes the use of computational features for querying databases, rather than browsing per se. Despite that, the performance evaluations they employed – epsilon queries and nearest-neighbour queries – may have relevance in the event we wish to test our data for perceptual retrievals at some stage in our project.

Lim et al [60] describe two methods of generating layouts for browsing a texture image database. As these both involve deriving similarity information from computational features and not perceptual information, we can discard these for the moment. Also, like many articles reviewed so far in this chapter, they describe the use of MDS and PCA in reducing the dimensionality of highly dimensional feature vectors, again endorsing the suitability of these approaches in the literature, and strengthening these as candidates in our project. The authors project their dataset of textures into 2 dimensional space to inspect mapping for structure and thereby assess the success of the approach. However, no formal testing was done to measure the efficiency of browsing.

Many authors have attempted to bridge the gap between query/search (where we have a query feature vector, image or term to begin with) and browsing (where we may be unsure at the start what we are looking for. Integrated browsing and querying/searching has been proposed by Pecenovic et al [81] and Santini et al [97]. Pecenovic et al argued that a fully interactive *real-time* display of a hierarchically

clustered collection, projected into a two dimensional space can bridge the gap between the user and the system. Rather than supplying the browsing feature as a solution in itself, however, it was merely supplied to allow users to select a query image against which to retrieve similar images.

We failed to discover any specific research in the area of browsing texture databases but we can draw from other researches in browsing environments as to candidate techniques and approaches. Browsing invariably seeks to exploit obvious structure in a dataset so we must examine approaches for the perceptual organisation of texture.

Ordinary image presentation has historically been in the structure of a grid. This can be evidenced by the presence of such grids:

1. in operating system image file explorers,
2. in design/drawing/painting packages,
3. on web pages for thumbnail visualisation, and
4. as a GUI component in most graphical software development environments.

This is usually a good solution for a limited number of items, but it is not scalable when dealing with many more images, in the order of hundreds or thousands.

Alternative methods for image browsing have been proposed by Demontis et al [20]. The five most visualisations from their scholarly article *Experimental Interfaces for Visual Browsing of Large Collections of Images* were:

Cube: images are connected to the vertices of one or more virtual cubes, whose rotation allows pictures in the foreground to continually change.

Snow: similarly to “snow flakes”, images “rain down” from the upper area of the screen and disappear as the bottom is reached.

Snake: images move, with a perspective effect, along a sinuous path reminiscent of a snake.

Volcano: like lava erupted by the crater of a volcano, images are “emitted” at the centre of the screen and slide down along virtual slopes.

Funnel: images appear at the screen edges and disappear in the centre with a perspective effect.

Experiments showed that these methods can reduce the browsing time with respect to traditional solutions. The work was extended by one of the researchers, Porta [87], who suggested a further seven visualisations in his scholarly article *Browsing Large Collections of Images through Unconventional Visualization Techniques*:

Elastic: a grid that can be independently scrolled and scaled horizontally and vertically.

Shot: images are fired like bullets at the upper part of the screen and, with a perspective effect, progressively reach the lower part of the screen.

Spot: images rapidly appear in random positions on the screen (also known in other literature as Rapid-Fire Image Preview [118] or Rapid Serial Visual Presentation (RSVP) [59]).

Cylinder: images are randomly arranged on the lateral surface of a virtual rotating cylinder.

Rotor: images are arranged within four grids on four different planes rotating around a central axis.

Tornado: images move as if they were in a vortex.

Tornado of Planes: applies the Tornado principle to grids of images rather than to single images.

Whilst the interfaces proposed by Demontis et al, or UI (user interface) components as they are more widely recognised, allow many more images to be displayed on the screen than the traditional grid structure, their research did not place any focus on the organisation of the images within the display structure so as to assist in the browsing experience. Instead the images were displayed in random order.

Rapid Serial Visual Presentation (RSVP) is the electronic equivalent of riffling a book in order to assess its content. Evaluation RSVP in video-on-demand browsing systems [59] suggests that electronic RSVP can be applied successfully within that context. Most work on information navigation sheds little light on the world of image browsing as the usual approach to navigating information spaces, such as internet pages and websites, involves assigning or extracting keywords or analysis of textual content. However one group, Wittenburg et al [118], investigated the use of rapid-fire image previews to provide cues to users as to where they wish to navigate. This was also the basis of Porta's *Spot* visualisation technique described above. A modified version of this component may well be a candidate for the hierarchical navigation of texture/image databases.

In their paper on perceptual image similarity experiments, Rogowitz et al [95] compared similarity matrices built from two different psychophysical scaling experiments and two different algorithmic approaches in an attempt to gain insight into how the dimensions human observers use for judging similarity differ from the algorithmic methods. Although their findings in this respect are interesting, the method by

which they visualised the data provides a good candidate user interface component for work on browsing environments. Multidimensional scaling (MDS) was employed to reduce the ND similarity matrix to 2 or 3 dimensions and either displaying them in a 2D canvas or a 3D VRML (virtual reality modelling language) scene.

A survey of browsing models for content based image retrieval by Heesch [39] gives an alternative discussion of browsing environments, including image retrieval, CBIR, human-computer interaction, data visualisation, browsing, networks, clustering and dimensionality reduction. Although it does not examine the specific works we have selected here, it can be regarded as complimentary further reading for those who would benefit from a wider overview.

2.3 Existing Datasets

Our search of the browsing environment literature failed to discover any projects that had specifically produced a database of textures that we could use for this project. We must therefore look to other areas of texture research in order to identify possible candidate texture databases for use in our research. In section 3.2.1, we detail the stimuli specification in full, but for the purposes of identifying suitable candidate databases, the criteria can be summarised as follows.

1. Dataset must be **sufficiently large** to facilitate non-trivial browsing, i.e. provide opportunity to fill several screens with texture images (say circa 500 samples)
2. Dataset should consist of **surface textures** – with variation due only to surface relief
3. Textures should be **homogeneous** – it would be non-obvious from where on any texture a small patch had been sampled
4. As **illumination and viewpoint conditions** are known to affect texture perception, these need to be **constant**
5. Existing **perceptual similarity data** would be desirable

Brodatz (Textures - A Photographic Album for Artists and Designers)

The Brodatz [7] dataset was hitherto considered the *de facto* database for training and testing retrieval models for texture. It consists of one-hundred and twelve 640×640 pixel texture images, but the capture conditions are unspecified. As we

know that changes in illumination conditions can significantly affect the perception of observers and the values of computational features [8], and as we intend to use both human judgements and feature extraction in this thesis, it is unlikely that the Brodatz dataset will satisfy our requirements.

CuRET (Columbia-Utrecht Reflectance and Texture Database)

The CuRET database [17] consists of three specific texture databases for the investigation of the visual appearance of real-world surfaces:

- BRDF (bidirectional reflectance distribution function) database
- BRDF parameter database
- BTF (bidirectional texture function) database

This database has been used for visual appearance, texture analysis and synthesis. However, the stimuli have been captured under varying illumination and viewpoint conditions, as well as containing specular and diffuse surfaces suggesting different reflectance models have been used in rendering. These variables may contribute to bias in human judgements leaving the database unsuitable for our purposes in its full form, but we may be able to select a subset from the database which share the same viewpoint and illumination conditions and reflectance model. However, it is unlikely that a subset satisfying our selection criteria would contain sufficient numbers to be considered large.

MeasTex

MeasTex [69] is a collection of 2D texture images with unknown viewpoint and illumination conditions. It is supplied with a quantitative measurement framework for image texture analysis and synthesis. As viewpoint and illumination conditions affect texture perception we cannot consider MeasTex as a candidate.

OuTex (University of Oulu Texture Database)

Generated to test texture segmentation and classification algorithms, the OuTex database [75] reflects changes in illumination, surface rotation and resolution. Images captured at three illumination positions are available but these are coplanar and cannot be used to recover the surface height map using photometric stereo. However, like the CuRET database, a subset of images could be selected which share the same capture condition, rendering OuTex a possible candidate in all but the dataset size

Dataset	Criteria					
	Sufficiently Large	Constant Viewpoint	Constant Illumination	Surface Textures	Homogeneous Textures	Similarity Data Available
Brodatz	✗	✗	✗	✓	✓	✗
CuRET	✗	✗	✗	✓	✓	✗
MeasTex	✗	✗	✗	✓	✓	✗
OuTex	✓	✗	✗	✓	✓	✗
PhoTex	✗	✓	✓	✓	✓	✗
VisTex	✓	✗	✗	✓	✓	✗

Table 2.1: Eligibility of Existing Datasets by Criterion

criterion. A suitable subset would be significantly fewer than the 320 available, and therefore could not be considered large.

PhoTex (Photometric Texture Database at Texture Lab)

The PhoTex database [84] consists of height maps that allows us to render controlled texture stimuli, but its limitation of representing only one category of texture (namely rough surfaces such as plaster or rock) causes it to fail our criterion that the database should represent a wide range of textures and therefore it cannot be considered a candidate dataset.

VisTex (Vision Texture Lab Database at MIT)

The motivation behind VisTex [113] was to provide a large set of high quality textures for a range of texture processing applications. However, having been captured under inconsistent studio lighting types and viewpoint it would be unsuitable as a candidate in our research.

Table 2.1 summarises that none of the existing datasets satisfies all our selection criteria. Even those which contain subsets of textures which may be considered suitable would not provide sufficient numbers to be regarded as a large dataset. We resolve to generate a new database, the development of which is described in chapter 3.

2.4 Identifying Structure in Texture Databases

Although there are domain specific taxonomies, there are no generally accepted taxonomies for texture [109, 50]. We have established that browsing environments

must exploit obvious structure in a dataset for organisation and navigation. In this section we examine candidate approaches for capturing human judgements and methods of examining that raw data for structure.

2.4.1 Capturing Human Judgements

In order to develop perceptually relevant browsing environments, we must resolve to involve human subjects in the process of acquiring a perceptual description of any dataset we utilise. Human perception of similarity between objects has been successfully recorded by cognitive scientists through psychophysical experiments in the past, but a range of different approaches are available for capturing these judgements. In this section we consider the candidate approaches for designing psychophysical texture experiments.

Pairwise Comparison

Pairwise comparison involves the presentation of two stimuli to an observer who is asked to compare some characteristic and offer their measure of the characteristic. This could be as simple as stating whether the stimuli are similar (yes/no) or to value the strength of the similarity on a scale. This is a popular approach in Content Based Image Retrieval (CBIR) and has been used by Rogowitz *et al.* [95] to investigate the perceptual similarity between each pair of a set of ninety-seven images. The number of possible comparisons is dictated by combination theory as shown in equation 2.1.

$$\begin{aligned} c &= \frac{n(n-1)}{2} & (2.1) \\ &= \frac{97 \times 96}{2} \\ &= 4656 \end{aligned}$$

To reduce the number of comparisons, Rogowitz *et al.* modified the approach and asked observers to compare a query texture with eight other textures from which they were asked to select the four most similar in descending order of similarity. A similar approach was undertaken by Payne *et al.* [80] using the Brodatz database. Issues with scalability discount this approach from consideration for our project.

Perceptual Ordering

Perceptual ordering is also part of Gestalt psychologists' *laws* of perceptual organisation. Here, the interest is in how the human mind orders the perceptual environment with respect to a visual stimulus [64]. Perceptual ordering tends to rely on *a priori* knowledge of the query stimulus by the subjects, and their ability to recall having seen something like it in the dataset before. In Content Based Image Retrieval (CBIR) this human ability has been used to inform and improve the performance of Query-by-Example (QBE) retrieval engines. Although this approach may be useful to us in the ordering of textures having an equal similarity value to a given query or example texture, this is unlikely to prove useful in obtaining a basic perceptual description of our dataset.

Perceptual Grouping

Perceptual grouping was a term coined by Gestalt psychologists to represent the ability of humans to group similar structural elements within images. Gestalt theory also related to grouping with respect to characteristics such as similarity, proximity, continuation, closure and symmetry [64]. Julesz [47] used the theory to investigate how humans segregate homogeneous texture regions within an image. Lowe [64] explained that perceptual grouping refers to the human visual ability to derive groupings or structures from images without any *a priori* knowledge of the image content. Beyond segmentation, some researchers such as Rao *et al.* [89] employed the approach as a technique for grouping together images that are visually similar in an attempt to identify the high level features of texture perception. Well used in the literature, and providing sufficiently high resolution data, this is an acceptable candidate for capturing human judgements in this project. This method was also utilised by the author with Clarke *et al* [13] on *The role of Wallpaper Groups in Perceptual Texture Similarity* in 2011.

2.4.2 Identifying Structure in Psychophysical Data

Typically, perceptual similarity is represented in a similarity matrix denoting the strength of similarity between each pair of stimuli. As similarity space is sparse and high-dimensional, it can be difficult to visualise or assess in any meaningful way. To render the space accessible to analysis we must first reduce and compact the space. Several techniques to facilitate this transformation are available to us, each having particular merits depending on the type of analysis we wish to conduct.

So that we might increase our potential for identifying a range of possible browsing models, we would ideally like to select a collection of contrasting analysis approaches, each of which is capable of reducing and compacting similarity space and is popular and well regarded in the literature.

Cluster Analysis

Cluster analysis facilitates the partitioning of our data into meaningful subgroups regardless of prior knowledge concerning the number of clusters or their composition [31]. When the data in question is acquired from perceptual judgements, cluster analysis provides a quick and reliable method of verifying that the perceptual data is meaningful. In particular it helps discover whether sensible or believable groupings exist within the dataset that can offer insight about its structural composition.

A survey by Jain *et al.* [44] identified two categories of clustering approaches: *hierarchical* and *partitional*. In the pursuit of a taxonomy of texture categories, psychophysicists have largely subjected their data to hierarchical cluster analysis. This involves generating a sequence of data partitions where each sequence corresponds to a particular number of clusters. Depending on whether the process involves merging clusters to produce fewer clusters or splitting clusters produce more clusters, the respective methods are known as *agglomerative* or *divisive* [33]. Agglomerative approaches are more attuned to the way human observers create groups.

Dimensionality Analysis

Visualisation of multivariate data requires a dimension reduction to a two or three dimensional representation [54]. It is important, if this visualisation is to be meaningful, that the distances between points in low-dimensional space correspond to the (dis)similarities between points in the original space [24, p. 573]. When attempting to visualise multivariate data, it has been shown that the most suitable dimensionality reduction algorithm or technique may only be discovered after studying the results of all the others [100]. Here we consider a variety of dimensionality reduction methods.

Psychometric Method was an early method of tracking the correspondence between human and computational rankings of texture and was used by Tamura *et al.* [108] and Amadasun *et al.* [3]. By computing a representative ranking for the texture features being considered, a comparison can be made with the rankings captured from human judgements. The result is an indication of how well a texture feature corresponds to human perception. The technique was more recently adopted

by Abbadeni [1] in testing the performance of autocovariance-based features with respect to human texture perception.

Principal Component Analysis (PCA) extracts the principal components of a feature space by performing variance optimising rotation of the space. It was initially applied to psychophysical data by Rao *et al.* [90] to investigate how much of the total variance of physical texture space was accounted for by a set of twelve perceptual properties. A more recent application saw Payne *et al.* [80] applying PCA to ranking scales drawn by human observers to compare the similarity of regular textures and discover any structure in the ranking scales.

Classification & Regression Trees (CART) is a non-parametric regression technique for selecting variables and their interactions from a large set of variables based on how well the variables model an expected outcome. It has been successfully used by Rao *et al.* [90] to determine whether a prediction could be made as to cluster membership given a series of responses on a sliding scale by which observers described a texture.

Multidimensional Scaling (MDS) is a method presented by Kruskal [57] of reducing the dimensionality of ordinal data. It has been extensively used in an attempt to identify the principal perceptual dimensions of texture by projecting selected low dimensional data in 2D or 3D space. The result is a visualisation of the data which can be inspected for structural information that may indicate some trend or progression. The assumption in this application of MDS is that perceived similarity space can be translated into a form of *psychological space* where the proximity of textures in psychological space approximate to their perceived similarity.

Motivation for the application of MDS to texture perception came after its successful application to colour perception by Shepard [101]. He demonstrated after applying MDS to perceptual colour space that he could organise the perceptual space in only two dimensions, known as the *Colour Wheel*. In texture perception, MDS has become the exploratory technique for bootstrapping the process of theorising about mental representations of texture [38].

Direct Magnitude Estimation (DME) is a standard psychophysical rating procedure that assumes the human mind processes information as magnitudes and that cognitive categorisation is a means of delimiting magnitude information [21]. An example in texture perception is to ask observers to rate a texture for some

characteristic compared to a reference texture with a pre-assigned rating of the characteristic in question.

Self Organising Maps (SOM) are an alternative dimensionality reduction technique where similarity proximity can be preserved in two dimensions [53]. The SOM is also known as the Kohonen map, after the Finnish professor, Teuvo Kohonen, who first described it [53]. There have been numerous applications of the method, from speech to finding patterns of poverty in the world [24]. Korpipaa [55] made a visualisation of information space using the SOM. His multivariate data was a vector of relative percentages of keyword occurrences in web page content. The tool facilitated web site navigation by clicking nodes in the SOM representing the keyword a user was interested in finding content about. An implementation of Kohonen's work is the *SOM Toolkit for Matlab* by Vesanto et al [112].

As multidimensional scaling (MDS) is a popular method that has had much success in discovering structure in multivariate data we regard it as an obvious first choice for preliminary inspection of the dataset for structure. Many of the other dimensionality reduction approaches generate similarly structured output as MDS, with the exception of self organising maps (SOM). We can therefore satisfy the criterion that our collection of approaches should provide significant contrast by selecting MDS, SOM and hierarchical clustering, although this is a largely pragmatic decision.

2.4.3 Image Collection Annotation

The part of this survey seeking to describe approaches for perceptually organising a dataset would not be complete without a treatment of the web-based annotation tools which have grown recently in popularity. These provide a way of building large annotated datasets by relying on collaborative effort of a large population of users. In the case of the Google Image Labeller [32] and Flickr Photo Sharing Service [30] the goal is to improve image search by keyword. A somewhat more sophisticated project, LabelMe [96], seeks to label objects in cluttered scenes, the labels providing information about the object's identity, shape, location, and possibly other attributes such as pose. Their goal is to provide a dynamic dataset that will lead to new research in the areas of object recognition and computer graphics, such as object recognition in context and photo-realistic rendering. Fergus et.al. [27] used the results of image keyword searches in Google Images to form datasets for their research, although problems can arise when polysemes (e.g. "iris" can be iris-flower, iris-eye, Iris-Murdoch) return images unrelated to the intended category.

Bernard et.al. [5] remark that while text and images are separately ambiguous, jointly they tend not to be. They offer that this is because writers of text descriptions of images tend to leave out what is visually obvious (the colour of flowers, etc.) and instead mention properties that are difficult to visually infer (the species of the flower, say). The annotation task is also often a lengthy process and needs to be repeated for images later added to a collection. Annotation also suffers from not providing a predictable organisation of the collection for browsing. When investigating how people organise their personal photograph collections, Schaffalitzky et. al. [99] found that collections were often grouped by scene, whereas Rodden et. al. [94] found chronological ordering was favoured by many.

Other researchers such as Kadobayashi et al [48] and Snavely et al [103] proposed methods for 3D viewpoint-based photo search but as their work was based on collections of images of buildings and scenes where a successful search would return images of a scene from alternative viewpoints, it is not an applicable approach for our dataset of homogeneous or near homogeneous textures.

Gordon [34] investigated the use of subject terms in the cataloguing of images. He claimed that subject access to image collections in the online environment had faired poorly due to difficulties in matching the vocabulary that people use to describe their retrieval needs to the way that collection materials are catalogued by reference librarians. He concludes that developing a rich browsing space of image subject terms is a problem best solved by a thorough, manual analysis of the subject terms.

Given our aim of investigating intuitive and perceptually relevant browsing environments for texture databases it would seem that image collection annotation, which is largely based around labelling images with descriptions or keywords, rather than capturing the strength of relationships between members of a dataset in some perceptually relevant scale, it is unlikely to provide a good basis for organising our dataset for browsing.

2.5 Mapping Computational Features to Perceptual Space

In generating a perceptually relevant feature-based organisation of a dataset it is important that we integrate human judgements in the organisation generated by the system. Known as the training stage, the system *learns* how suitable the computational features are for predicting the perceptual organisation obtained from psychophysical experiments. Used in combination, these features can achieve an

increasingly accurate prediction model.

Little attention has been given to the mapping of similarity in perceptual space to similarity in feature space and the main focus of previous research can be categorised as classification and retrieval. Payne *et al.* [79] used Kendall’s tau to correlate human rankings of Brodatz textures with rankings of a number of different features. Here, no mapping was undertaken but rather the psychophysical data was used as an evaluation of the features.

Long *et al.* [63] presented a neural network that was trained to optimize invariant and perceptual mappings. Tests to assess the performance of the invariant network showed that the invariant network can perform invariant and perceptual mappings accurately and invariant and perceptual mappings improve the performance of texture image retrieval.

Petrou *et al.* [83] used groupings by human observers to compute a measure reflecting the stability of computational features. The measure accounts for the variability in each class while applying a range of features and the method allows for the assignment of weights to each feature representing how well it models each perceptual class.

The recent and most comprehensive work on mapping feature space to perceptual space was by former member of the Texture Lab, Emrith [25]. Although his approach was in pursuit of a perceptual retrieval engine, the output of his system was essentially a similarity matrix, which has the possibility of being translated into a suitable data organisation for use in a browsing environment. Rather than repeating his work, we shall seek to make an evaluation of its applicability to the area of browsing environments.

2.6 Crowdsourcing

Psychophysical experiments are vital to any research where perceptual descriptions of a dataset underpin the analysis. Traditionally, observers for cognitive psychology experiments are recruited from within the undergraduate student population at the university where the research is carried out but it can be difficult to recruit in sufficient numbers where extended trials are required. Kittur *et al.* [52] demonstrated that hundreds of observers can be recruited for highly interactive tasks for marginal costs within a time frame of days or even minutes using micro-task markets such as Amazon’s Mechanical Turk (AMT) [4].

The micro-task market is a system in which small micro-tasks are made available for selection and completion by users for some reward (micro-reward). Micro-tasks can

typically be completed in a few minutes or even seconds while micro-payments may range from a few cents to a dollar or two. In the context of AMT, the jobs are known as human-intelligence tasks (HITs) and the users, some one hundred thousand from over one hundred countries are often known as *Turkers*.

Gordon *et al.* [35] found that untrained *Turkers* evaluating natural language verbalisations of an open knowledge extraction system will generally give ratings that correlate strongly with those of artificial intelligence (AI) researchers. Snow *et al.* [105] found that in AMT natural language annotation tasks, only a small number of non-expert annotations per item are necessary to equal the performance of an expert annotator. These projects showed that many large tasks can be effectively designed and carried out using AMT at a fraction of the usual expense.

When outsourcing a collection of tasks directly to individual workers via public solicitation we must attempt to understand the relationship between financial incentives and performance. Mason *et al.* [67] found when researching the *performance of crowds* that increased financial incentives increased the quantity, but not the quality, of work. The most important factor in work quality was the design of the compensation scheme (e.g. a quota scheme versus a piece rate) even to the extent that better work can be accomplished for less pay. Greater rewards were found to get the work done faster, but not better.

Dekel *et al.* [19] introduced a data cleaning approach for datasets that are labelled by crowds. They estimated the effect an observer has on a classifier by removing their contribution, retraining and measuring the change in the classifier. A significant change may indicate that the observer should be removed from the study. This algorithm benefits from requiring no prior knowledge.

Mason *et al.* concluded that crowdsourcing permits broader and more representative participation than the traditional pool of university students and could become a useful tool for studying questions of interest to behavioural and social scientists as well.

2.7 Discussion

With this survey chapter we have examined the state of the art with reference to browsing environments with a view to discovering current browsing models that may inform this research project. We have also made a thorough search of the human computer interaction (HCI) literature for components that may contribute to alternative browsing models for navigating large image/texture databases.

As we are interested in perceptually relevant browsing we have reviewed techniques for capturing human judgements and analysing psychophysical data. Perceptual grouping has become a popular means of obtaining perceptual descriptions of datasets but we anticipate that modifications to this approach may be required where the target dataset is large. Scalability issues with pairwise comparisons and perceptual ordering may discount these approaches from consideration. We have also discussed the merits of a variety of approaches to analysing psychophysical data and there is little evidence for singling out any particular approach. Instead it would be prudent to use several approaches in order to make comparisons between the individual approaches and to give the greatest potential for the discovery of new browsing models for navigation.

Image collection annotation has grown in recent years with the collaborative nature of the internet. However, our search of the relevant literature has discounted this approach as a candidate for informing the data organisation of our dataset. Annotation is much more applicable to databases of composite images where the aim is to annotate each image with a list of keywords based on objects contained within the image. This facilitates search by keyword or synonym which is highly effective in language driven search engines but it would have little merit in obtaining a perceptual organisation of a dataset based on strength of similarities between member textures.

Most texture research has been in the area of content-based image retrieval (CBIR). While it is sensible to use techniques and approaches previously used by researchers in that area, we also draw attention to the problems associated with CBIR and why we wish to pursue the investigation of browsing environments as a distinct research area.

We have reviewed the range of publicly available texture databases used by texture researchers in the past and discovered that they are largely captured under unknown viewpoint and illumination conditions. This poses problems with human perception of texture and with deriving computational features from surfaces. We also failed to discover a database that was sufficiently large to pose a suitable challenge to perceptual browsing. We shall use this review to inform our decisions in section 3.2 on a way forward for identifying a suitable database to underpin our research.

Crowdsourcing has been used by researchers to recruit hundreds of observers for highly interactive tasks at relatively low cost. It is therefore a candidate resource where there are difficulties recruiting observers from the traditional pool of university students. This is particularly true where we require a large number of trials and may be of help in developing scalable models for capturing human judgements. Given

the issues we highlighted concerning obtaining high quality results, we must develop strategies that link observer compensation to performance.

Most conventional navigation in computing relies on classification of information and the utilisation of wayfinding to move between the classified information. Way points that indicate the type or nature of data to be encountered are important for dividing the process up into manageable activities. There is a distinct gap in research where navigation is directly related to any obvious perceptual structure in the dataset being browsed. Such structure may be useful to users in eliminating large sections of a dataset as irrelevant or in remembering where particular types of textures/images may be found. Preserving a link between proximity in the navigation model and perceptual similarity of data members may be helpful but must be balanced against available screen real estate and quantity of interaction required of observers for navigation.

A more comprehensive collection of works that span the breadth of knowledge in texture analysis can be found in the book *Handbook of Texture Analysis* edited by Mirmehdi *et al.* [70]. For a collection of articles that address the issues that concern feature selection, *Evolving Feature Selection* [61], with foreword by Huan Liu, is of value.

2.8 Conclusions

We have identified that browsing environments seek to exploit obvious structure within a dataset for organisation and navigation. Since we wish our browsing environments to be organised perceptually, we must identify a suitable means of capturing human similarity judgements. Of the candidates considered we discovered the most widely adopted approach to be perceptual grouping. However, we have some concern over the scalability of this approach as all of the browsing and retrieval projects that used the technique did so with relatively small datasets of around one hundred samples. We anticipate that a modification may be required to introduce scalability as our dataset must be considerable larger in order to sufficiently test the efficiency of our browsing environments to the search tasks. In section 3.3 we design and implement a pilot experiment and describe an adaptation that results in a scalability improvement.

This project is focussed on perceptually relevant browsing of large texture databases, and in our survey of the body of scholarly work we identified many papers, which titularly suggest they are concerned in browsing, but on examination were found to use the word browsing to mean examining retrievals from databases based on

image or texture content, commonly referred to as content-based image retrieval or CBIR. As we wish to browse a dataset without beginning with a query (commonly an example texture, or a prototype or description of one), and indeed without knowing ahead of time what the user might be looking for, we have been able to dismiss some of these works in relation to browsing. However, we have included some of these papers in our survey as other components of their work provided insight into other widely adopted techniques, such as dimensionality reduction and hierarchical clustering.

As browsing environments consist of data, a browsing model and a means of displaying the data to users we also included a limited search of research on candidate GUI (Graphical User Interface) components. Many of the recent novel approaches attempt to maximise screen real estate by animating (position, perspective and scale) visual stimuli so as to expose users to as many as possible. However, these were often found to be one dimensional in terms of the data model linking images together (for example Snake / Shot [20, 87]) and may be most suited to modelling the result of a retrieval (in descending order of similarity) than non-query browsing. Others were more suitable for modelling several dimensions and our search revealed three which may be compatible with the approaches of examining raw similarity data above. These are (respectively), projection in 2D or 3D space using VRML (virtual reality modelling language), rapid-fire image previews [118, 87] and the classic grid layout widely used in drawing packages, operating systems or web pages.

In order to design browsing environments that vary significantly in navigation type we must identify approaches for examining raw similarity data that could produce contrasting data organisations for browsing. Among other approaches that we reviewed, we found that MDS (Multidimensional Scaling), Hierarchical Clustering and SOM (Self-Organising Maps) satisfy this criterion, and that these were by far the most widely adopted in recent works. In chapter 4 we use these techniques to make a preliminary analysis of the results from our perceptual grouping environment and describe their integration into browsing environments in chapter 5.

For completeness we made an brief examination of image collection annotation. As this is largely used for labelling image content and associating meta-data for textual based search we considered it unsuitable for gathering perceptual similarity data.

A well known problem with projects collecting data on human perception is finding observers in sufficient numbers. We reviewed a number of projects where researchers had made use of crowdsourcing communities to recruit large numbers of observers to take part in short human intelligence tasks (HITs) using a variety of experimental stimuli, and generally for low levels of compensation. We also surveyed a number of

scholarly articles which had investigated the effectiveness of using crowdsourcing communities for academic research, tackling themes like the compensation-quality trade off. A description of our use of crowdsourcing for capturing similarity judgements is given in chapter 10 and an experiment to measure the effectiveness of the approach is described in chapter 11.

Chapter 3

Dataset Development & Capturing Human Judgements

3.1 Introduction

In our survey, a search of the browsing environment literature (section 2.2) failed to discover any projects specifically for browsing texture. We then extended our search to other areas of texture research in order to identify possible candidate texture databases for use in our project (section 2.3). Broadly, our criteria for identifying suitable databases were:

1. the dataset must be large enough to facilitate non-trivial browsing (circa 500 samples),
2. the dataset should consist of surface textures, and
3. the textures should be homogeneous. Additionally,
4. existing structural data would be desirable.

After considering the Broadatz, CuRET, MeasTex, OuTex, PhoTex and VisTex datasets, none was found to be suitable with respect to our criteria. Our project would require the development of a new dataset and in this chapter we describe the detailed specification and capture of the dataset that shall be referred to in this thesis as *Tex500*, and the design of an experiment to capture perceptual similarity judgements from which we construct a similarity matrix describing the dataset. The similarity matrix is vital to developing browsing environments for navigation of the dataset as we will use a variety of interpretations of this data to form logical organisations of the dataset for use as navigation schema. This data analysis will be described in chapter 4.

Throughout this research project we will use and display the *Tex500* dataset in a variety of ways. These include:

- printing on photographic paper for use in table-top sorting experiment,
- displaying at a variety of scales in browsing environments on computer screens, and
- extracting computational features for automatic generation of similarity matrices.

These, and particularly the last, give rise to a number of exacting requirements that each texture in the dataset must meet. These requirements are described in full, and we illustrate that none of the existing datasets featured in section 2.3 meets our requirements.

Capturing human judgements of similarity can be done in a variety of ways as discussed in section 2.4.1. We have elected to use a perceptual grouping experiment where observers are asked to make groups of textures they perceive to be similar. We discuss problems that arose due to the large size of the dataset and how we overcame these problems by designing a scalable version of the grouping experiment.

3.2 Dataset Development

The aim of this research project is to investigate browsing environments for large texture databases. But how do we define what constitutes a large database? Previous research projects using textures as stimuli have typically utilised small datasets so to simply aim to acquire more textures than that would not necessarily fulfil the objective. To represent a significant improvement over previous datasets it was decided that the dataset must contain more than one hundred and fifty textures but limited to five hundred or fewer in order to facilitate the necessary psychophysical experiments. Given a dataset of that size, we should reasonably expect to discover any of the pitfalls and issues associated with developing browsing environments for much larger collections. We begin our dataset development by discussing the specifications of the stimuli.

3.2.1 Specifications

We decided to keep our stimuli as simple as possible, controlling the environmental conditions to as great an extent as they could be, and to focus on our central issue:

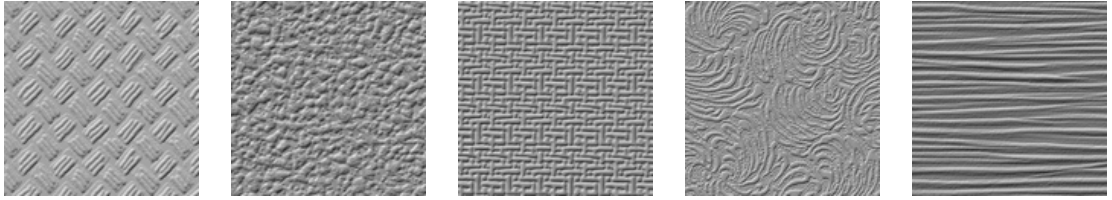


Figure 3.1: Example Textures from *Tex500* Dataset

perceptual organisation of texture. In the simplest terms we want variation only due to surface relief (ignoring albedo and reflectance function). Figure 3.1 shows some examples from the *Tex500* dataset.

Wide Range As large a variety of textures as possible and not limited to any particular application domain

Even Sampling Uniform sampling of the texture space

Surface Textures Surface textures without confusing surface markings, i.e. monochrome, constant albedo, lambertian surfaces

Homogeneous Samples should contain a single homogeneous or near homogeneous texture to avoid problems with observers using different segments of a texture for judging similarity

Constant Scale Samples should be of approximately the same granularity and roughness

Real Surfaces Captured from real surfaces and not synthesised to prevent unstable feature responses

Resolution High enough resolution to capture exact detail without giving rise to storage difficulties

Constant Illumination Images for viewing must be rendered under a single set of illumination conditions

Captured as Height Maps To ensure feature extraction is unbiased by illumination conditions

Believable Surface Texture

The human visual cortex is highly non-linear and optimised for textures that originate from our environment or that can be thought of as originating from our environment. In order to produce consistent results observers need to believe that the stimuli are

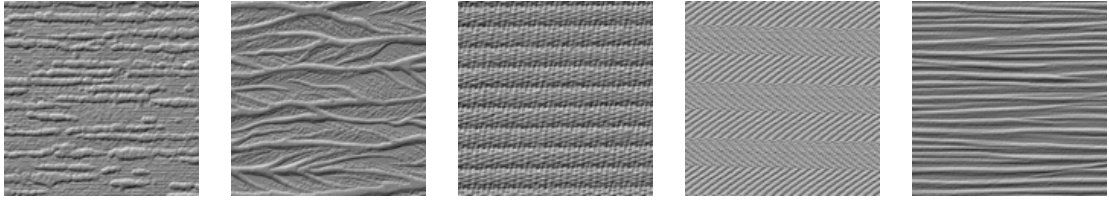


Figure 3.2: Example Directional Textures from *Tex500* Dataset

ecologically valid. We can satisfy this by capturing the digital images from real textured surfaces.

Scale and Roughness

Scale and roughness have been identified in previous studies as important dimensions of texture. We try to eliminate discrimination by observers using these dimensions by limiting variation of scale and roughness between stimuli where at all possible.

Controlled Illumination, Viewpoint & Orientation

In chapter 2 we demonstrated that:

- illumination conditions can bias the outputs of computational features, and
- human perception of a surface can be significantly influenced by illumination, viewpoint orientation.

Prior to the work of Emrith [25] on identifying computational features for texture retrieval, most psychophysical experiments in texture research used datasets comprising texture images with unknown illumination and viewpoint conditions or with too few member textures to be considered a large dataset. We shall replicate his constraint of these properties by capturing our stimuli as digital height maps and rendering these representations of our surfaces under constant illumination and viewpoint conditions.

Emrith also discovered that when presented with directional textures, observers would group textures according to whether the directional quality was largely horizontal or largely vertical. To avoid the risk of observers separating similar textures due to the principal directional orientation we shall rotate all directional textures such that the principal directional component is horizontal as shown in figure 3.2.

Variety of Texture Samples

As we are attempting a general investigation of browsing environments for unbounded texture space (and not an application-specific area of texture) we must follow the work of Rao *et al.* [89, 90] and Emrith [25] in attempting to produce a dataset which includes as much variety of texture types as possible and samples these types as evenly as possible.

Matte Surfaces

It has been shown by Ho [40] that the degree of ‘glossiness’ of a surface can significantly affect human perception of the surface characteristics. Specifically, the presence of gloss can cause the surface to be perceived as being more curved [110]. To avoid bias resulting from the perception of surface properties we shall render our surfaces for experimentation using the simple Lambertian reflectance model [74] which produces a matt surface.

Constant Albedo

Real surfaces are composed of patches that have different light energy absorption capabilities. An area of high energy absorption will reflect less light than one of low energy absorption. These reflectance differences are referred to as surface albedo [71] and variations of albedo across a surface can influence human judgements when making comparisons between stimuli. To avoid introducing such influences we must render the dataset under constant albedo giving a monochrome appearance.

Image Size & Resolution

As we shall be using our images in printed form as well as on screen at full size and thumbnail size we must give consideration to the print sizes and resolutions we expect will be required to produce good quality display of our textures. We also plan to perform computational feature extraction on our texture height maps, giving rise to quality, storage and processing time considerations.

We wanted to ensure that our dataset represented an improvement in resolution over the datasets discussed in chapter 2 which ranged from 256×256 pixels to 384×384 pixels. Anticipating that some processing we may wish to carry out might be more efficient with image resolutions that are $2^n \times 2^n$ pixels, we considered 512×512 pixels and 1024×1024 pixels. We discounted the possibility of using $2048 \times$

No.	Component	Summary
1	Number of samples	Several hundred
2	Surface reflectance	Lambertian, monochrome, constant albedo
3	Homogeneity	Homogeneous or near-homogeneous
4	Realism	Digitally captured real surfaces
5	Resolution & size	1024 pixels ² (8 bit), 4 inches ²
6	Source capture type	Height map
7	Environmental conditions	Constant illumination and viewpoint
8	Scale & roughness	Approximately constant

Table 3.1: *Tex500* Dataset Requirements

2048 pixels as this could not be displayed at 1:1 scaling on the average computer screen. We set the source image size to the highest of those considered suitable, 1024×1024 pixels.

Previous studies discovered that a printed image size of 4×4 inches was sufficient for table top sorting. Given that we hope to have a much larger dataset than used in previous texture perception experiments, increasing this size may make the table top experiment unmanageable due to space constraints. We also did not want to make the printed images smaller than previous projects. We therefore decided to make printed stimuli 4×4 inches giving a printed resolution of 256 dots per inch (DPI).

3.2.2 Acquisition of Stimuli

This subsection describes the sourcing, sampling and preparation of stimuli for experimentation. The stimuli comply with the requirements discussed in subsection 3.2.1 and summarised in table 3.1.

Sourcing the Samples

The author set about the task of collecting and digitising as many textured surfaces as possible. These included embossed, blown vinyl and woven wall coverings, carpets and rugs, window blinds and soft furnishings fabrics, building materials, product packaging and any other moveable item bearing a suitable texture for capture. The sourcing of samples was approached systematically by networking with a wide variety of suppliers in the central belt of Scotland until the law of diminishing returns dictated few additional samples would be discovered by continuing the search. Items were either purchased, hired or borrowed then removed to the lab for digitisation. This was a time consuming and challenging task, but the effort was rewarded with a large, diverse, application non-specific dataset. The final count of surfaces obtained

was five hundred and the dataset was labelled *Tex500* to reflect the type of stimuli (textures) and the size of the dataset.

Digital Capture of Surfaces

Using height maps to model our textures provides two significant benefits:

- texture features derived from height maps are independent of any imaging conditions used to view the surfaces, and
- height maps derived from samples with glossy surfaces or surfaces with variable albedo can be rendered to satisfy requirement no. 2 (table 3.1).

The capture method we selected was R. J. Woodham’s photometric stereo [119]. It assumes:

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

Figure 3.3 illustrates the set up used to capture our surfaces. To recover the surface topology, at least three images are required, taken with illumination at non coplanar angles. By solving three simultaneous equations we can estimate the per-pixel scaled surface normals, from which we can derive the unit surface normals and albedo values. For each surface we captured four images, all at slant angle 60° and tilt angles 0° , 90° , 180° and 270° as shown. An example of the images obtained can be seen in figure 3.4 (for stimulus 067).

Figure 3.5 shows the height map resulting from the photometric stereo integration and the rendered surface under Lambertian conditions at slant angle 45° and tilt angle 135° for stimulus 067. Although the surface albedo for each surface was derived after capture, constant albedo was applied in rendering to remove albedo variation.

Preparation of Samples for Experimentation

The psychophysical grouping experiments used throughout this thesis employ texture samples presented to observers in the form of photographic prints. Each print is scaled to 4×4 inches at a resolution of 1024×1024 pixels using a monochrome HP LaserJet printer. A white space containing the stimulus reference number is included at the bottom of the image so that observers can ensure they make comparisons from

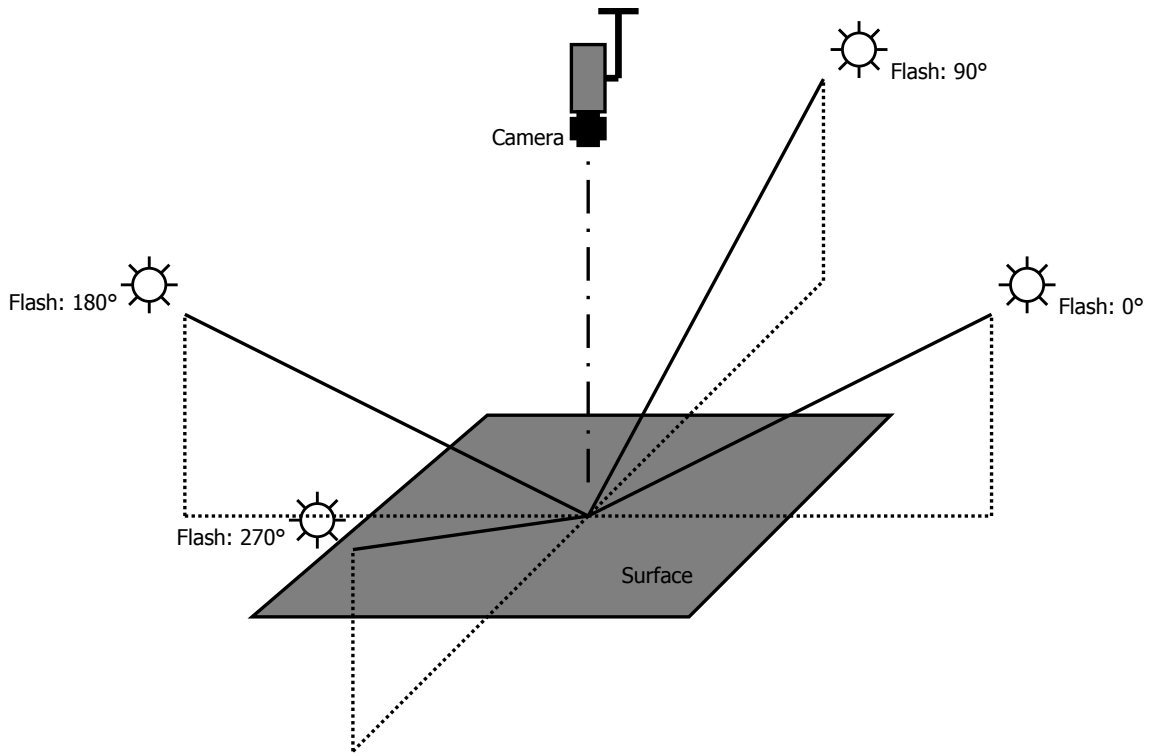


Figure 3.3: Photometric Capture Set Up

the same viewpoint and illumination angles. A barcode is printed on the reverse of each stimulus to facilitate efficient recording of the result using a barcode scanner. An example of the front and reverse of stimulus 067 can be seen in figures 3.6 and 3.7. Both are shown at actual size.

3.3 Experimental Design

3.3.1 Perceptual Grouping Experiment

In chapter 2 we considered the different assessment techniques for capturing human perceptual similarity judgements used by researchers in the past. We were persuaded of the case for using perceptual grouping by its well known advantages, namely:

- perceptual grouping has previously been used in the field of texture perception to determine perceptual dimensions,
- no complex set up is required,
- observers make their judgements in the context of the entire dataset which they can see in full at all times,
- the user does not need to remember previous judgements such as would be

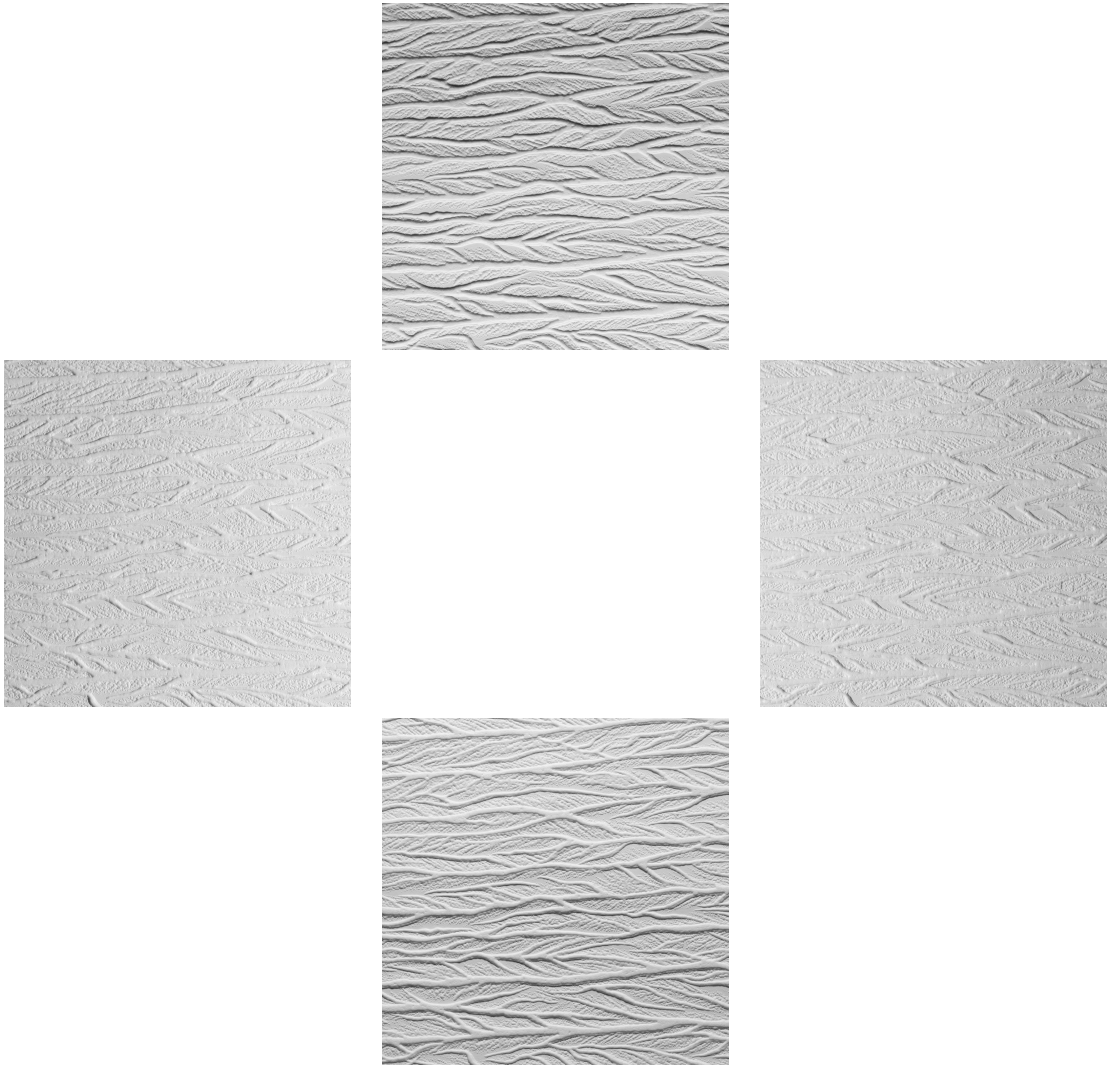
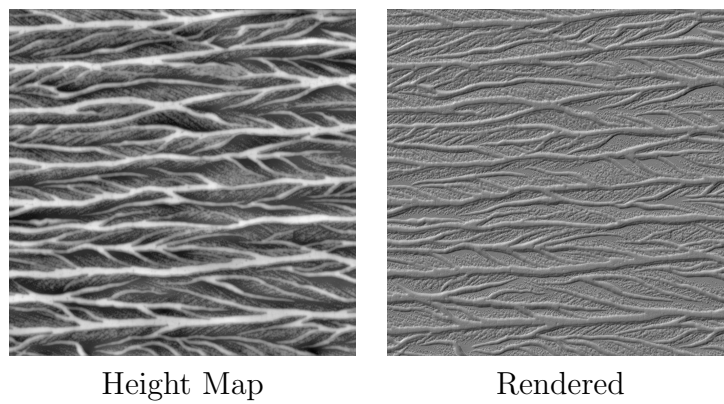


Figure 3.4: Photometric stereo images of stimulus 067 illuminated at constant slant angle 60° and tilt angles 0° (left), 90° (bottom), 180° (right), 270° (top)



Height Map

Rendered

Figure 3.5: Height Map and Rendered Surface (Lambertian Slant 45° and Tilt 135°) for Stimulus 067

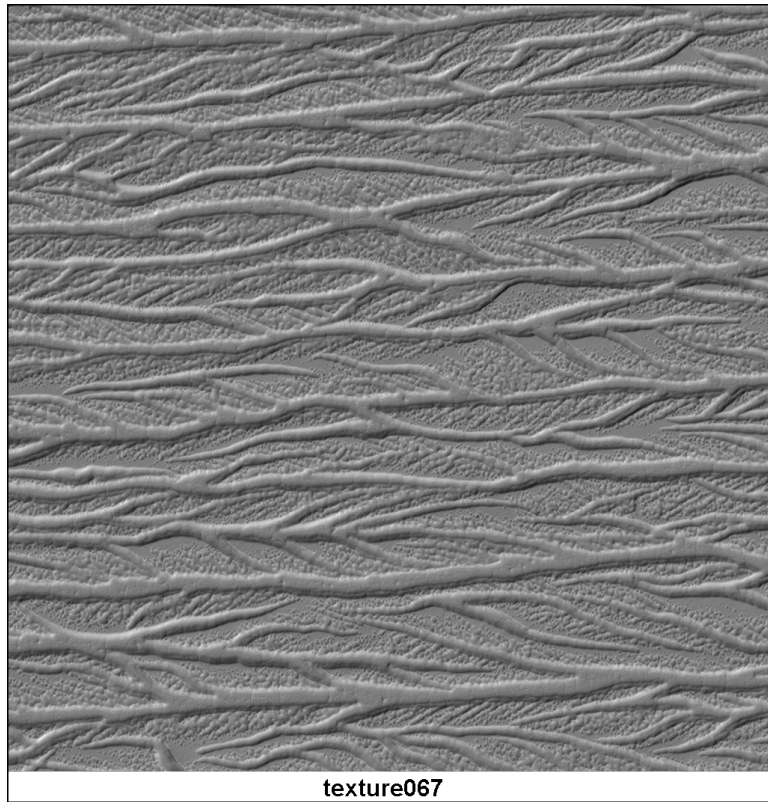


Figure 3.6: Printed Stimulus 067 (Front)

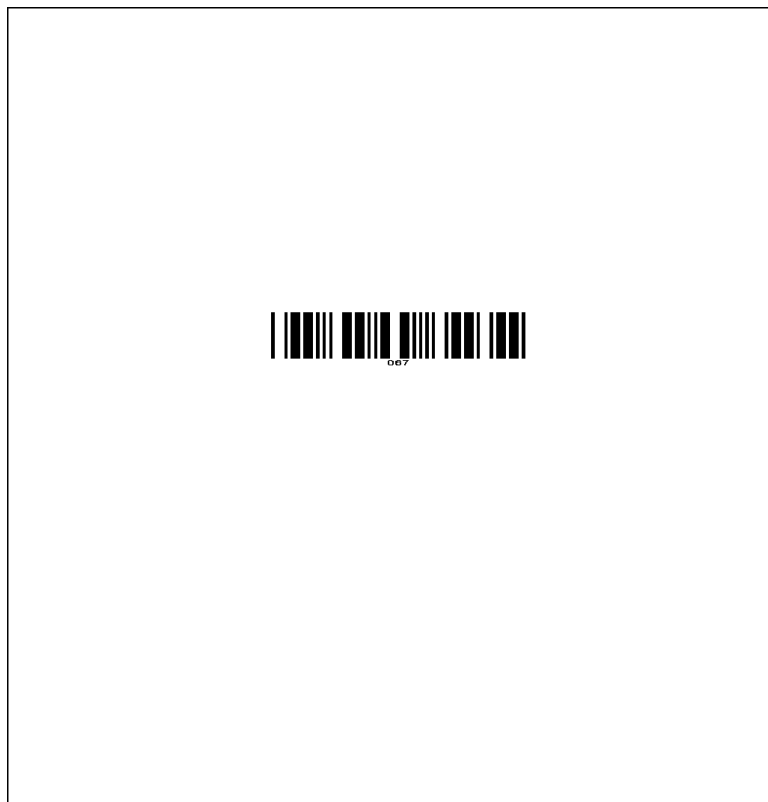


Figure 3.7: Printed Stimulus 067 (Reverse)



Figure 3.8: Grouping Experiment Setting

necessary with pairwise comparisons,

- grouping is a time efficient method.

Implementation

In his thesis [25], Emrith, adopting the method of Rao et al [89], reported that a grouping task on a dataset of one hundred and twenty textures could be carried out by observers in 30–40 minutes. As our dataset is considerably larger, and the number of possible comparisons grows exponentially with respect to number of comparators in the set, we must anticipate that our experiment should take considerably longer. We therefore set about conducting a pilot experiment with six observers who we recruited for their sympathy to the work of the Texture Lab. This group consisted of members or former members of the Texture Lab who could be relied upon to maintain their concentration and effort over a relatively long experimental session.

Emrith was able to use a standard office sized desk for his grouping experiment which, again due to the increased dataset size would not be suitable for our grouping experiment. We commissioned a customised sorting surface made from two sheets of medium-density fibreboard (MDF) each measuring $2400\text{mm} \times 1200\text{mm}$ which were joined along the short edge by a piano hinge to create a single flat sorting surface measuring $4800\text{mm} \times 1200\text{mm}$. This was placed on top of several classroom desks to support it at a comfortable height for observers to work. Figure 3.8 shows an observer taking part in the pilot grouping experiment at the sorting surface described.

Instructions

All observers were asked to complete a standard experiment consent form before being issued with the following instructions for completing the experiment. Observers were compensated at the rate of 5 GBP for every completed half hour of their time by way of Amazon gift vouchers.

1. Researcher presents (next) fifty randomised textures to observer for grouping.
2. Make groups that you perceive to be similar and then stop.
 - (a) There is no restriction on the number of textures you place in any group.
 - (b) Do not group singletons or outliers together.
 - (c) The working surface is not large enough for all the textures so you will need to overlap group members.
 - (d) You may split or merge groups, or move textures between groups at any time.
3. If more textures need to be presented, return to (1), otherwise go to (4).
4. There are no more textures to present. Make sure you are happy with your groups.

The groups made by each observer were recorded in a similarity matrix, normalised to the number of observers to give values in the range 0–1. Equations 3.1, 3.2 and 3.3 show the properties of a similarity matrix where s_{ij} is the similarity coefficient of textures t_i and t_j .

$$0 \leq s_{ij} \leq 1 \tag{3.1}$$

$$s_{ii} = 1 \tag{3.2}$$

$$s_{ij} = s_{ji} \tag{3.3}$$

Figure 3.9 shows an excerpt from the similarity matrix featuring the similarity coefficients of the first nine textures of the dataset. Notice that a texture always has a similarity relationship with itself of 1 ($\frac{6}{6}$).

6/6			2/6					
	6/6							
		6/6		1/6			1/6	
2/6			6/6			1/6		
		1/6		6/6			4/6	
					6/6			
			1/6			6/6		
		1/6		4/6			6/6	
								6/6

Figure 3.9: Pilot Experiment Similarity Matrix (Excerpt)

Problems

The observers who took part in the pilot experiment took between $2\frac{1}{2}$ and 4 hours to complete the task. This long session time gave rise to the following problems and concerns:

Mental Fatigue The extended session time was too long for the observer to remain focussed on the task

Physical Fatigue Observers reported tiredness from having to walk back and forth, and handling the textures over an extended period

Quality Risk If observers lose concentration they may be less willing or able to discriminate effectively

Recruitment It was anticipated that it would be difficult to find observers willing to take part in such a long experiment

Ethics The experiment simply did not represent a best practice ethical means of capturing human judgements

We reached agreement that subjecting observers to such an arduous experiment was not a viable option and to avoid the problems and concerns described above we must find an alternative approach. We present this approach in subsection 3.3.2.

3.3.2 Scalable Grouping Experiment

All of the problems that emerged during the pilot experiment could be eliminated by finding an approach that would be less time consuming for an observer to complete. If we are to retain the grouping experiment as our means of capturing human judgements then this can only be achieved by reducing the number of textures we ask each observer to sort. Had we not already conducted the pilot experiment then

we would not be in possession of any similarity data and we may have taken the view that experiments should be conducted using random subsets of the *Tex500* dataset, of a size that we know to be possible for an average observer to complete in say one hour.

However, we are not so disadvantaged as we do have the similarity matrix from the pilot experiment, even if it does not have the resolution that we may have liked. By using hierarchical analysis of this data (please see section 4.2 in the next chapter for full details), we can partition the dataset into three clusters, and ask future observers to sort only one of these clusters in a grouping experiment. We must also find a way of aggregating the data from these experiments into the original pilot similarity matrix.

Implementation

Essentially the task is exactly the same as the pilot grouping experiment with the exception that observers will each work with around one third of the textures. This allows us to dispense with the necessity to present observers with 50 textures at a time. Instead, all the textures for sorting can be randomly placed on the sorting table prior to the observer starting to produce their groups.

Sample Size

When we partitioned the dataset using the pilot similarity data we produced three clusters of roughly equal size (153, 171 and 176 textures). Most observers were able to complete a sorting experiment in under an hour and only a small number of observers took slightly longer. By working with these smaller subsets of the dataset we were able to eliminate all the problems experienced in the pilot experiment, while retaining the valuable data from the pilot experiments. Each of the subsets was presented to eight different observers for sorting, in addition to the six observers who took part in the pilot experiment.

Aggregation of Data

To aggregate the new experimental data with the similarity data already collected in the pilot experiment then we must consider two aspects of the relationship between each texture and all other textures in the dataset:

Occurrence When an observer groups a texture with another texture, and

14	14	6	6	14	6	6	14	6
14	14	6	6	14	6	6	14	6
6	6	14	14	6	14	14	6	6
6	6	14	14	6	14	14	6	6
14	14	6	6	14	6	6	14	6
6	6	14	14	6	14	14	6	6
6	6	14	14	6	14	14	6	6
14	14	6	6	14	6	6	14	6
6	6	6	6	6	6	6	6	14

Figure 3.10: Opportunity Matrix (Excerpt)

14/14			2/6					
	14/14							
		14/14		1/6			1/6	
2/6			14/14		1/14	2/14		
		1/6		14/14			6/14	
			1/14		14/14	2/14		
			2/14		2/14	14/14		
		1/6		6/14			14/14	
								14/14

Figure 3.11: Augmented Similarity Matrix (Excerpt)

Opportunity When an observer is given the opportunity of grouping a texture with another texture (in other words they are presented together in the same experiment).

In the pilot experiment, the opportunity value for all texture pairs is six. This is because all observer had the opportunity of pairing any texture in the set and there were six observers. But in the scalable grouping experiment, this was not the case. Observers only had the opportunity of pairing textures with the ones that appeared in their pre-determined subset. Figure 3.10 shows an excerpt of the final opportunity matrix. There are two possible values for each texture pair. The value six represents those textures where they only had the opportunity of being paired in the pilot experiment, while the value fourteen also means they were available for observers to pair in the scalable grouping experiment.

By dividing the occurrence matrix by the opportunity matrix we obtain a new normalised similarity matrix in the range 0–1 for all of the observations. An excerpt of the final similarity matrix is shown in figure 3.11. To assist readability, we have continued to show the fractional values rather than a decimal approximation. Note the continued compliance with the properties described in equations 3.1, 3.2 and 3.3.

3.4 Conclusions

In this chapter we described an experiment designed to capture human judgements of similarity by using a perceptual grouping experiment. We described the problems that emerged as a result of using this approach on a large dataset, and highlighted the compromise between observer fatigue and precision. We proposed a new approach that allowed us to experiment with smaller subsets of our dataset and to aggregate the similarity matrix obtained from the pilot experiment with the data obtained in the scalable experiments.

In chapter 4 we shall inspect the similarity data with a view to discovering any obvious structure that might allow us to logically organise the dataset for display in browsing environments.

Chapter 4

Preliminary Analysis of Results

4.1 Introduction

The goal of this chapter is to validate the results from the psychophysical grouping experiments in chapter 3 (section 3.3), identifying that meaningful structure emerged from the data. Should we find some easily perceived structure we can employ it to:

- investigate the validity of the grouping experiment,
- suggest navigation methods that can exploit these structures, and
- determine how best to represent structure and images in browsing environments.

We will source the analysis approaches from those commonly used in the literature for examining multivariate data. Candidate approaches were discussed in full in section 2.4.2 of chapter 2. We shall not be carrying out any additional psychophysical experiments in our effort to evaluate these visualisations and there is no ground truth that we can measure perceived structure against. However, we shall be carrying out experiments to compare browsing environments that take advantage of the products of this chapter in chapter 6.

The result of the experiments described in chapter 3 is the set of similarity coefficients between each pair of textures in the *Tex500* dataset, represented by a 500 dimensional similarity matrix. By analysing the quantisation of the similarity matrix we can discover the range of possible similarity coefficients by opportunity incidence. We can see in table 4.1 that pairings with an opportunity incidence of 6 (textures appearing in different subsets in the scalable grouping experiment) had a range of values from $\frac{0}{6}$ to $\frac{3}{6}$. This tells us that all texture pairs with a value of $\frac{4}{6}$ and over were later placed in the same subset, their opportunity incidence incrementing to 14 by the end of the experiment.

Opportunity Incidence	Possible Similarity Coefficients
6	0/6, 1/6, 2/6, 3/6
14	0/14, 1/14, 2/14, ..., 12/14, 13/14, 14/14

Table 4.1: Range of Possible Similarity Coefficients by Opportunity Incidence

The similarity matrix is a fairly raw representation of the perceptual data but it still has its uses. For example, we might want to compare the similarity coefficient of two known textures t_i and t_j to discover how alike, in the context of the variability in the *Tex500* dataset, observers considered them to be. This can be done by simply looking up the s_{ij} value. A more sophisticated use of the similarity matrix would be to find the set R , in descending order of similarity, of non-dissimilar textures to a given query texture t_i . Commonly known as a perceptual retrieval, the result set can be seen in equation 4.1.

$$R_i = \{t_j : s_{ij} > 0\} \text{ ordered by } s_{ij} \text{ descending} \quad (4.1)$$

The similarity space itself, being sparse and high dimensional, cannot provide us with any obvious way of organising or navigating the dataset. As organisation and navigability is essential to the use of our perceptual data in browsing environments, we must perform further analysis to translate the data into a useful format. The criteria for selecting suitable approaches are that they must:

1. be capable of reducing and compacting similarity space,
2. be recognised in literature for visualising/organising multivariate data, and
3. as a collection, provide contrasting organisation schema (no two similar manifolds).

A collection of approaches drawn from section 2.4.2 that satisfy all four of the above criteria are as follows. In the remainder of this chapter we describe the use of these approaches to identify structure in our dataset.

Hierarchical Cluster Analysis A linkage algorithm is used to generate a hierarchical cluster tree representing the similarity matrix

Dimensional Analysis Multidimensional scaling (MDS) is used to project our high dimensional data in low dimensional space

Neural Network Analysis The self-organising map approach is used to populate a grid of neurons with our data members

4.2 Hierarchical Analysis

Hierarchical clustering groups data across a range of scales by creating a hierarchical cluster tree or *dendrogram*. The resulting tree is not a single set of clusters, but instead a multilevel hierarchy where clusters at a lower level are joined to form larger clusters at a higher level, and so on until at the top, or root, we have a single cluster representing the whole dataset.

4.2.1 Dendrogram

A dendrogram allows us to partition a dataset into clusters of a given number or dissimilarity level and thereby to inspect the data for any evident structure. Dendrograms are a crisp approach to clustering, meaning that at any level of clustering, a texture can be a member of only one cluster. We use the most straightforward method of tree construction, UPGMA (Unweighted Pair Group Method with Arithmetic Mean) [104, pp. 230-234] to produce a dendrogram as shown in figure 4.1. To facilitate inspection, figure 4.2 shows a scaled view of the rightmost fifty textures from the dendrogram (10% of the dataset).

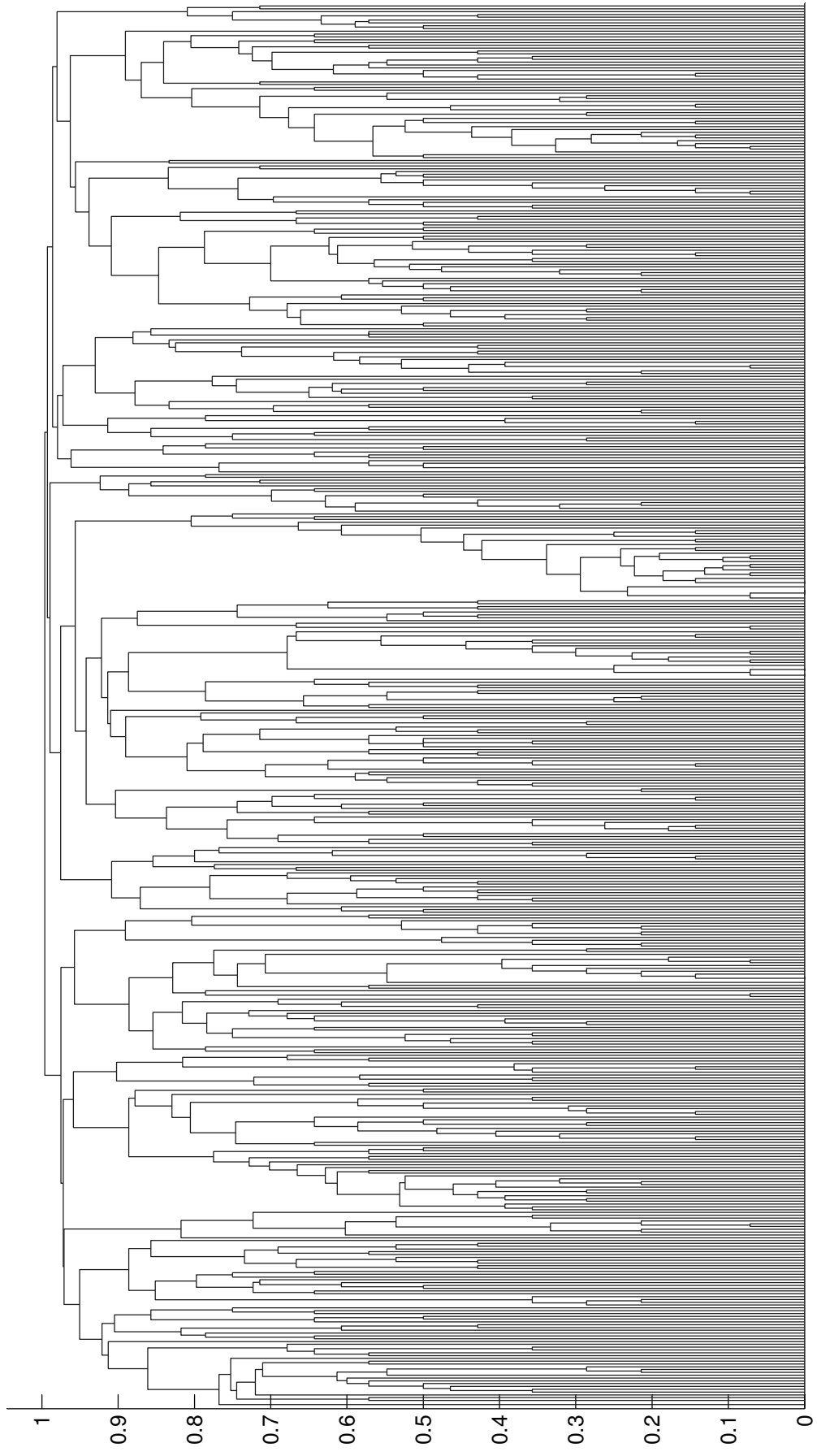


Figure 4.1: Dendrogram

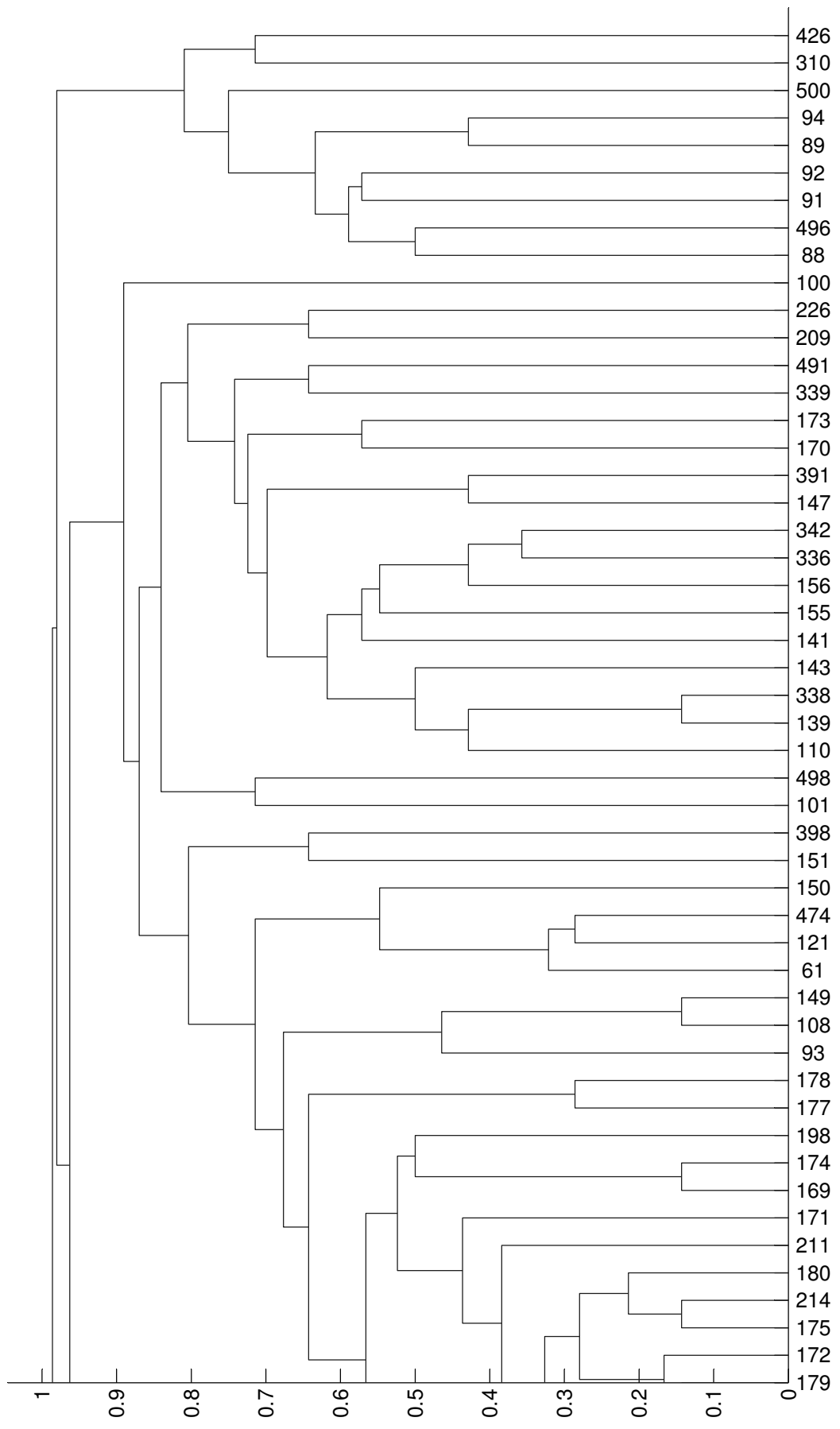


Figure 4.2: Partial Dendrogram (rightmost 50 textures from Figure 4.1)

This affords us the opportunity to qualitatively appraise the result of our psychophysical grouping experiment. Visual inspection by the author of the partial dendrogram discovered an obvious cluster containing the rightmost nine textures at a dissimilarity level of 0.8. This cluster is detailed in figure 4.3 and the similarity between textures is obvious to the author. We were able to verify by cluster analysis, and inspection of the clusters for similarity between cluster members, that the psychophysical experiment described in chapter 3 produced a good description of the *Tex500* dataset with respect to perceptual similarity.

Despite the descriptive and pictorial nature of the hierarchical cluster tree, it does not afford us any particularly obvious means of navigating the dataset. If we attach a thumbnail of each texture to the leaf nodes of the tree, which is essentially what those nodes represent, then we'd find it difficult to make use of the instrument to browse the dataset. To the knowledge of the author, the dendrogram in its raw state has not been used for direct navigation of a dataset.

However, if we recursively partition the dataset into smaller and smaller clusters, we can use these various levels of clustering to refine our search through the dataset towards very small and consistent clusters of similar textures. As the clusters at any level can be joined to form the previous cluster in the hierarchy it is simple to provide a reverse navigation to generalise our navigation at any point. We discuss the application of hierarchical clustering in full in section 5.2.

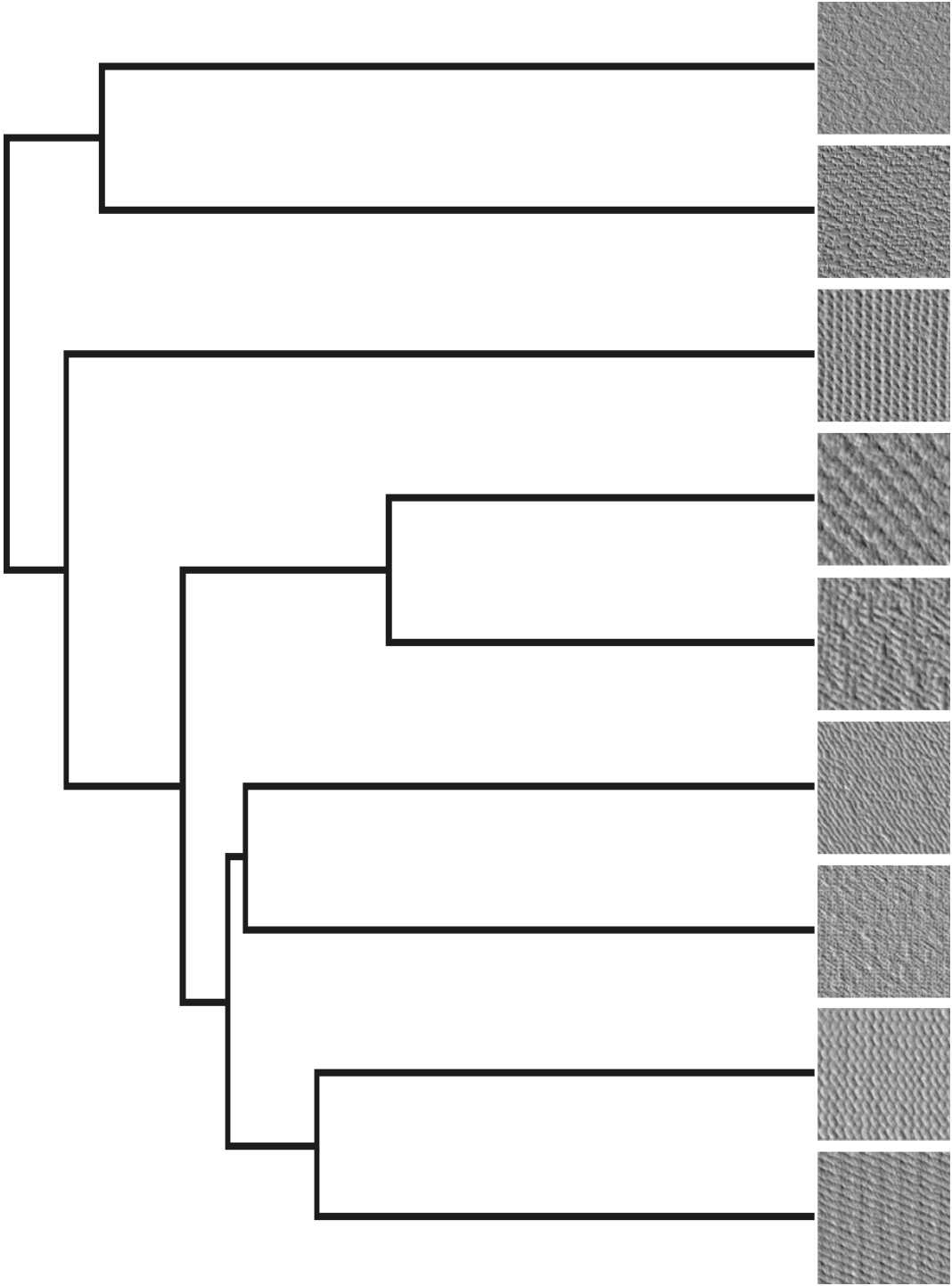


Figure 4.3: Partial Dendrogram with Images (rightmost 9 textures from Figure 4.2)

4.3 Dimensional Analysis

Although the data obtained from our experiments is organised in a similarity matrix, it is distance-like [117], rather than numerically metric in the way we would expect, say, of data modelled in n -dimensional Euclidean space. Nonmetric multivariate data such as ours can be reduced in dimensionality and transformed into genuine Euclidean distances using multidimensional scaling (MDS) [57].

MDS operates on dissimilarity data so we must first convert our similarity data using equation 4.2. We perform non-metric multidimensional scaling on our d_{pilot} (pilot experimental data) and $d_{\text{pilot+scalable}}$ (pilot and scalable experimental data) using Kruskal’s classic algorithm [57] which has been long regarded as the *de facto* approach. It returns a N -dimensional configuration for each member of the dataset. For the purposes of our analysis we generated 1–8 dimensions for each dissimilarity matrix. We also generated a test dissimilarity matrix of 4 dimension metric random data to help evaluate the method. The processing time and stress values for each result generated is shown in table 4.2.

$$d = 1 - s \tag{4.2}$$

4.3.1 Stress

Stress is the term coined by Kruskal [57] to denote the loss function used to minimise non-metric MDS models. Stress is defined in equation 4.3. 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. These terms are defined in equations 4.4 and 4.5.

$$\text{Stress} = \sqrt{\frac{S^*}{T^*}} \tag{4.3}$$

$$S^* = \sum_{r,s} (d_{rs} - \hat{d}_{rs})^2 \tag{4.4}$$

$$T^* = \sum_{r,s} d_{rs}^2 \tag{4.5}$$

\hat{d}_{rs} represents the dissimilarity values defined on an $N \times N$ dissimilarity matrix such that the mapping is always monotonic where as d_{rs} represents the distances computed

Data	Source	Dims	Time (s)	Stress
Experimental	d_{pilot}	..1	21	0.40
		..2	86	0.26
		..3	116	0.19
		..4	292	0.15
		..5	584	0.12
		..6	718	0.10
		..7	837	0.08
		..8	1274	0.07
	$d_{\text{pilot+scalable}}$..1	13	0.20
		..2	182	0.16
		..3	275	0.13
		..4	454	0.11
		..5	634	0.09
		..6	716	0.07
		..7	896	0.07
		..8	1310	0.06
Generated	d_{random}	..1	22	0.65
		..2	324	0.44
		..3	415	0.29
		..4	433	0.16
		..5	441	0.00
		..6	685	0.00
		..7	1013	0.00
		..8	1410	0.00

Table 4.2: Convergence times and stress values from MDS

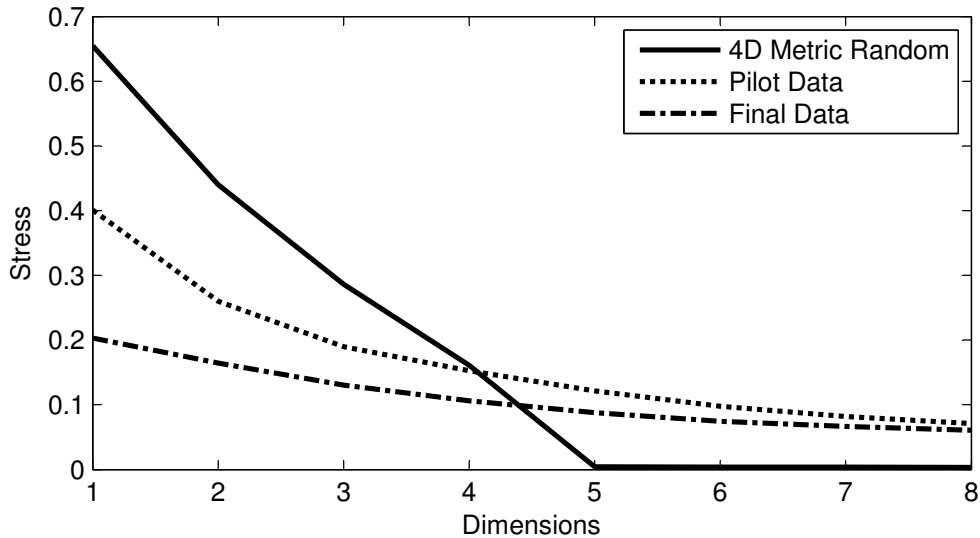


Figure 4.4: Stress Plot

Stress Value	Goodness Measure
Above 0.200	Poor
0.100	Fair
0.050	Good
0.025	Excellent
0.000	Perfect

Table 4.3: Stress Value Goodness of Fit Interpretation

from points in the spatial configuration. Since its conception by Kruskal, stress has been widely used as a measure for the goodness of fit of a chosen configuration [68].

Figure 4.4 shows a plot of the stress values for each of the three dissimilarity matrices against the dimensions extracted. To place these stress values in some context, table 4.3 shows the goodness of fit interpretations for MDS stress values. Although there was no identifiable ‘elbow’ in the final experimental data stress plot, it can be described as approaching ‘fair’ at 3 dimensions.

Stress can also be interpreted as a measure for how much variability in the dataset is modelled at a particular number of dimensions. As we can see from the generated 4-dimensional random metric data, the addition of a 5th dimension when performing multidimensional scaling contributed nothing further in terms of variability described. This was to be expected as we already knew ahead of processing that there were a maximum of 4 dimensions to be found.

Our stress result for 3 dimensions on our $d_{\text{pilot+scalable}}$ dissimilarity matrix (0.11) tells us this may be suitable for use in projecting the dataset in 3 dimensional space. We can also determine that 3 dimensions describe 89.4% of the variability in the data

and that the addition of a fourth dimension would only contribute a further 1.8% to the variability described. We describe the use of this approach in constructing a data organisation for a browsing environment in section 5.3.

4.3.2 Meaningful Dimensions

By projecting our 3-dimensional data onto three single 2-dimensional planes we might expect to discover some significant trend in how the textures change as we move in particular directions across these planes. Various researchers have in the past been able to detect some tenuous trends by using this subjective analysis but it seems likely to hold only for the most under sampled texture spaces or perhaps for severely constrained datasets with a focus on a single class of textures that vary over few parameters. The author’s findings for the *Tex500* dataset was that we could clearly observe a relationship between proximity in low dimensional space and similarity between textures but that no obvious trends or ‘dimensions’ emerged as we moved in any direction through low-dimensional space.

4.4 Neural Network Analysis

A well documented Artificial Neural Network approach to data analysis is the Self Organising Map (also known as the Self-Organising Feature Map or Kohonen map after its inventor, Teuvo Kohonen [53]). It is a vector quantisation method consisting of neurons organised on a regular low-dimensional grid. Each neuron is represented by a d -dimensional weight vector $m = [m_1, \dots, m_d]$ where d is equal to the dimension of the input vectors (in our case five hundred). The SOM training algorithm is similar to vector quantisation algorithms like *k-means* [36]. In contrast, in addition to the best matching weight vector, its topological neighbours on the map are updated and the area around the best matching vector stretched towards the training sample (see figure 4.5). Ultimately, neurons become ordered with neighbouring neurons sharing the same weight vector.

During each step in the training phase one sample input vector x is chosen from the input data set. The neuron with weight vector closest to the sample input vector is called the best matching unit (BMU), denoted by c in equation 4.6 where $\|\cdot\|$ is the distance measure (we use the typical Euclidian distance).

$$\|x - m_c\| = \min_i \{\|x - m_i\|\} \quad (4.6)$$

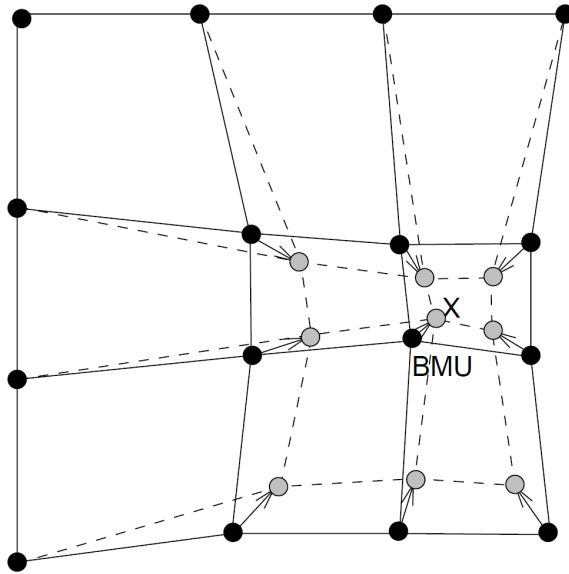


Figure 4.5: Updating the Best Matching Unit (BMU)

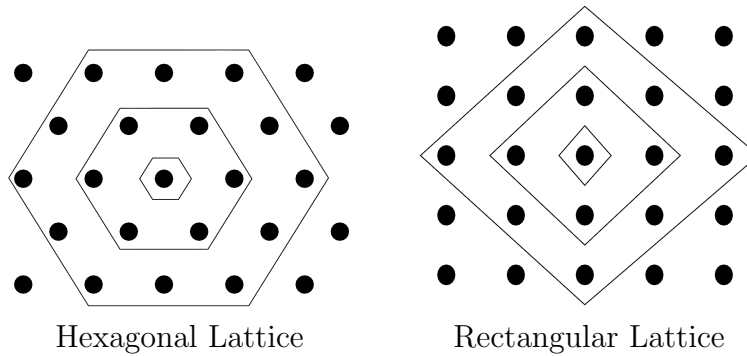


Figure 4.6: Local Lattice Structures

The most appealing feature of the SOM with respect to our aim of data organisation for navigation is that the neurons are connected to adjacent neurons by a neighbourhood relation which dictates the map's topology. Not only are the data assigned to a neuron based on some common strength of neural response to their input vector, but neurons are positioned across the grid relative to their weight vector. The benefit here is that we can envisage the possibility of navigating neurons then examining their contents (in our case textures).

A well received implementation of the SOM method is the SOM Toolbox for Matlab by Vesanto *et al.* [112]. Their implementation allows for two possible local lattice structures as shown in figure 4.6 and three possible map shapes illustrated in figure 4.7.

In terms of topology, we can easily select an arrangement which has become familiar to us through the use of any file browsing interface where thumbnails are displayed on a computer screen. We are almost always presented with a rectangular lattice

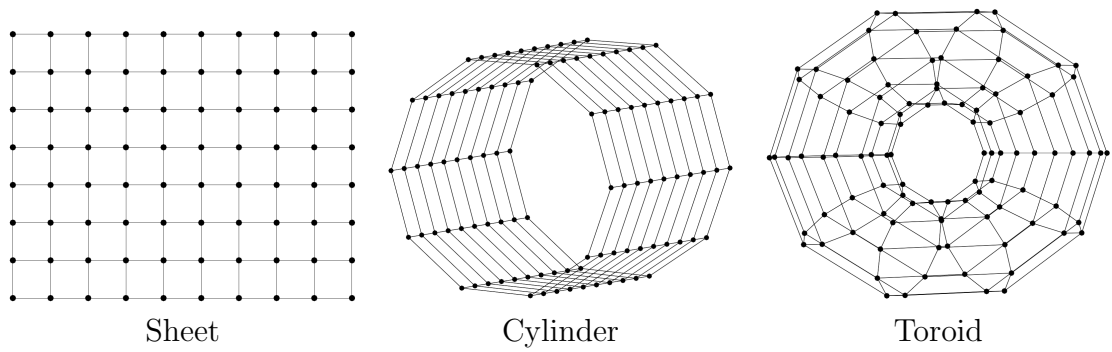


Figure 4.7: Map Shapes

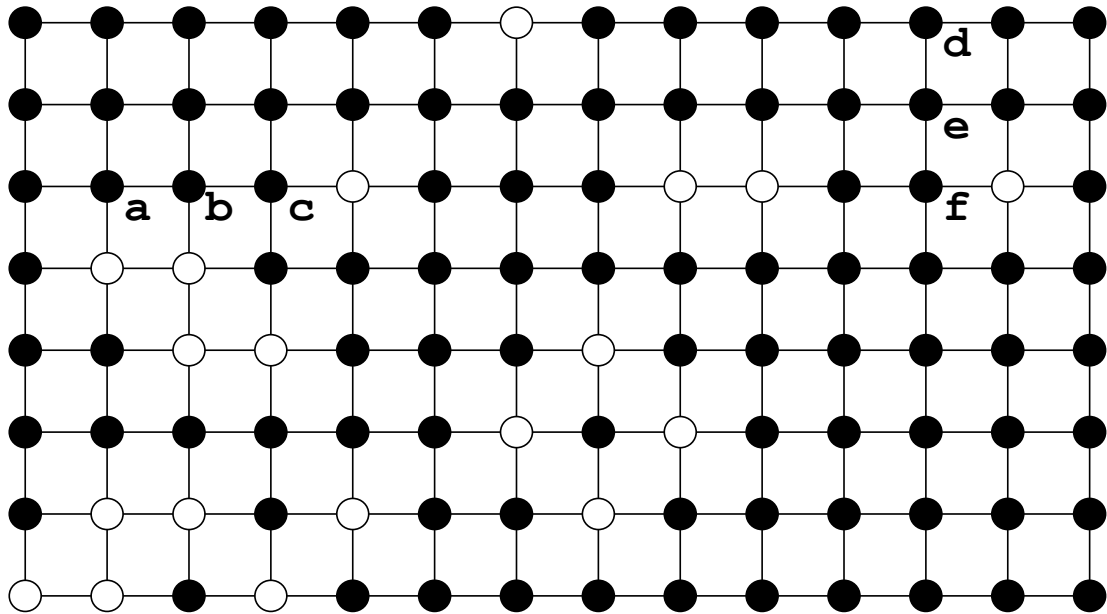


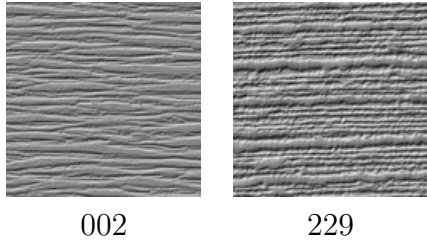
Figure 4.8: Neural Network Produced by SOM Analysis

inside a rectangular window, best represented by the sheet map shape shown. That is not to say it would be impossible to navigate any of the other combinations on a computer screen, but to keep interface development simple, and for all “neurons” to appear on screen at all times and without occluding other neurons, we shall restrict our investigation to the rectangular lattice sheet.

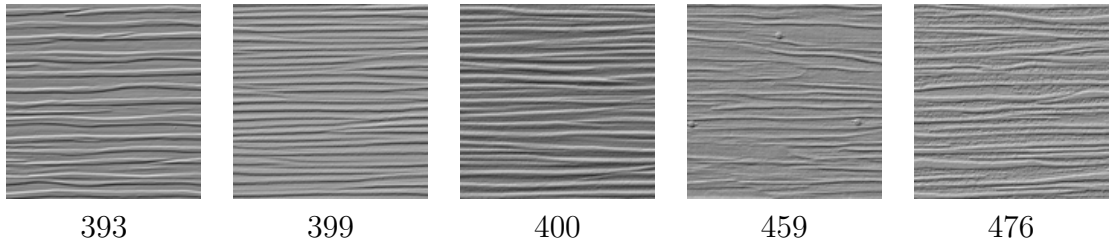
Figure 4.8 shows the SOM resulting from our perceptual similarity data. Neurons whose weight vector did not maximise the input vector of any member of our dataset are shown as unfilled circles while neurons containing one or more textures are shown filled. The empty neurons signify areas on the manifold where there are gaps in the texture space represented by the *Tex500* dataset.

Figures 4.9 and 4.10 show the contents of the numbered neurons at the top left and top right of figure 4.8 respectively. We can clearly see similarities between the textures within each neuron, as well as their close relationship with their neighbouring

Neuron (a)



Neuron (b)



Neuron (c)

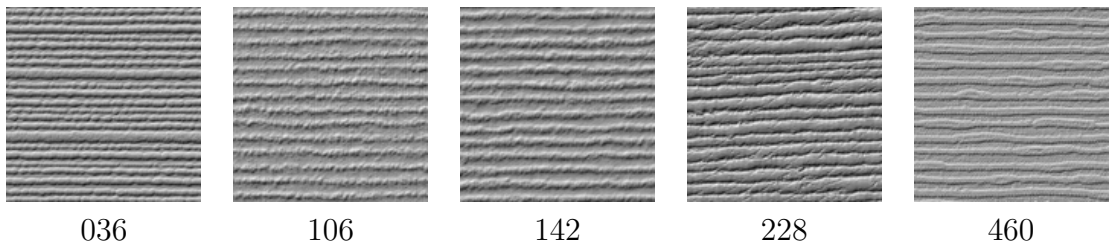


Figure 4.9: Top Left Numbered Neurons from Figure 4.8

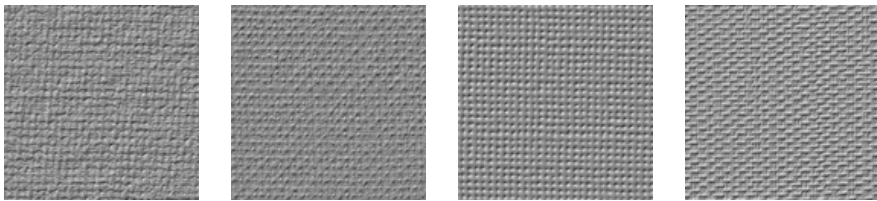
neurons.

Table 4.4 shows the outputs from processing the *Tex500* SOM. The quantisation error is the average distance between each data vector and its best matching unit (BMU). It measures map resolution. Topographic error is the proportion of all data vectors for which first and second BMUs are not adjacent units. It measures topology preservation.

Measure	Value
Quantisation error	1.188
Topographic error	0.054
Optimal Map size	14×8

Table 4.4: SOM Data Processing Result

Neuron (d)



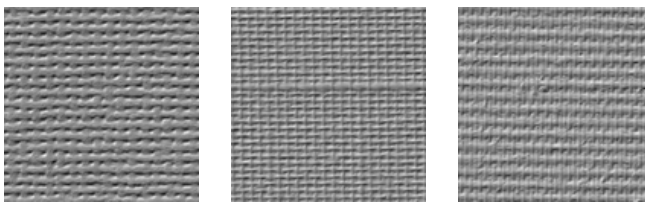
438

439

452

461

Neuron (e)

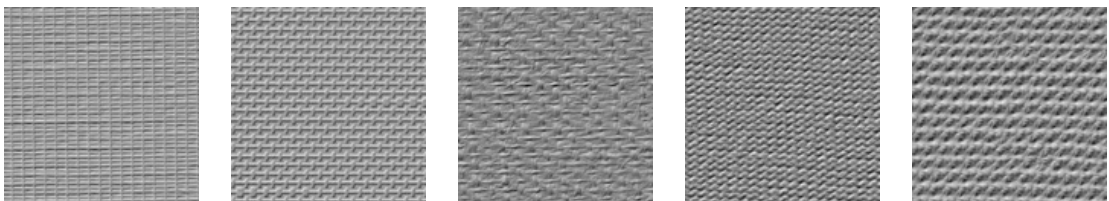


391

141

155

Neuron (f)



100

363

491

498

500

Figure 4.10: Top Right Numbered Neurons from Figure 4.8

4.5 Conclusions

In this chapter we set out to validate the results of our perceptual grouping experiments and to discover whether any obvious structures could be found in similarity matrix describing the *Tex500* dataset that we might exploit for navigation. We began by selecting three contrasting approaches for the analysis of multivariate data, two in the area of dimensionality reduction (multidimensional scaling and self organising map) and one hierarchical clustering approach.

With all three of the approaches selected, inspection by the author detected the presence of structures that suggested suitability for the organisation of the textures in three different browsing models. Our initial analysis is subjective and conclusions tentative, but it seems to validate the results. More rigorous investigation is required, and we will address this in chapters 5 and 6 where we discuss the design of three browsing environments (RF_G , MDS_G and SOM_G) and an experiment to compare the three in terms of accuracy and mean task time for a given task.

Although we have identified three possible browsing models, a number of issues need to be addressed as we refine these into usable browsing environments:

- Limited screen real estate:
 - How small can thumbnails be while still allowing users to correctly identify the texture?
 - What is the optimal SOM grid size: minimising topographic error vs. scrollbar free display of grid?
 - In hierarchical and SOM approaches, how should we represent collections of textures represented by clusters or neurons?
 - How do we manage scaling and occlusion when projecting textures in 2D/3D space using MDS co-ordinates?
- Navigation:
 - For each model, what is the optimal behaviour when selecting a texture/cluster at *navigating* levels in the structure?
 - Should we consider in advance the potential cost of *wrong turns* in the navigation?
 - Would a model benefit from redundancy in populating clusters/neurons, i.e. a texture appearing in more than one location?

Being particular to the design of browsing environments, these questions and issues will be addressed in chapter 5.

Chapter 5

Design Browsing Environments

5.1 Introduction

An early motivation for this thesis was a sparsity of commercial tools or learned research in the area of computer browsing environments. A search of the browsing environments literature in section 2.2 of the survey revealed few research projects in the area of browsing texture databases. Many scholarly articles that made mention of browsing in the title have been found to describe browsing in the context of displaying the result of a content-based image retrieval (CBIR) query, rather than browsing an image database per se. Some research was discovered about browsing environments for image databases which were organised using content based features such as the colour histogram distance metric [95], but there were no obviously transferable approaches that would assist with our navigation of *Tex500* in a perceptually relevant way.

Chapter 4 described our preliminary inspection of the similarity matrix using three visualisation techniques, from which we were able to identify the presence of plausible structure in the *Tex500* database. With this chapter we aim to develop three browsing environments that:

- exploit the structures identified in chapter 4,
- organise textures plausibly (observers remember or anticipate organisation),
- optimise use of screen real estate,
- facilitate fast elimination of unwanted textures,
- links texture proximity to similarity, and

- as a collection, provides sufficiently contrasting browsing environments for worthwhile comparison.

Given that we have already produced three structures describing the relationship between textures in the *Tex500* dataset, two key questions present themselves:

1. could our structures be used to develop a variety of perceptual space browsing environments? And if so,
2. could experiments be designed to discover which environment proves the most efficient to use?

Recall from chapter 4 that we have processed our raw perceptual similarity data into three distinct structural forms:

1. hierarchical data in the form of a dendrogram,
2. reduced dimensional co-ordinates using MDS, and
3. an artificial neural network using SOM.

In sections 5.2, 5.3 and 5.4 we describe the design of three browsing environments based on these structures, including details of human-computer interaction (HCI) component selection, navigation schema and layout design. The browsing environments will be referred to as SOM_G , MDS_G and RF_G^1 , the subscript ‘G’ denoting that the basis of the data organisation within these browsing environments is our initial grouping experiment in chapter 3.

5.2 Rapid-Fire Browsing Environment (RF_G)

If we had a dendrogram describing a sufficiently small dataset, it is possible to envisage a navigation schema based around the user descending through the hierarchy towards a particular texture or textures they may have in mind, refining their navigation where each node divides into two or more branches. At each decision point the user would have to be presented with all textures available along each branch in order to decide which path to select. If they decide they have gone in the wrong direction then an option to reverse one node at a time will allow them to ascend the structure until they find a more preferable route. This would be inconceivable with a large dataset, particularly at the higher levels of the hierarchy.

In order to navigate a tree structure, the user must know at each node what lies

¹The RF_G browsing environment is named after the *rapid-fire* component used in its implementation. This will be introduced later in this chapter.

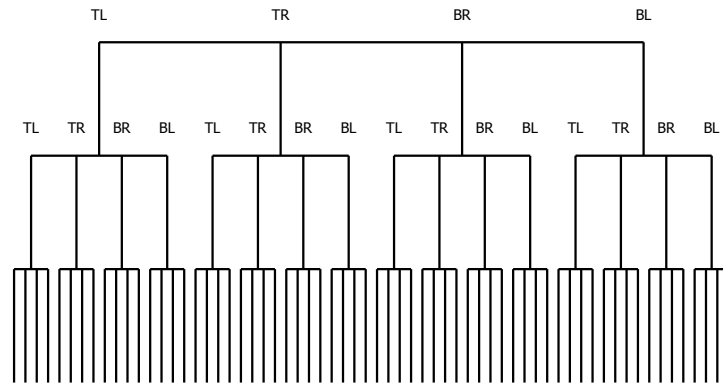


Figure 5.1: Rapid-Fire Tree Structure

down each of the branches ahead in order to decide which route they prefer. It would be difficult to model a complex hierarchy such as our dendrogram using the limited screen real estate available on the average computer monitor but there are steps we could take to simplify the structure. Since a computer monitor is rectangular we can easily segment it into smaller rectangles, and present in each an alternative route downwards through the hierarchy. Such sub-division could result in two, four, six, eight or more segments, but for reasons of simplicity and to provide sufficient space to represent each navigation on screen we have opted for segmentation into four quadrants.

We could make our perceptual hierarchy ready for such representation by partitioning the dataset into four clusters by drawing a horizontal line across the dendrogram at a point where it crosses four vertical branches. Hanging from each of these points is a sub-dendrogram on which we can repeat the process. By continuing to apply this process we will obtain a hierarchy of the form shown in figure 5.1. Finally, we can map each level of refinement to an on-screen position, TL (top-left), TR (top-right), BR (bottom-right), & BL (bottom-left).

The obvious problem here is that each of the clusters, particularly at the highest level of partitioning, could contain large numbers of textures. Indeed at the highest level we would have to somehow display four clusters for the user to choose from, each of which contains an average of one hundred and twenty-five textures. Of course we could choose to show only a random subset from the cluster such that all four clusters could be easily displayed on the user's screen. This is a poor solution in the case of a cluster containing a large variety of texture types since it is plausible that some key textures that may provide the essential cue for the user navigation may be omitted. This is particularly likely to occur at the highest levels in the hierarchy where the whole texture space encoded in the dataset must be represented. To find a solution to this problem, the author investigated the area of advanced

visual interfaces.

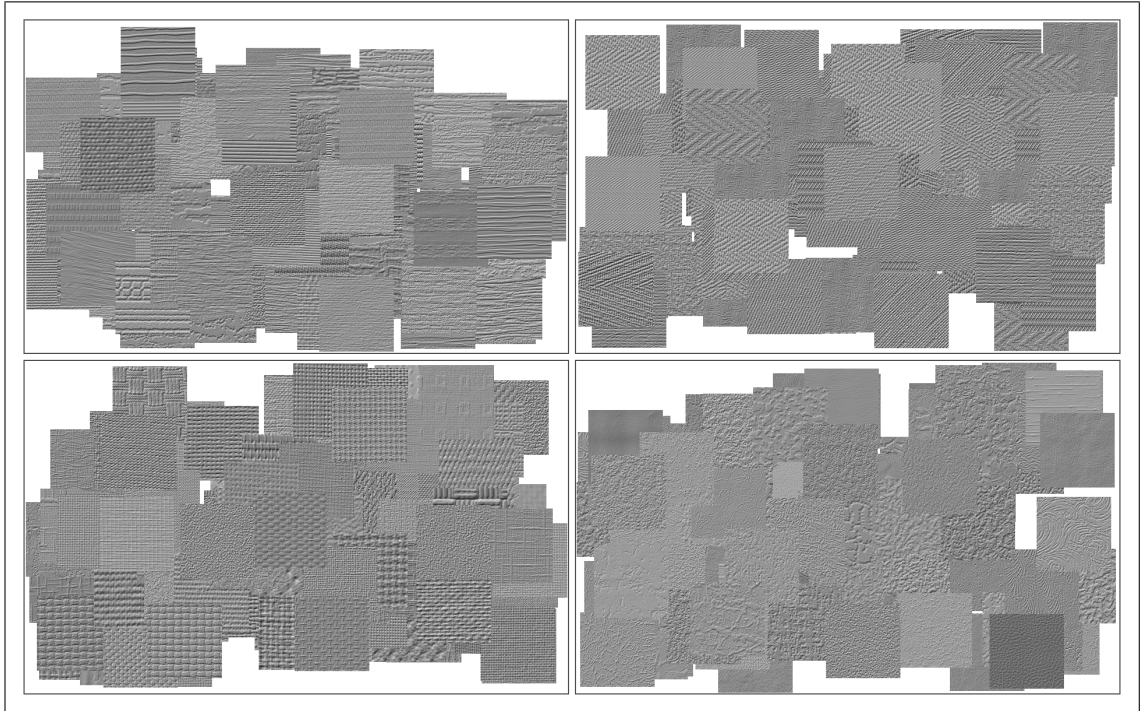


Figure 5.2: RF_G Browsing Environment Top Level Screen Shot

Wittenburg et al. [118] proposed prototypes for a means of navigating information spaces (such as collections of Internet sites) by presenting images that represent areas of the information space to the user in rapid succession within a preview window. The images, which are randomly positioned in the previewer, can be clicked upon by the user to proceed to that location in the information space. The approach goes some way to vastly increasing the number of possible out-links the user can follow in relation to the available screen real estate. The visual metaphor is of photographs being thrown randomly onto a table top. Over time the topmost photographs will occlude those underneath but the user will already have a flavour of the particular selection modelled. Although Wittenburg did not attempt to exploit any natural structure within the data his interface provided out-links to, it does suggest a user interface component which may be of high value for displaying example textures from a particular cluster or subset or out dataset within a browsing environment.

Our approach is an adaptation of the approach proposed by Wittenburg et al. We would display a distinct preview window for each cluster to be represented (see figure 5.2). Clicking on a particular preview window (not an individual image within the preview as described by Wittenburg et al.) would navigate downwards by one level in the hierarchy, and present four new preview windows from which to choose (see figure 5.3). In the event that the user clicks on a cluster with four or fewer member textures then they are presented with a detail view containing all member textures

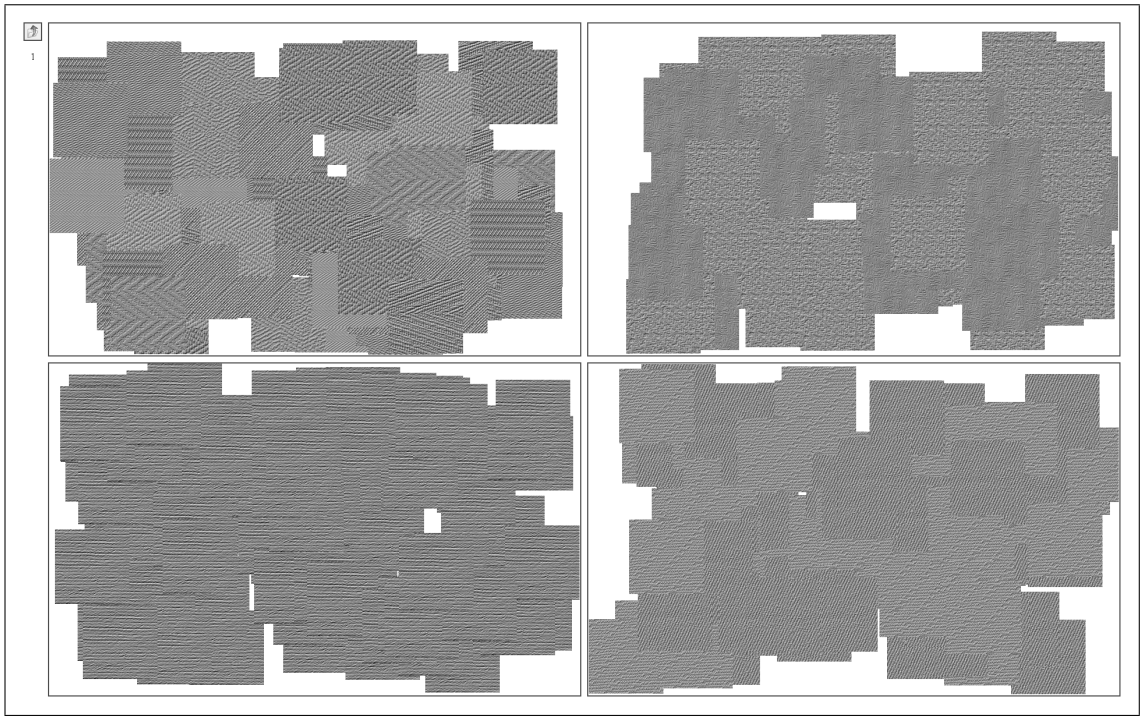


Figure 5.3: RF_G Browsing Environment Next Level Screen Shot

(see figure 5.4). In all but the highest level, an up button is provided for the user to reverse or step back up the hierarchy.



Figure 5.4: RF_G Browsing Environment Detail Level Screen Shot

5.3 MDS Browsing Environment (MDS_G)

The reduced dimensionality information obtained by performing multidimensional scaling (MDS) on our perceptual similarity data is an obvious possible source of structure for organising data within a browsing environment. For example, low dimensional MDS data has been successfully used in the past to map texture space in a 3D manifold [62]. It is also known to preserve the proximity information from the original data [24, ch. 10]. The proximity of similar textures within a browsing environment may suggest important navigational cues to assist the user.

Rogowitz et al. [95] conducted a variety of psychophysical experiments to measure the similarities between members of a dataset of ninety-seven digitised photographic images. They plotted thumbnail images first in flat 2D space using the 2D MDS

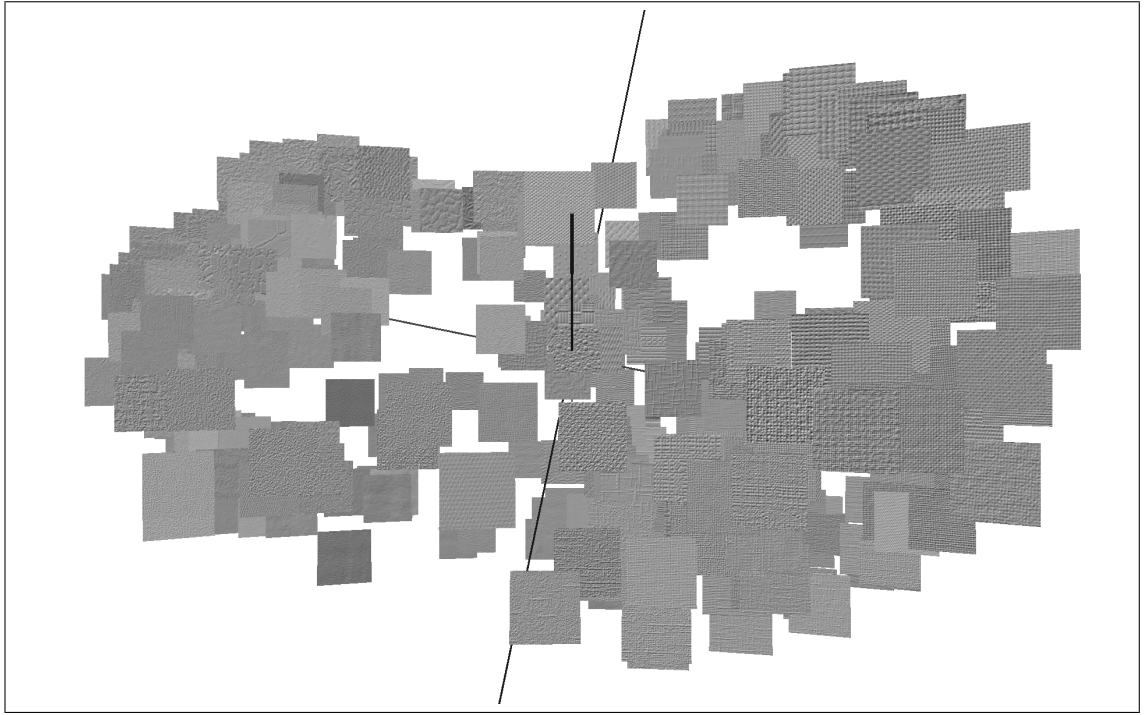


Figure 5.5: MDS_G Browsing Environment Top Level Screen Shot

co-ordinates obtained in order to perform analysis of the trends to be found in these two dimensions. They also plotted thumbnails in 3D space using virtual reality modelling language (VRML) which was viewed in an appropriate internet browser plug-in. In both cases, they found that images appearing similar to each other appear near to each other. As this fulfils our criterion that a browsing environment should arrange data logically and predictably, we proceeded to plot our surfaces in 3D space using the co-ordinates obtained from 3D MDS in an X3D (XML 3D) scene.

In order to make the 3D environment more intuitive, the texture thumbnails were rendered using X3D billboard components. This causes the thumbnails to always be displayed facing the scene camera (the user viewpoint). Figure 5.5 shows a screen capture of the 3D rendered scene. From this level in the interface, the user can rotate the scene in 3D space until they find the area containing the type of texture they wish to find. They can then click on an individual thumbnail which takes them to the detail level associated with the selected texture.

In the 3D scene three problems became evident while selecting textures:

1. thumbnails distant from the camera were barely distinguishable from one another due to scaling,
2. densely populated areas of the scene resulted in some thumbnails being obscured by others, and

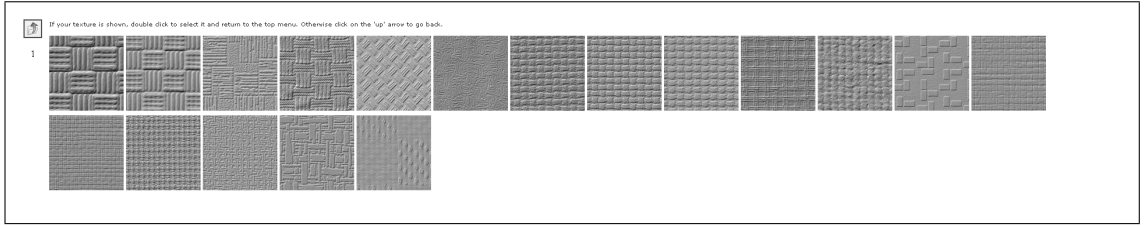


Figure 5.6: MDS_G Browsing Environment Detail Level Screen Shot

3. thumbnail edges are sometimes difficult to determine resulting in a nearby thumbnail being accidentally selected.

In cases where the wrong thumbnail was selected, the result is usually that one of its nearest neighbours is selected. The impact of a wrong selection here was reduced by displaying in the detail level (figure 5.6) the selected thumbnail along with all other textures with a non-zero similarity value to the selection, in descending order of similarity. Therefore, users need not be minutely accurate in selecting a texture from the 3D scene to still find the texture they desire.

5.4 SOM Browsing Environment (SOM_G)

An alternative approach to dimensionality reduction is the use of artificial neural networks. These are normally used in classification tasks and an oft-used example is the classification of Iris flowers into sub-species given the length and width measurement of sepal and petal leaves. Data points representing each set of measurements can be projected onto a variety of manifolds for analysis.

A particular artificial neural network approach is the self-organising map (SOM) which has been successfully used to organise documents and website pages based on relative percentages of keyword occurrences [55, 58, 43]. Prototype vectors are placed on a regular low-dimensional grid. These prototype vectors, or neurons, are d -dimensional weight vectors, where d is equal to the dimension of the input vector, in our case, five hundred.

Neurons are connected to adjacent neurons by a neighbourhood relation, which dictates the topology of the map. The topology consists of two factors: local lattice structure, which can be hexagonal or rectangular, and global map shape, either sheet, cylinder or toroid.

The most obvious choice of topology to be modelled on a computer screen would be the rectangular sheet (essentially a grid). By supplying our similarity matrix to a SOM implementation, a grid of neurons will be established which represents

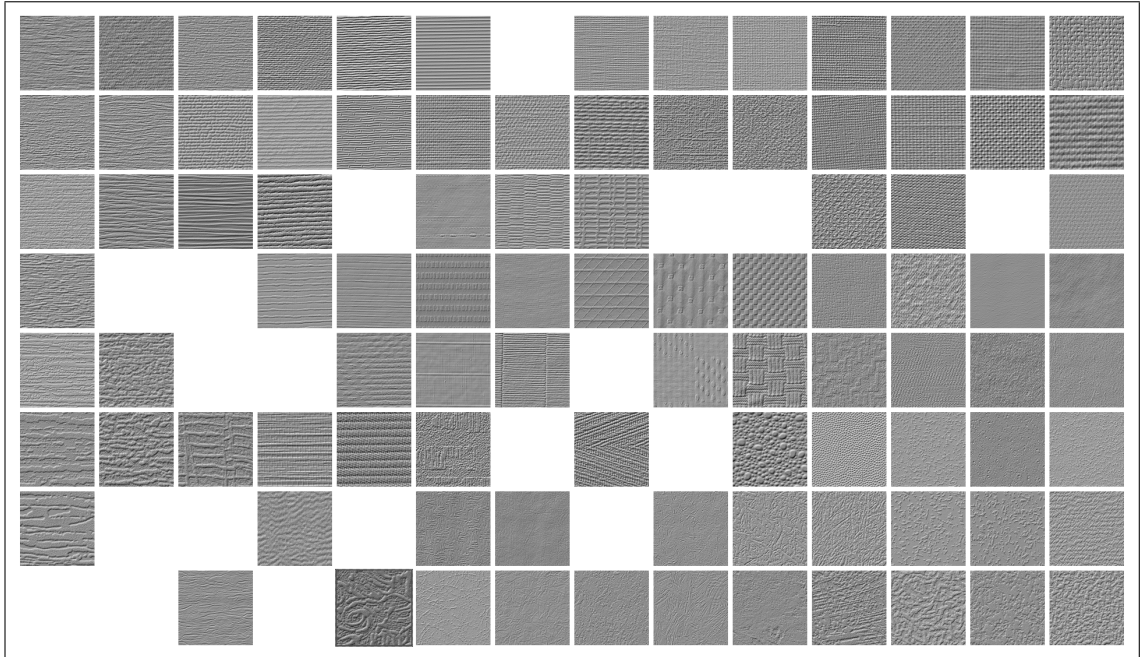


Figure 5.7: SOM_G Browsing Environment Top Level Screen Shot

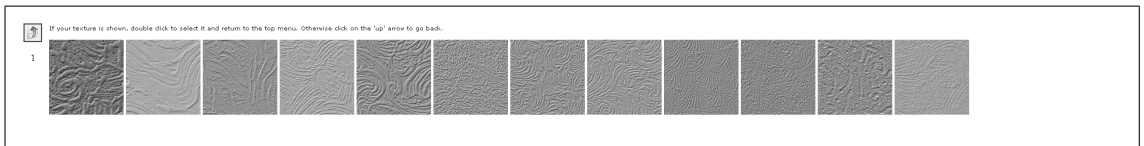


Figure 5.8: SOM_G Browsing Environment Detail Level Screen Shot

the variability within the dataset. The size of the grid can be automatic (reflecting dataset variability) or a particular grid size can be specified. In our case, the optimal grid shape generated by the SOM implementation coincided with the optimal grid shape that maximises screen real estate on the displays we selected for psychophysical experimentation. Each data point will be allocated to a single neuron depending on which neuron gives the strongest response. Each neuron (or grid position) will contain zero or more data points. Empty neurons can be thought of as areas where the measurement (texture) space was under-sampled. As each grid position represents zero or more textures we must make a decision as to how that collection is to be represented visually at the neuron level. The most straightforward solution is to display a thumbnail of the centroid texture of the cluster, using the Euclidean distance function. This provides a map view for our *Tex500* dataset as shown in figure 5.7.

On clicking one of these thumbnails, the user is taken to the detail level (figure 5.8) where they can see all member textures of the selected neuron, in distance order from the centroid texture whose thumbnail image they clicked on.

For a full description of the SOM please refer to Kohonen's book [53]. The author

used an implementation of SOM for Matlab by Vesanto et al. [112].

5.5 Conclusions

In chapter 4, we discussed the identification of structures in the similarity data derived from our sorting experiments. With this chapter we described how these structures were exploited to form data organisations and navigation schema in three contrasting browsing environments:

RF_G The rapid-fire browsing environment was based on the hierarchical clustering structure identified in section 4.2. The dataset is repeatedly partitioned into four clusters until the root nodes are reached and at each level the user is presented with four rapid fire previews representing the clusters. Users navigate by selecting a preview to descend further in the hierarchy or selecting an icon to reverse to the previous level. The author found this the most preferable browsing environment and found that he could easily navigate the dataset, and was generally able to anticipate his position within the structure making few wrong turns. Possible difficulties with this browsing environment are that it may be hard to identify poorly represented clusters of textures, particularly at higher levels in the hierarchy, and that wrong turns may be costly to operating time.

MDS_G The multidimensional scaling browsing environment was based on the dimensional analysis structure identified in section 4.3. Here the user is presented with a 3D scene representing all of the textures in the dataset. Whilst there certainly seems to be a strong correlation between proximity and similarity there is often no obvious wayfinding cues between clusters of similar textures, necessitating rotation of the view until the correct type of textures are found. This may have a detrimental effect on operating time. There is also the problem of occlusion, which is particularly problematic in areas where there is high representation of certain texture types. The author found this interface convenient for general browsing but users may have problems searching for a particular remembered texture.

SOM_G The self-organising map browsing environment was based on the neural network analysis structure identified in section 4.4. Users are presented with a grid of textures, each of which is representative of a neuron containing similar textures. By clicking on a texture they are presented with the neuron contents. At the grid level similarity between neuron contents is reflected by spatial proximity between grid items. A benefit of this model is that there is

generally low cost in terms of wrong turns but there may be issues around how representative the textures at grid level are of the neuron contents, particularly where there is lower cohesion between neuron contents. The author found this novel browsing environment easy to use but still favoured the rapid-fire browser.

Although the author’s perception of the advantages and disadvantages of each browsing environment is helpful in identifying potential issues for future users, we must subject the browsing environments to further scrutiny before coming to any conclusions. The author has considerable knowledge of the dataset, structures on which the browsing environments were based and how the browsing environments exploit these structures, so cannot be regarded as naive to the task of browsing environment assessment. We can, however, use his findings to form the hypothesis that the RF_G browsing environment will perform best. We test this hypothesis in chapter 6.

Given the high cost of obtaining perceptual data about a dataset using large grouping experiments like those described in chapter 3, a desirable product would be a technique for adding new textures to a dataset without repeating the grouping experiment. Browsing environments may provide us with a basis for this function and in chapter 7 we propose a simple dataset augmentation approach that features one of our browsing environments in a central role.

Chapter 6

SOM_G, MDS_G & RF_G Browsing Environments Comparison

6.1 Introduction

In this chapter we describe the experiment used to test the performance of the three browsing environments introduced in chapter 5. Having extensively used each interface to browse the dataset, the author found a preference for the rapid-fire image preview environment and formed the hypothesis shown in equation 6.1 (with respect to mean task time). In fact these experiments disproved that hypothesis.

$$\text{Hypothesis: } \text{RF}_G < \text{SOM}_G, \text{RF}_G < \text{MDS}_G \quad (6.1)$$

6.2 Experimental Design

The purpose of our experiment is to assess the efficiency of each of the browsing environments in question and to discover which, if any, is statistically the most efficient. We must therefore consider an appropriate task, stimuli and experimental approach to make that assessment.

6.2.1 Task Specification

As we wish to assess the efficiency of various browsing environments, we must model the kind of task that would be a core reason for using such a browsing environment in the first place. It would therefore seem appropriate to ask users to find a particular

texture from the dataset. The task stimulus could be a description of the surface to be found but this would be open to the interpretation of individual observers and it would be difficult to assess whether the observer had arrived at the correct, or at least a plausible selection. We therefore decided that the task stimulus, or query, should be an actual surface from the dataset that is present in the browsing environment.

6.2.2 Interface Presentation

As observers would have a range of cognitive abilities in relation to the task we decided it best to ask all observers test all three interfaces. To eliminate any learning effects from the experiment the order of presentation would be balanced by equalising the permutations available. As there were six possible permutations for three items, we planned to use each permutation twice, giving a total of twelve observer sessions.

6.2.3 Stimuli Selection

A pilot experiment revealed that observers could comfortably complete a total of twenty-four tasks in one hour without becoming fatigued. It was therefore decided that twenty-four stimuli would be used in the experiment, eight for each browsing interface.

To stabilise possible differences in task difficulty related to individual stimuli, the same set of stimuli would be presented to all observers. These would be presented in random order.

To ensure the stimuli represented a cross section of the dataset, the dataset was partitioned into twenty-four clusters using the dendrogram from subsection 4.2.1 and a random texture selected from each cluster. This stimuli set has been labelled *Tex500[024]* and can be found in full in Appendix A.1.

6.2.4 Sample Size

As we have already discussed, twelve observers would each carry out eight trials on each browsing interface, giving a total of ninety-six trials per browsing environment. Although we can never be certain from the outset that this is sufficient to reach a statistically significant conclusion, it does seem like a fair sample size.

6.2.5 Performance Measurement

Mean Task Time

Although we considered a number of ways to measure each task, by far the simplest is the time, in seconds, from presentation of the stimulus to the selection by the observer of their elected surface. We also recorded every navigation by users in case analysis of actual navigations versus minimum possible navigations might prove useful. However, the complexity of this analysis made it unsuitable for comparing task efficiency due to the variability in the effects of ‘wrong turns’ in each of the browsing environments.

Preferred Browsing Environment

For completeness, and in case measurement of mean task time proved not to be statistically significant, users were also asked to identify their preferred browsing environment at the end of the experiment.

Accuracy Assessment

As we have perceptual similarity data describing the relationship between all surfaces in the dataset, we can easily make an assessment of the accuracy of observers’ selections. In the event they find the exact match to the stimulus, a score of one is assigned. Otherwise the similarity value of the stimuli to their selection is assigned.

6.2.6 Observer Selection

Volunteers were invited from the student population of the School of Mathematical and Computer Sciences at Heriot Watt University. No age or other restrictions were placed on volunteers. Observers were informed they would have to use a mouse and VDU and that they should make sure if they needed to use corrective eyewear that they wore this for the experiment. No age data was collected from observers but in the opinion of the author they all fell within the 18–35 years age group.

6.2.7 Instructions to Observers

Prior to taking part in the experiment each observer was asked to complete a consent form with their name, address and email address and to sign an agreement in the

following terms.

I have agreed to take part in an experiment on the navigation of large texture databases. The procedure has been explained to me and I understand that I am free to leave the experiment at any time. In exchange I will receive 10 GBP worth of high street vouchers.

The same form continued with the following instructions which were read to the observer by the author.

1. The researcher will present you with three different interfaces for navigating a large texture database
2. You will be presented with 8 random query textures to search for using each of the interfaces
3. The researcher will describe each navigation interface prior to your using it
4. Your mouse clicks will be recorded throughout the experiment for later analysis
5. At the end of the experiment the researcher will ask you which interface you preferred using

The following description and an author-lead demonstration was given of each of the interfaces at point of use.

SOM_G This interface has a top level map where each grid square represents a collection of one or more textures. You should notice that similar collections of textures will be close to each other on this grid. Click on any grid square to see the member textures [author clicks top left texture]. To return to the grid, click on the arrow [author clicks arrow]. Double click a texture to select it in the detail level [author demonstrates]

MDS_G This interface has a 3D view showing every texture in the database. Navigate to the type of texture you are looking for by clicking on white space and rotating the 3D environment with your mouse [author demonstrates]. Click on the texture you are looking for, or a similar texture, and you will be presented with similar textures to the one you clicked. To return to the 3D view, click the arrow [author clicks arrow]. Double click a texture to select it in the detail view [author demonstrates]

RF_G This interface is a multi-layered hierarchy split consecutively into four parts at a time (the quadrants of the screen). Each quadrant has a rapid fire view of the textures it represents. Click a quadrant to descend one level [author demonstrates]. Each level has an arrow to return to the level before [author

clicks arrow to demonstrate]. At the top level there is no arrow. Once there are fewer than four textures left you will see all remaining textures in a grid view. From here you can double click a texture to select it or return to the hierarchical view by clicking the arrow. You do not need to wait until you see the particular texture you are looking for to select a quadrant, as you may be able to determine the correct route by identifying similar textures to the query

6.3 Analysis of Result

6.3.1 Mean Task Time

Result

Figure 6.1 shows the mean task time for each of the three browsing environments. These were 51, 112 and 143 seconds respectively. Standard error bars are also indicated on the plot, the standard error ($\sigma\bar{x}$) being calculated as shown in equation 6.2 where s is the sample standard deviation and N is the sample size. The intervals are plotted as shown in equations 6.3 and 6.4. A table of mean task times for each observer can be found in Appendix B.1.1.

$$\sigma\bar{x} = \frac{s}{\sqrt{N}} \quad (6.2)$$

$$\text{lower bound} = \bar{x} - \sigma\bar{x} \quad (6.3)$$

$$\text{upper bound} = \bar{x} + \sigma\bar{x} \quad (6.4)$$

Statistical Significance and Experimental Effect

Although we can clearly see the differences in the mean task time between browsing environments, we cannot draw any conclusions from this until we test the statistical significance of the differences [28]. The statistical test we use is called the *dependent t-test*, which is calculated from the mean difference between our samples (\bar{D}) and the standard error of differences ($\sigma_{\bar{D}}$) as shown in equation 6.5.

$$t = \left| \frac{\bar{D}}{\sigma_{\bar{D}}} \right| \quad (6.5)$$

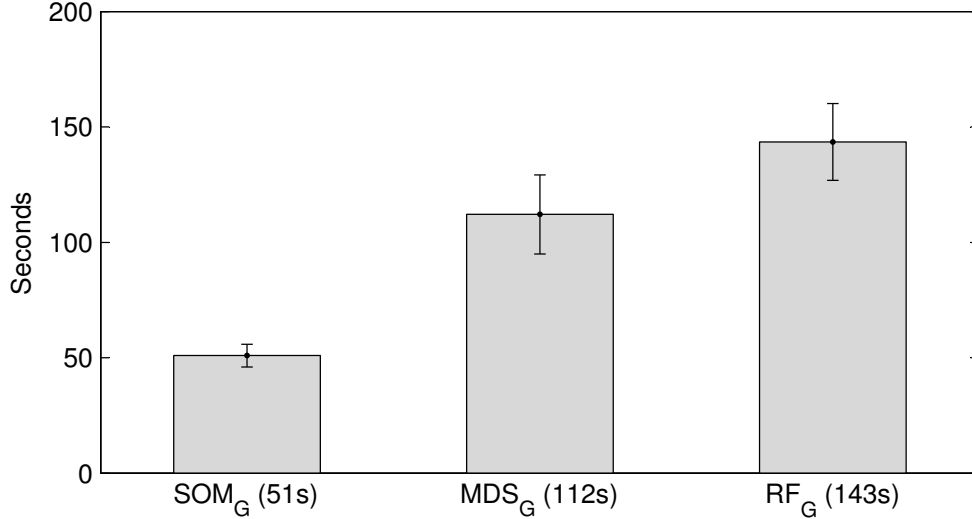


Figure 6.1: Mean Task Time (with Standard Error Bars) from SOM_G, MDS_G & RF_G Browsing Environments Comparison Experiment

We estimate the standard error of differences from the standard deviation of differences obtained within the sample (s_D) and the sample size (N) as shown in equation 6.6.

$$\sigma_{\bar{D}} = \frac{s_D}{\sqrt{N}} \quad (6.6)$$

Once we have a value for t we can use this to calculate the effect size, r , as shown in equation 6.7. df denotes degrees of freedom, which is derived from the sample size N as shown in equation 6.8.

$$r = \frac{t^2}{t^2 + df} \quad (6.7)$$

$$df = N - 1 \quad (6.8)$$

The analysis of our experimental results is summarised in table 6.1. As we are interested in an overall result reflecting 95% confidence (a p value of 0.05), and we are comparing three pairs of browsing environments, we must achieve a p value for each comparison of $p < 1 - \sqrt[3]{0.95}$ ($p < 0.01695$). As this p condition is achieved for the SOM_G over the other interfaces, it can be said that the differences between the mean task time in the SOM_G browsing environment and the others is statistically significant. This is not the case for the RF_G and MDS_G environments, however, so we cannot rank these with confidence. In terms of the r -values, by cross referencing with table 6.2, we can also see that there is a large effect size of the SOM_G over the

Interfaces Compared	t-test	p-value	Stat. Sig.	r-value	df
SOM _G vs. MDS _G	2.91	0.01	Y	0.66	11
SOM _G vs. RF _G	5.33	0.00	Y	0.85	11
MDS _G vs. RF _G	1.71	0.12	N	0.46	11

Table 6.1: Statistical Analysis of SOM_G, MDS_G & RF_G Mean Task Time

r-value	Effect
0.1	Small
0.3	Medium
0.5	Large

Table 6.2: Experimental Effect Sizes

other two browsing environments, indicating that the differences in mean task time was due largely to the differences in the browsing environments, rather than variance within the sample.

For detailed experimental results please see appendix B.1.1.

Preferred Browsing Environment

Table 6.3 shows the votes each browsing environment received when observers were asked to name which they preferred. This measure correlates exactly with the mean task time analysis.

6.3.2 Accuracy

Result

Of the twenty-four stimuli presented to observers, only four resulted in a non-exact match being selected by observers. As can be seen in figure 6.2, the non-exact selections (right of each pair) have a very high resemblance to the stimuli presented (left of each pair), and also high perceptual similarity values. There were a total of eight non-exact selections throughout all three browsing environments.

Browsing Environment	Votes
SOM _G	9
MDS _G	3
RF _G	0

Table 6.3: Observer Preference from SOM_G, MDS_G & RF_G Browsing Environments Comparison Experiment

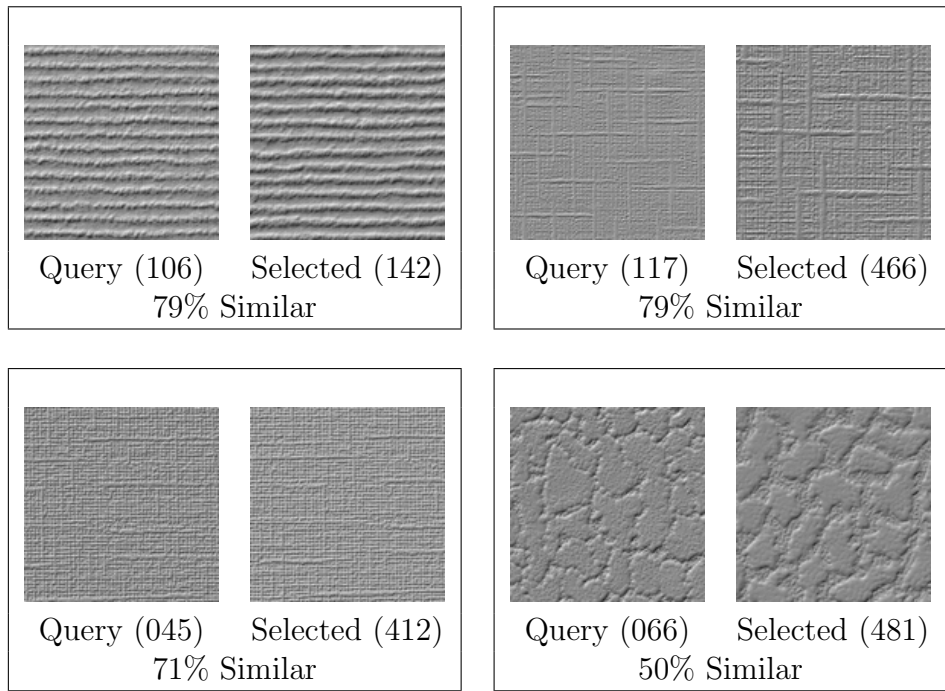


Figure 6.2: Non-Exact Observer Selections in SOM_G , MDS_G & RF_G Browsing Environments Comparison Experiment

Interfaces Compared	t-test	p-value	Stat. Sig.	r-value	df
SOM_G vs. MDS_G	0.24	0.81	N	0.07	11
SOM_G vs. RF_G	0.27	0.79	N	0.08	11
MDS_G vs. RF_G	0.50	0.62	N	0.15	11

Table 6.4: Statistical Analysis of SOM_G , MDS_G & RF_G Accuracy

By assigning a value of 100% to each accurate observer selection, and the percentage similarity value to non-exact selections, we can analyse the mean accuracy of each browsing environment as shown in figure 6.3.

Statistical Significance and Experimental Effect

Analysis of the statistical significance of differences in the accuracy of browsing environments is shown in table 6.4. In all cases, $p > 1 - \sqrt[3]{0.95}$ ($p > 0.01695$) indicating no statistical significance in the accuracy of any interface over the others. The r value shows very little experimental effect, suggesting the differences were naturally present in the sample and not as a result of the issue under examination.

For detailed experimental results please see appendix B.1.2.

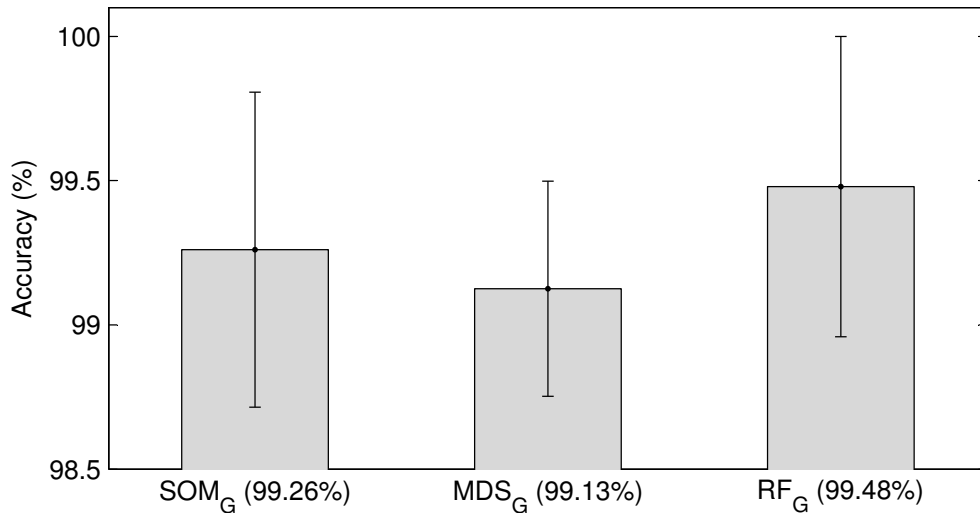


Figure 6.3: Accuracy (with Standard Error Bars) from SOM_G, MDS_G & RF_G Browsing Environments Comparison Experiment

6.4 Discussion

In keeping with many PhD research projects, the author recruited observers for this and the other browsing environment comparison experiments in this thesis from the student population of his own department, in this case, the School of Mathematical and Computer Sciences at Heriot Watt University. Although there is no evidence to suggest this group were not naive to the task of texture search, it could be argued they universally hold higher than average computer use skills and therefore would perform better than population in the use of browsing environments. Potential negative effects of this were minimised by the exclusive use of such observers for these experiments and by using a repeated measures approach where all observers tested all variables.

At the point of recruitment, observers were made aware of the estimated duration of the experiment and the fixed payment they would receive to compensate them for their participation. Although there was no observable behaviour on the part of participants to modify their performance to say, fill the time, we have no way of knowing what effect the payment structure might have had on the outcome of our experiment.

All of our experiments were based on a database of five hundred textures, which we consider by comparison to other texture research projects to be large. However, there is no guarantee that we would obtain similar results if the dataset size was significantly increased or decreased. We also limited our browsing to a texture database, but our results may well have been different given an alternative class of

visual stimuli.

6.5 Conclusions

Our experiment revealed that the SOM_G browsing environment was more efficient than the MDS_G and RF_G interfaces, but that the MDS_G and RF_G interfaces could not be ranked with any statistical confidence. There was an exact correlation between this result and the number of votes received by each browsing environment when observers were asked to identify which they preferred. Non-exact selections made by observers revealed no statistically significant differences between the accuracy of the browsing environments.

The success of the SOM_G browsing environment in this experiment may be explained by its relative simplicity and clarity. In terms of navigation this is positive as users can quickly navigate between the two states of top (grid) level and group level in a single click. This contrasts with the RF_G browser where there are up to six levels of hierarchical navigation between the root level and a particular group. This can prove costly in the event of a wrong turn.

Although the MDS_G browser also shares with SOM_G the property of having only two levels, its poorer performance may be explained by the organisation of the 3D projection level. Occlusion is an obvious problem where textures near the viewpoint mask or partially mask textures behind. Also, textures that are far from the viewpoint are scaled according to perspective and properties important to their identification may become undetectable with reduced resolution.

As the SOM_G browsing environment has been identified as the most efficient of those tested here, the remainder of the research in this thesis will use that model for the implementation of browsing environments.

Although this chapter identifies an efficient browsing model with respect to mean browsing time for a particular task, it does not in itself provide us with any scalable solution for capturing perceptual similarity data for browsing large databases. In this thesis we examine two scalable methods for obtaining perceptual similarity data. The first is a technique for augmenting similarity matrices with additional data members (chapters 7 and 10). The second is using computational features to calculate the similarity between textures (chapter 8).

Chapter 7

Pilot Dataset Augmentation

7.1 Introduction

In chapter 3 we demonstrated that obtaining similarity data about a texture dataset can be costly, particularly as the dataset size increases. Although we described a method of conducting grouping experiments on hierarchically partitioned subsets of our *Tex500* dataset, this approach was only possible as we had already obtained similarity data about the whole dataset from 6 pilot grouping experiments where users sorted the entire *Tex500* dataset. These pilot grouping experiments were not scalable and had to be abandoned due to problems (see subsection 3.3.1), and although we retained the pilot data to develop our pseudo-scalable grouping approach, it would be much more valuable to find an inexpensive method for inserting a new texture into a dataset for which we already hold perceptual similarity data.

If we could find a method of augmenting an existing dataset with new previously unseen textures we could avoid two significant problems:

- Initial grouping experiments could be restricted to a manageable size for the average observer, and
- The addition of new textures would no longer require expensive grouping experiments to be repeated.

Until now we have asked users to find a given query texture that they already know to be present in the browsing environment, but there is no reason why the task cannot be modified to browsing for similar textures to a given query texture. An algorithm could then be developed for translating the similarity information we hold about the selected textures into similarity information about the new surface.

In this chapter we describe a simple method for augmenting a dataset with new

textures, design an experiment to simulate adding some new textures to the dataset then build a browsing environment (SOM_A) using the augmented dataset. Later, in chapter 9, we shall compare the efficiency of the resulting browsing environment with the browsing environment derived from our initial grouping experiment (SOM_G) and a browsing environment derived in chapter 8 from computational features (SOM_F).

7.2 Relationship Between New Textures and Existing Textures

In considering approaches for inserting new textures in the dataset, we must first understand the data we hold that describes the existing members of the dataset and the type of translation we need to perform to describe the new texture in an equivalent form. The *Tex500* dataset is described with a 500×500 perceptual similarity matrix where each texture has a similarity relationship with every other texture in the dataset, including itself. The similarity values are in the range $[0 - 1]$ where 1 is maximal similarity. Every texture has a similarity value with itself of 1.

Each texture in the dataset therefore has a 500-dimension similarity vector describing its similarity with all textures in the dataset. Similarity vectors for the first three textures in the dataset can be seen in equations 7.1, 7.2 and 7.3. Actual similarity values have been replaced with variable s with a subscript denoting the texture indexes of the similarity value.

$$sv_1 = [1, s_{(1,2)}, s_{(1,3)}, s_{(1,4)}, \dots, s_{(1,500)}] \quad (7.1)$$

$$sv_2 = [s_{(2,1)}, 1, s_{(2,3)}, s_{(2,4)}, \dots, s_{(2,500)}] \quad (7.2)$$

$$sv_3 = [s_{(3,1)}, s_{(3,2)}, 1, s_{(3,4)}, \dots, s_{(3,500)}] \quad (7.3)$$

By asking observers to select the most similar existing textures to a new texture with which we wish to augment the dataset, we can derive a similarity vector for that new texture using the similarity vectors from the selected existing textures. An obvious approach is to take the mean of each element in the similarity vectors selected. The following paragraphs describe three simple approaches to deriving the new similarity vector using a simple mean, ordered mean or weighted mean as the basis for the derivation.

Mean Similarity Vector

The simplest conceivable solution to adding a new texture to the dataset would be to assign it a similarity vector which represents the mean similarity values of the textures selected by a user as those most similar to the new texture. If, for example, a user is presented with new texture numbered 501 and they select textures 1, 2 and 3 as the most similar existing textures in the dataset, then texture 501 would be assigned the similarity vector shown in equation 7.4. Observe that each time a new texture is added, the length of all similarity vectors increase by 1 and the added texture is assigned a similarity with itself of 1.

$$\begin{aligned}
 sv_{501} &= [d_{(501,1)}, d_{(501,2)}, d_{(501,3)}, d_{(501,4)}, \dots, d_{(501,500)}, 1] \\
 &= \left[\frac{1 + s_{(2,1)} + s_{(3,1)}}{3}, \frac{s_{(1,2)} + 1 + s_{(3,2)}}{3}, \frac{s_{(1,3)} + s_{(2,3)} + 1}{3}, \right. \\
 &\quad \left. \frac{s_{(1,4)} + s_{(2,4)} + s_{(3,4)}}{3}, \dots, \frac{s_{(1,500)} + s_{(2,500)} + s_{(3,500)}}{3}, 1 \right] \quad (7.4)
 \end{aligned}$$

Ordered Similarity Vector

There may be occasions where an observer perceives that the most similar existing textures to the new texture are not equally so. In a refinement to the above method, we could give the observer the opportunity to order their response by strength of similarity or to freely attribute weights to their responses. Ordering would necessitate some predefined and somewhat arbitrary weighting method being applied to the ordered texture selection. An example is shown in equation 7.5 where the textures are weighted in decrements of 1 from the total number of textures selected. This ordered approach may or may not offer improvement over the simple averaging approach since the weights may not accurately reflect the observers perceived decrement in similarity. Indeed these weights would be applied even in the unlikely event that the observer considers their entire selection to be of equal similarity to the stimulus.

$$\begin{aligned}
 sv_{501} &= [d_{(501,1)}, d_{(501,2)}, d_{(501,3)}, d_{(501,4)}, \dots, d_{(501,500)}, 1] \\
 &= \left[\frac{3 + 2s_{(2,1)} + s_{(3,1)}}{6}, \frac{3s_{(1,2)} + 2 + s_{(3,2)}}{6}, \right. \\
 &\quad \frac{3s_{(1,3)} + 2s_{(2,3)} + 1}{6}, \frac{3s_{(1,4)} + 2s_{(2,4)} + s_{(3,4)}}{6}, \\
 &\quad \left. \dots, \frac{3s_{(1,500)} + 2s_{(2,500)} + s_{(3,500)}}{6}, 1 \right] \quad (7.5)
 \end{aligned}$$

Weighted Similarity Vector

Equation 7.6 shows the algorithm where the observer is able to freely assign their own weights to their selection. Although this approach could result in greater accuracy and lower quantisation in the resulting similarity vector, it places considerable additional judgement effort on the observer and may increase observer fatigue and have a negative impact on our ability to recruit observers.

$$\begin{aligned}
 sv_{501} &= [d_{(501,1)}, d_{(501,2)}, d_{(501,3)}, d_{(501,4)}, \dots, d_{(501,500)}, 1] \\
 &= \left[\frac{w_1 + w_2s_{(2,1)} + w_3s_{(3,1)}}{w_1 + w_2 + w_3}, \frac{w_1s_{(1,2)} + w_2 + w_3s_{(3,2)}}{w_1 + w_2 + w_3}, \right. \\
 &\quad \frac{w_1s_{(1,3)} + w_2s_{(2,3)} + w_3}{w_1 + w_2 + w_3}, \frac{w_1s_{(1,4)} + w_2s_{(2,4)} + w_3s_{(3,4)}}{w_1 + w_2 + w_3}, \\
 &\quad \left. \dots, \frac{w_1s_{(1,500)} + w_2s_{(2,500)} + w_3s_{(3,500)}}{w_1 + w_2 + w_3}, 1 \right] \tag{7.6}
 \end{aligned}$$

Selected Method

Because of the potential problems highlighted in the ordered and weighting approaches detailed above, and to keep this pilot augmentation experiment as simple as possible, we decided to use the mean similarity method. We will aim to minimise the problems in achieving the average human response by acquiring 10 individual observations per stimulus.

Figure 7.1 shows an abbreviated similarity matrix illustrating the relationship between existing dataset members (textures 1-500) and added textures (501-503). Here, s values denote similarity values for existing dataset members while d values denote the similarity values derived from the existing similarity values of those textures judged by observers to be most similar to added textures. This is the same indexing used in the first line of equations 7.4, 7.5 and 7.6.

7.3 Inter-Relationship Between New Textures

Having measured the perceptual similarity between pairs of textures in our existing dataset, and proposed a method (section 7.2) of deriving the similarity values of new textures with existing textures from these measured values, we need to develop a method for deriving the similarity values reflecting the inter-relationship between new textures to complete the augmented similarity matrix.

1	$m_{(1,2)}$	$m_{(1,3)}$	\dots	$m_{(1,500)}$	$d_{(1,501)}$	$d_{(1,502)}$	$d_{(1,503)}$
$m_{(2,1)}$	1	$m_{(2,3)}$	\dots	$m_{(2,500)}$	$d_{(2,501)}$	$d_{(2,502)}$	$d_{(2,503)}$
$m_{(3,1)}$	$m_{(3,2)}$	1	\dots	$m_{(3,500)}$	$d_{(3,501)}$	$d_{(3,502)}$	$d_{(3,503)}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
$m_{(500,1)}$	$m_{(500,2)}$	$m_{(500,3)}$	\dots	1	$d_{(500,501)}$	$d_{(500,502)}$	$d_{(500,503)}$
$d_{(501,1)}$	$d_{(501,2)}$	$d_{(501,3)}$	\dots	$d_{(501,500)}$	1	$dd_{(501,502)}$	$dd_{(501,503)}$
$d_{(502,1)}$	$d_{(502,2)}$	$d_{(502,3)}$	\dots	$d_{(502,500)}$	$dd_{(502,501)}$	1	$dd_{(502,503)}$
$d_{(503,1)}$	$d_{(503,2)}$	$d_{(503,3)}$	\dots	$d_{(503,500)}$	$dd_{(503,501)}$	$dd_{(503,502)}$	1

Figure 7.1: Mapping of Measured (m), Derived (d) and Double Derived (dd) Similarity Values in the Augmented Similarity Matrix

Figure 7.1 shows our augmented similarity matrix mapped with the areas where the similarity values are measured, derived and double derived. In this section we will define the approach used to calculate similarity values in the double derived portion of the augmented similarity matrix.

For each new texture with which we augment the dataset, we first seek existing members of the dataset which most closely resemble these new additions. We can estimate the similarity between those new textures by taking the mean of the similarity values of all pairwise combinations of the textures they are judged to be similar to. If, for example, we augment the dataset with two new textures, 501 and 502, and observers judge that 501 is similar to existing textures 1 and 2, and 502 is similar to 5, 6 and 7, then the double derived similarity between 501 and 502 is as shown in equation 7.7.

$$dd_{(501,502)} = dd_{(502,501)} = \frac{m_{(1,5)} + m_{(1,6)} + m_{(1,7)} + m_{(2,5)} + m_{(2,6)} + m_{(2,7)}}{6} \quad (7.7)$$

7.4 Design Data Augmentation Experiment

7.4.1 Stimuli Selection

In section 6.2.3 we described the selection of a subset of the *Tex500* dataset representing a cross section of all available textures. This subset, known as *Tex500[024]* will also be used as the stimuli for this experiment. This stimuli set can be seen in full in appendix A.1. We must also discard the existing similarity data for the *Tex500[024]* textures in our perceptual similarity matrix.

7.4.2 Design SOM_R Browsing Environment

In section 6.2.5, we concluded that the SOM_G browsing environment was the most efficient of those tested. We will therefore use this browsing environment (and its underlying data organisation) as the instrument for capturing similarity perceptions from observers here. As the object of the task is to assign a similarity vector to textures not already part of the dataset, the *Tex500[024]* textures will be removed from the browsing environment. The resulting reduced SOM browsing environment will be referred to as SOM_R.

7.4.3 Task Specification

For each stimulus, observers will be asked to use the SOM_R browsing environment to select those textures they consider to be most similar to the stimulus. They must select two or more textures before continuing to the next trial. They record their response by adding similar textures to a palette from which textures may also be removed in the event they discover a more suitable selection. Once they have submitted their response to a trial, no changes can be made.

7.4.4 Sample Size

The experiment will be conducted with ten observers, each of whom will be exposed to all twenty-four stimuli. The stimuli will be presented in random order.

7.4.5 Observer Selection

Volunteers were invited from the student population of the School of Mathematical and Computer Sciences at Heriot Watt University. No age or other restrictions were placed on volunteers. Observers were informed they would have to use a mouse and VDU and that they should make sure if they needed to use corrective eyewear that they wore this for the experiment. No age data was collected from observers but in the opinion of the author they all fell within the 18–35 years age group.

7.4.6 Instructions to Observers

Prior to taking part in the experiment each observer was asked to complete a consent form with their name, address and email address and to sign an agreement in the

following terms.

I have agreed to take part in an experiment on perceptual similarity of textures. The procedure has been explained to me and I understand that I am free to leave the experiment at any time. In exchange I will receive 10 worth of high street vouchers in return for my participation.

The same form continued with the following instructions which were read to the observer by the author.

1. The researcher will present you with twenty-four query textures
2. Your task is to find the most similar textures using a grid interface. The query texture does not appear in the interface
3. Once you have found two or more similar textures, click submit on your palette to record your selection. Try to select all textures that are similar to the query texture
4. Click 'next stimulus' to proceed to the next query texture

The following description and an author-lead demonstration was given of each of the interfaces at point of use.

This interface has a top level map where each grid square represents a collection of one or more textures. You should notice that similar collections of textures will be close to each other on this grid. Click on any grid square to see the member textures [author clicks top left texture]. To return to the grid, click on the arrow [author clicks arrow]. Double click a texture to add it to your palette from the detail level [author demonstrates]. You must find two or more similar textures for each query texture.

7.5 Augmentation Experiment Data Analysis

7.5.1 Votes

The outcome of the similarity perception experiment is an array of similar texture votes for each of the stimuli. This result can be viewed in full in appendix B.2.

7.5.2 Build SOM_A Browsing Environment

By applying our algorithm (see equation 7.4 above) to the result of the similarity perception experiment we obtain a new similarity matrix which we use as the data organisation for a new browsing environment. The browsing environment built on this augmented data organisation is referred to as SOM_A .

7.6 Conclusions

In this chapter we described the design of a simple augmentation approach for inserting unknown textures in a dataset described by a perceptual similarity matrix. We demonstrated proof of concept by removing a small subset of textures from our dataset and conducting an augmentation experiment and assigned a similarity vector to these ‘unknown’ textures thereby creating a new augmented similarity matrix which we used to build browsing environment SOM_A .

By using the SOM_A browsing environment to browse the dataset, the author formed the opinion that the data organisation was logical and did not represent an inferior experience to the SOM_G browsing environment. He was therefore able to form the hypothesis shown in equation 7.8.

$$\text{Hypothesis: } SOM_A \approx SOM_G \quad (7.8)$$

This hypothesis, together with the hypothesis proposed in section 8.4 are tested in a browsing environment comparison experiment in chapter 9.

Chapter 8

Identifying Features for Data Organisation

8.1 Introduction

We described in chapter 3 the technique we used to capture human judgements on the perceptual similarity of the members of our *Tex500* dataset. Previous researches have sought to map this perceptual space to a corresponding space of computational features for the purpose of automatic texture retrieval. With this chapter we aim to borrow the techniques of these researches to produce a feature-based data organisation on which we can build a browsing environment. We demonstrated in chapter 6 that the most efficient browsing environment tested was that based on the self organising map, and we shall continue using only this browsing environment to navigate the resulting feature-based organisation. We shall label this browsing environment SOM_F .

One notable recent piece of work on identifying perceptually relevant computational features for surface texture is the PhD thesis of Emrith [25, ch. 5 & 6]. He describes in detail the process of automated feature selection from a large set of features (i.e. several thousands). He also investigates a number of candidate feature extraction methods for inclusion in this large set of features and examines their merits and weaknesses in relation to his selection criteria. As the main theme of our research is the investigation of a variety of browsing environments and underpinning data organisations, we shall be using the same pool of features and feature selection techniques used by Emrith as something of a *black box* without introducing any new ideas in his subject area. Our only difference in application is that we shall use his approach to produce a data organisation for use in a browsing environment rather than as a retrieval engine. Any work on texture retrieval and feature development

and selection is beyond the scope of our research and this chapter should be read as a précis of the two aforementioned chapters of his work and a report on the application of his work to our application.

In chapter 7, we produced a SOM based on our pilot dataset augmentation approach, SOM_A . In order to examine the efficiency of the alternative data organisations we have produced we shall compare browsing environments SOM_G , SOM_A and SOM_F in an experiment described in chapter 9. In each case the only changing variable is the specific data organisation underpinning the browsing environment. That experiment is expected to facilitate an evaluation of the applicability of our augmentation approach and Emrith's feature selection approach to the data organisation of browsing environments.

8.2 Identifying Features for Texture Description

8.2.1 Feature Selection Criteria

Emrith's feature selection criteria were developed for the avoidance of bias in selecting texture descriptors and the restriction of candidate feature descriptions to those already investigated and well described in the literature. Additionally the following specific criteria were described in full.

Phase Sensitive Features: The phase spectrum contains most or all of the structural information in an image and contributes immensely in helping people to recognise and interpret objects or structure within an image.

Power Spectral Sensitive Features: Different textures normally generate different energy distributions in the frequency domain and that variation can be efficiently captured within the power spectrum, which represents the strength of each spatial frequency.

Position Independent Features: This refers to the ability of a feature descriptor to recognise two samples with similar texture primitives as similar when the texture primitives are displaced by a certain amount. Human observers have no trouble in making such judgements but computational descriptors that mimic this behaviour are rare.

Large Pool of Features: The smaller the number of features employed, the greater the bias in representing the different texture categories in a dataset [83]. By starting from the idea that we cannot know what low-level features might contribute to the high-level descriptions of a texture then we must recognise

that the larger the set of computational features, the less prejudiced the potential high-level representation will be.

Avoidance of Redundant Features: Highly correlated features or features that contribute in the same way to describe a particular texture characteristic increase the computational complexity and degrade the performance of a description system and must be avoided.

Inexpensive and Simple to Compute: This characteristic follows from the requirement to have a large set of features.

8.2.2 Feature Extraction Methods

Emrith considered four classes of computational features as candidates for inclusion in his large set of features. The following lists and summarises the classes considered. For a complete description please see [25, pp. 90-100].

Local Binary Patterns: Generate binary codes that describe how the local texture pattern is built and was first introduced as a complimentary measure for local image contrast [72, 73]. Although LBP operators are simple to design and implement and computationally cheap, they cannot capture large-scale features and can result in histograms with a large number of bins.

Gabor Wavelets: Allow multi resolution (or multi spectral) decomposition through proper tuning of orientation and radial frequencies. Can be designed to be highly selective in both position and frequency [15, 23]. Although they share common Human Visual System properties, they suffer from position sensitivity.

Simoncelli's Features: Follow a number of texture models that are based on the application of oriented linear kernels at multiple spatial scales. Features are derived from fixed over-complete multi-scale complex wavelet representations [102, 88]. Although they can generate large feature sets through varying scale and orientation, they tend to contain considerable redundancy and many of the phase sensitive features are also position sensitive.

Trace Transform Features: Tracing an image with straight lines along which certain functionals of the image function are calculated. Different functionals that can be used may be invariant to different transformations of the image [49, 82]. By varying the types of functionals, the generation of thousands of features is possible, but results in significant memory utilisation and disk space requirements.

Features		Criteria					
		Position independent	Phase sensitive	Power Spectrum	No Redundancy	Inexpensive & Simple	Large Pool
Multi-scale LBP		✓	✓	✓	✗	✓	✗
Gabor	Power	✓	✗	✓	✗	✓	✗
	Phase	✗	✓	✗	✗	✓	✗
Simoncelli		✗	✓	✓	✗	✗	✓
Trace Transform		✓	✓	✓	✓	✓	✓

Table 8.1: Eligibility of Feature Extraction Methods by Criterion

8.2.3 Feature Set Selection

By cross referencing his feature set criteria with the candidate feature extraction approaches (see table 8.1), Emrith was able to eliminate all but the Trace Transform features from his large set of features. We used exactly the same set of features in generating our feature-based data organisation.

8.2.4 Feature Normalisation

When developing a large pool of features there is a risk of introducing a large variation in the span between individual features. To address the potential problem of computing distance values dominated by features with wide value ranges it is sensible to apply some normalisation to the features included. After consideration of a range of normalisation approaches [2], Emrith selected an approach where all features have zero mean and unit variance. The features were finally scaled to the range [0, 1]. His transformation is shown in equation 8.1.

$$\tilde{x} = \frac{(x - \mu)/3\sigma + 1}{2} \quad (8.1)$$

8.3 Producing a Feature-Based Dataset Description

8.3.1 Similarity Matrix Dimensionality Reduction

Recall figure 4.4 from chapter 4 plotting the stress values for our 4D metric random, pilot and final data after running multidimensional scaling (MDS). We again wish to use this data to decide upon the correct number of dimensions to which we reduce our final data before running linear regression for feature selection. We must be

Dimensions	Stress
1	0.427
2	0.232
3	0.169
4	0.127
5	0.102
6	0.085
7	0.075
8	0.065

Table 8.2: Stress Values for Final Data from Figure 4.4

confident that the number of dimensions sufficiently describes the variability in the similarity matrix being reduced such that we can have confidence in the linear regression to follow. Figure 4.3 (chapter 4) details that a *fair* goodness of fit can be achieved at a stress value of 0.1. Table 8.2 details the actual stress figures for the final experimental data after subjecting it to MDS. We can see that a figure of 0.1 is approached at 5 dimensions and we will use this number of dimensions to encode the final experimental data before subjecting it to linear regression.

8.3.2 Feature Selection Training Set

It is customary when training a feature-based description of a dataset to train the system using a subset of the textures in the dataset and to test the system using a different subset. As we shall be using a browsing environment comparison experiment to evaluate the effectiveness of the feature-based data organisation produced here, and we shall be using the previously defined (section 6.2.3) *Tex500[024]* as the stimuli for that experiment, it seems a reasonable approach to use the *Tex500[476]* subset for training the system and *Tex500[024]* as the test subset.

8.3.3 Optimise Number of Features

If no limit was placed on the number of features used to model each dimension our system would suffer from over-fitting, in other words we would continue selecting features until a near exact representation of perceptual space would be modelled in feature space. To avoid over-fitting, we should implement some means of limiting the number of features per dimension such that there would be little additional benefit to introducing further features. We can do this using Procrustes Analysis - checking the alignment of the feature matrix with the perceptual matrix, and stopping when the alignment error reduction begins to diminish.

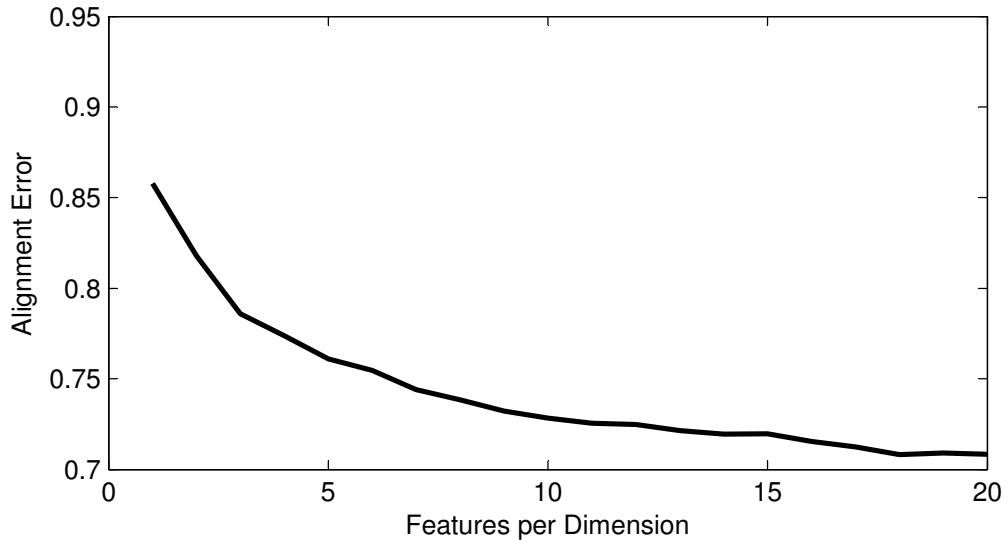


Figure 8.1: Alignment Error Analysis Plot

Figure 8.1 shows the alignment error for the first twenty features per dimension when comparing our perceptual matrix with the resulting feature-space matrices. We can see by inspection that the alignment error reduction begins to diminish after about nine features per dimension. We can therefore confidently limit our feature-space model to nine features per dimension.

8.3.4 Build SOM_F Browsing Environment

Having used Emrith’s approach to select nine Trace Transform features per dimension for the *Tex500*[476] training subset, we must complete the process by applying these features to the remaining test subset of the *Tex500* dataset (*Tex500*[024]). These two subsets can then be reunited in a similarity matrix in feature space for use as a data organisation on which we can build a feature-based browsing environment.

The resulting browsing environment is known as SOM_F and the top level grid screen can be seen in figure 8.2. Initial inspection of this top level view reveals that the texture space appears fairly uniformly sampled across the extent of the grid and that similarity/proximity relationship seems to hold for the exemplar textures of the neurons. However, more in-depth inspection of the neuron contents where the *Tex500*[024] stimuli can be found reveals a more varied level of similarity of textures within these neurons, and in some cases, questionable representativeness of the exemplar texture to the neuron contents. Figure 8.3 shows an example where there is apparently good group similarity and an intuitive exemplar whereas figure 8.4 shows an example where there is apparently poor group similarity and an unintuitive exemplar. Bearing in mind that this data organisation is formed from computational

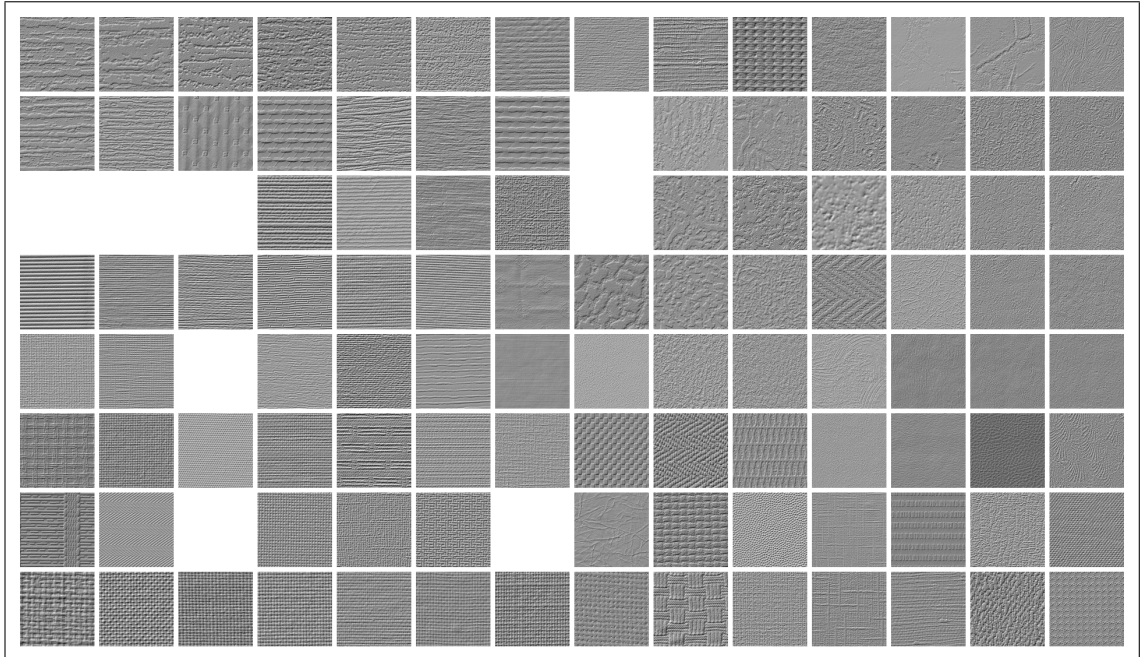


Figure 8.2: SOM_F Browsing Environment Top Level Screen Shot

features and that generally computational features do not take into account long range interactions in the textures, we may hypothesise that the strong long range interactions in the query texture in figure 8.4 might account for its apparently poor placement in the data organisation we have developed.

A full set of results for the *Tex500[024]* stimuli in a similar format to figures 8.4 and 8.3 can be found in appendix B.3.

8.4 Conclusions

In this chapter we have précied the work of Emrith [25] in setting criteria for developing a large set of computational features, identifying a number of candidate classes of features, and summarising the eligibility of those classes against the criteria. We went on to apply Emrith’s approach in feature selection to our *Tex500* dataset in order to produce a feature-based data organisation with which we can arrange our dataset within a browsing environment. After building our SOM_F browsing environment we made some initial analysis of the data organisation resulting from applying the feature selection approach and formed the hypothesis that some members of the *Tex500[024]* stimuli set that we shall later use to test the feature selection approach may, particularly in the case of stimuli with strong long-range interactions, prove unintuitive for observers to find within the data organisation generated. We therefore formed the hypothesis shown in equation 8.2 (with respect to mean task

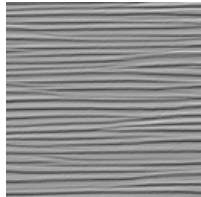
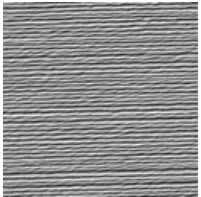
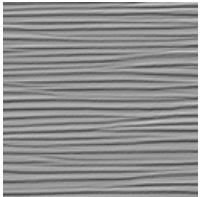
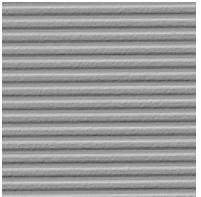
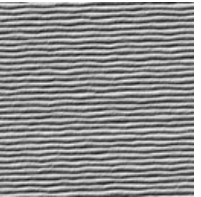
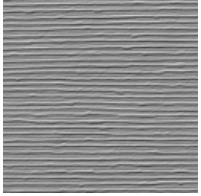
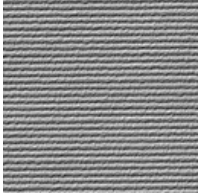
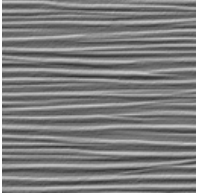
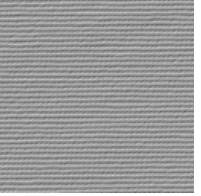
Stimuli	Neuron Contents (Starting with Neuron Exemplar)			
				
399	236	399	362	232
				
	287	194	400	306

Figure 8.3: Neuron from SOM_F with Similar Members and Intuitive Exemplar

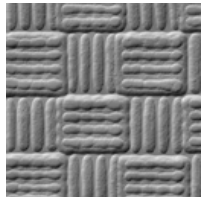
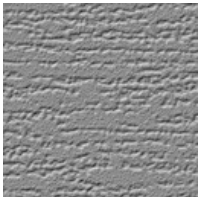
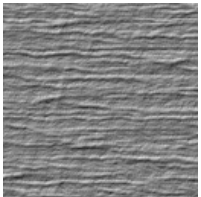
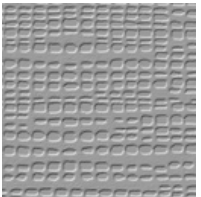
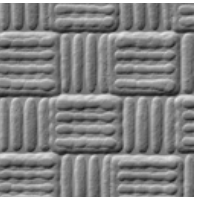
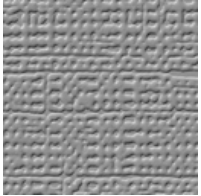
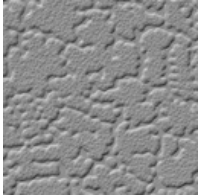
Stimuli	Neuron Contents (Starting with Neuron Exemplar)			
				
022	375	429	360	022
				
	447	376		

Figure 8.4: Neuron from SOM_F with Dissimilar Members and Unintuitive Exemplar

time), which we shall test in the browsing environment comparison experiment described in chapter 9.

$$\text{Hypothesis: } \text{SOM}_G < \text{SOM}_F \quad (8.2)$$

Chapter 9

SOM_G, SOM_A & SOM_F Browsing Environments Comparison

9.1 Introduction

In this chapter we describe the experiment used to test the performance of the SOM_G, SOM_A and SOM_F browsing environments introduced in chapters 5, 7 and 8 respectively. Table 9.1 gives a summary of the data organisations on which these are based.

Having extensively used each interface to browse the dataset, the author found that the data organisation in SOM_A was largely as plausible as the data organisation in SOM_G while the data organisation in SOM_F caused that browser to be much less intuitive to navigate. This informed the hypothesis shown in equation 9.1 (with respect to mean task time).

$$\text{Hypothesis: } \text{SOM}_A \approx \text{SOM}_G < \text{MDS}_F \quad (9.1)$$

Browser	Underlying Data Organisation
SOM _G	Perceptual similarity data for <i>Tex500</i> obtained using our initial grouping experiment (chapter 3)
SOM _A	Perceptual similarity data for <i>Tex500</i> [476] obtained using our initial grouping experiment (chapter 3) augmented using the pilot dataset augmentation experiment (chapter 7)
SOM _F	Computational feature-based representation of the dataset (chapter 8)

Table 9.1: Summary of SOM_G, SOM_A & SOM_F Underlying Data Organisations

9.2 Experimental Design

The purpose of our experiment is to assess the efficiency of each of the browsing environments in question and to discover which, if any, is statistically the most efficient. We must therefore consider an appropriate task, stimuli and experimental approach to make that assessment.

9.2.1 Task Specification

As we wish to assess the efficiency of various browsing environments, we must model the kind of task that would be a core reason for using such a browsing environment in the first place. It would therefore seem appropriate to ask users to find a particular texture from the dataset. The task stimulus could be a description of the surface to be found but this would be open to the interpretation of individual observers and it would be difficult to assess whether the observer had arrived at the correct, or at least a plausible selection. We therefore decided that the task stimulus, or query, should be an actual surface from the dataset that is present in the browsing environment.

9.2.2 Interface Presentation

As observers would have a range of cognitive abilities in relation to the task we decided it best to ask all observers test all three interfaces. To eliminate any learning effects from the experiment the order of presentation would be randomised.

9.2.3 Stimuli Selection

For the reasons detailed in section 6.2.3 of chapter 6, and to maintain consistency between browsing environment comparison experiments, the same stimuli set, *Tex500[024]*, will be used here.

9.2.4 Sample Size

As we have already discussed, twelve observers would each carry out eight trials on each browsing interface, giving a total of ninety-six trials per browsing environment. Although we can never be certain from the outset that this is sufficient to reach a statistically significant conclusion, it does seem like a fair sample size.

9.2.5 Performance Measurement

In contrast to the previous browsing environment comparison experiment (see section 6.2.5), we have elected only to measure the mean task time. Accuracy has been incorporated in this measurement by restricting observers to selecting exact matches only. If an incorrect selection is made they are informed of this and asked to continue searching. As only the organisation of the data within each SOM browsing environment varies in this experiment, and as browsers were presented in blind random order, the observers could not be asked to identify their preferred browsing environment.

9.2.6 Observer Selection

Volunteers were invited from the student population of the School of Mathematical and Computer Sciences at Heriot Watt University. No age or other restrictions were placed on volunteers. Observers were informed they would have to use a mouse and VDU and that they should make sure if they needed to use corrective eyewear that they wore this for the experiment. No age data was collected from observers but in the opinion of the author they all fell within the 18–35 years age group.

9.2.7 Instructions to Observers

Prior to taking part in the experiment each observer was asked to complete a consent form with their name, address and email address and to sign an agreement in the following terms.

I have agreed to take part in an experiment on texture browsing environments. The procedure has been explained to me and I understand that I am free to leave the experiment at any time.

The same form continued with the following instructions which were read to the observer by the author.

1. You will be presented with twenty-four query textures (left screen)
2. Your task is to find an exact match using a grid browser (right screen)
3. Each texture in the grid browser represents a group of one or more similar textures. Click on a texture in the grid to see the group of similar textures. Click on the ‘up’ icon to return to the grid

4. Double click a texture at the group level to select your answer. It will appear in the answer box along with whether your selection is correct or not. If the selection is incorrect, continue browsing until you find the correct texture
5. After a correct selection, once you are ready, click ‘continue’ to see your next query
6. Note that for each query texture, the grid browser layout may change
7. After clicking ‘continue’ to see the next query texture, you should try to locate it in the grid browser as quickly as possible

The following description and an author-lead demonstration was given of each of the interfaces at point of use.

This interface has a top level map where each grid square represents a collection of one or more textures. You should notice that similar collections of textures will be close to each other on this grid. Click on any grid square to see the member textures [author clicks top left texture]. To return to the grid, click on the arrow [author clicks arrow]. Double click a texture to select it in the detail level [author demonstrates].

9.3 Analysis of Result

9.3.1 Mean Task Time

Figure 9.1 shows the mean task time for each of the three data organisations. These were 73, 94 and 179 seconds respectively. Standard error bars are also indicated on the plot, the standard error ($\sigma_{\bar{x}}$) being calculated as shown in equation 6.2 where s is the sample standard deviation and N is the sample size. The intervals are plotted as shown in equations 6.3 and 6.4. The complete results of this experiment can be found in Appendix B.4.

9.3.2 Statistical Significance and Experimental Effect

As with the previous browsing environment comparison experiment, we cannot draw any conclusions from the differences in the mean task time until we test this for statistical significance. The statistical test we use is called the *dependent t-test*, which is calculated from the mean difference between our samples (\bar{D}) and the standard error of differences ($\sigma_{\bar{D}}$) as shown in equation 6.5.

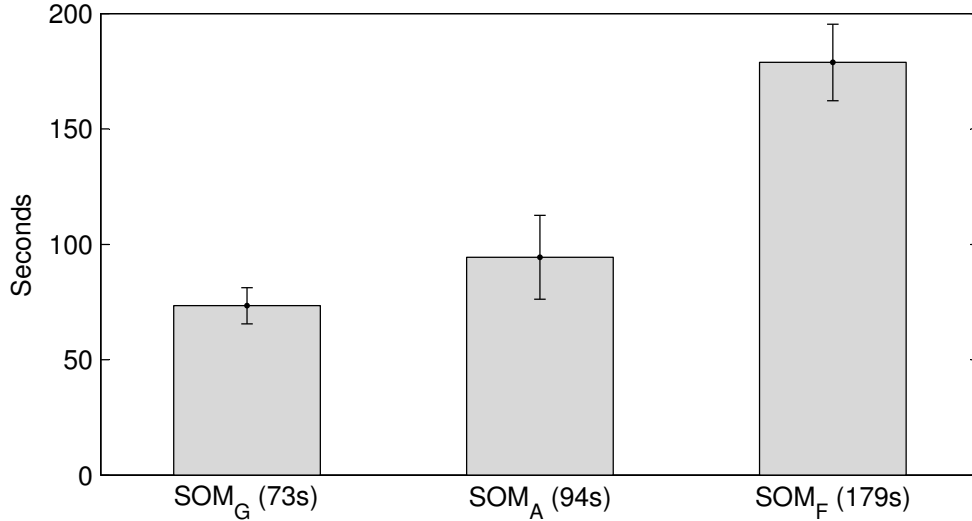


Figure 9.1: Mean Task Time (with Standard Error Bars) from SOM_G, SOM_A & SOM_F Browsing Environments Comparison Experiment

Interfaces Compared	t-test	p-value	Stat. Sig.	r-value	df
SOM _G vs. SOM _A	1.02	0.33	N	0.29	11
SOM _G vs. SOM _F	6.69	0.00	Y	0.90	11
SOM _A vs. SOM _F	3.92	0.00	Y	0.76	11

Table 9.2: Statistical Analysis of SOM_G, SOM_A & SOM_F Mean Task Time

We estimate the standard error of differences from the standard deviation of differences obtained within the sample (s_D) and the sample size (N) as shown in equation 6.6.

Once we have a value for t we can use this to calculate the effect size, r , as shown in equation 6.7. df denotes degrees of freedom, which is derived from the sample sizes N as shown in equation 6.8.

The analysis of our experimental results is summarised in table 9.2. As we are interested in an overall result reflecting 95% confidence (a p value of 0.05), and we are comparing three pairs of browsing environments, we must achieve a p value for each comparison of $p < 1 - \sqrt[3]{0.95}$ ($p < 0.01695$). As this p condition is achieved for SOM_G over SOM_F and SOM_A over SOM_F, but not SOM_G over SOM_A we can say that SOM_G and SOM_A are statistically equivalent and that both are statistically more efficient than SOM_F. In terms of the r-values, by cross referencing with table 9.3, we can also see that there is a large effect size of the SOM_G and SOM_A over SOM_F, indicating that the differences in mean task time was due largely to the differences in the respective data organisations, rather than variance within the sample.

For detailed experimental results please see appendix B.4.

r-value	Effect
0.1	Small
0.3	Medium
0.5	Large

Table 9.3: Experimental Effect Sizes

9.4 Conclusions

Our experiment revealed that the SOM_G and SOM_A browsing environments were statistically more efficient than the MDS_F browsing environment. This tells us that the features proposed by Emrith [25], which we used in our feature-based data organisation produced in chapter 8, do not sufficiently model human perception to be useful in organising textures in browsing environments. A possible reason for this is the failure of such features to account for long range interactions in the surfaces, almost certainly an important aspect of human discrimination of texture.

The experiment also revealed there was no statistical significance to the differences between mean task time in SOM_G and SOM_A . This suggests that the augmentation approach proposed in chapter 7 has merit and may be able to produce a data organisation for a SOM browsing environment that is as intuitive to users as the costly perceptual data obtained from the initial grouping experiments in chapter 3. We will examine the possibility of producing a scalable augmentation approach based on this pilot in chapter 10.

Chapter 10

Scalable Dataset Augmentation

10.1 Introduction

In chapter 7 we proposed a method of augmenting an existing dataset with unknown data members and proved this approach to be sound by conducting a browsing environment efficiency experiment (chapter 9). The augmentation approach was tested by removing twenty-four textures from our dataset (specified in appendix A.1), then using our technique to reintroduce these ‘unknown’ textures into the dataset. Although this appeared to be a fair sample to perform an initial proof of concept, the difficulties discussed in chapter 3 around obtaining perceptual similarity data using grouping experiments can only be solved by significantly reducing the number of textures we ask observers to sort.

As our data augmentation technique has a relatively low time overhead for each new texture presented to users, the key to making the approach scalable is to find a large pool of observers who would willingly analyse a relatively small number of textures over a short experimental session. If sufficiently motivated, individuals may be persuaded to take part in multiple experimental sessions. There already exist a variety of web based communities for exactly this purpose. Researchers feature tasks they would like completed in return for a small fee. Observers browse these tasks and choose to participate where they feel suitably interested, qualified, or remunerated. The collective term for such networks is *crowdsourcing* and this has proven useful in recruiting observers for a variety of research experiments. Crowdsourcing does not, however, come without its problems. In this chapter we discuss in detail how we designed a series of experiments to test the hypothesis that crowdsourcing could be harnessed to augment a small dataset, bootstrapped using a grouping experiment, with a large number of previously unknown textures. In chapter 11 we subject our

scalable dataset augmentation experiment to robust statistical analysis by conducting a browsing environment comparison experiment.

We utilised the Amazon Mechanical Turk [4] crowdsourcing community to employ a large pool of observers for the purpose of augmenting our dataset with textures unknown at the initial stage of bootstrapping.

10.2 New Grouping Experiment

In order to test the scalability of our augmentation approach, we must set realistic parameters for the number of textures in the initial set and the number of unknown textures to be inserted using the augmentation technique. In our pilot augmentation experiment (chapter 7) we removed the *Tex500[024]* stimuli from the dataset and used a perceptual similarity experiment and simple augmentation algorithm to introduce the stimuli to the dataset. Although this was sufficient to provide proof of concept, these figures would be unrealistic were we faced with the challenge of obtaining a perceptual description of our dataset without experiencing the problems associated with large grouping experiments.

As experience tells us that the average observer can easily perform a sorting task on one-hundred textures in under an hour, we select this as our initial dataset size. If we were starting this task with no previous perceptual description of the dataset then we must reproduce this limitation here by drawing these one-hundred textures randomly from the *Tex500* dataset (as opposed to some representative cross section). This randomly drawn subset is known as the *Tex500[100]* stimuli and can be seen in full in appendix A.2.

As we wish to test this approach from end to end without the prior knowledge about the dataset we already have, and because an observer's perception of similarity may be affected by the range of textures they are exposed to, we elected to conduct a brand new similarity grouping experiment using 14 observers (matching previous grouping resolution) rather than sub-sampling our original similarity matrix.

Figure 10.1 illustrates the dendrogram constructed from our new grouping experiment on the *Tex500[100]* data subset. By inspecting this in comparison with figure 10.2, illustrating the dendrogram constructed from a similarity matrix sub-sampled from our original grouping experiment similarity matrix, we can observe some broad similarities as detailed in table 10.1. This goes some way to reassuring us that given sufficient observers, the average human perception of similarity in texture is broadly consistent.

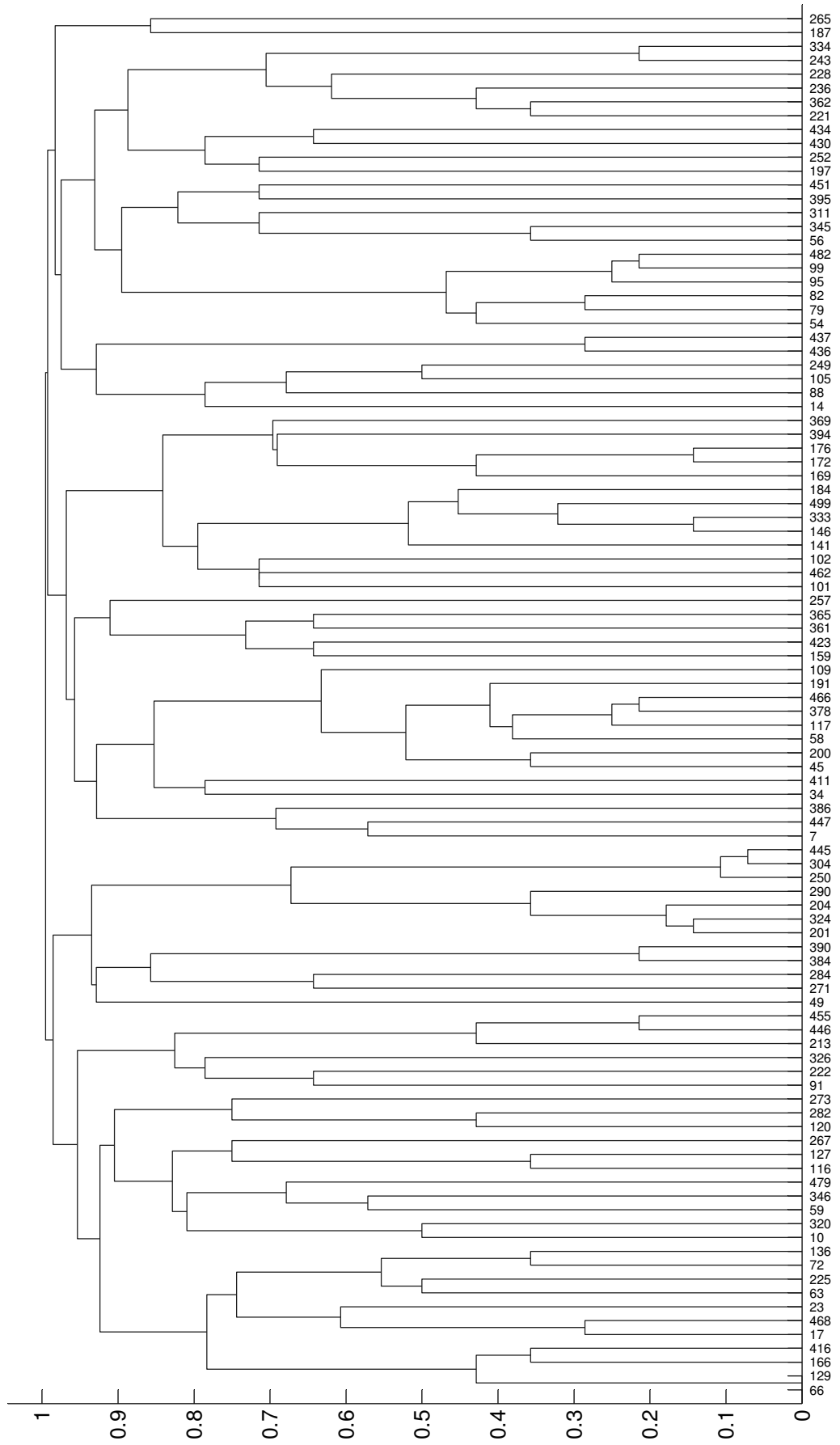


Figure 10.1: Dendrogram from Grouping Experiment on *Tex500*[100]

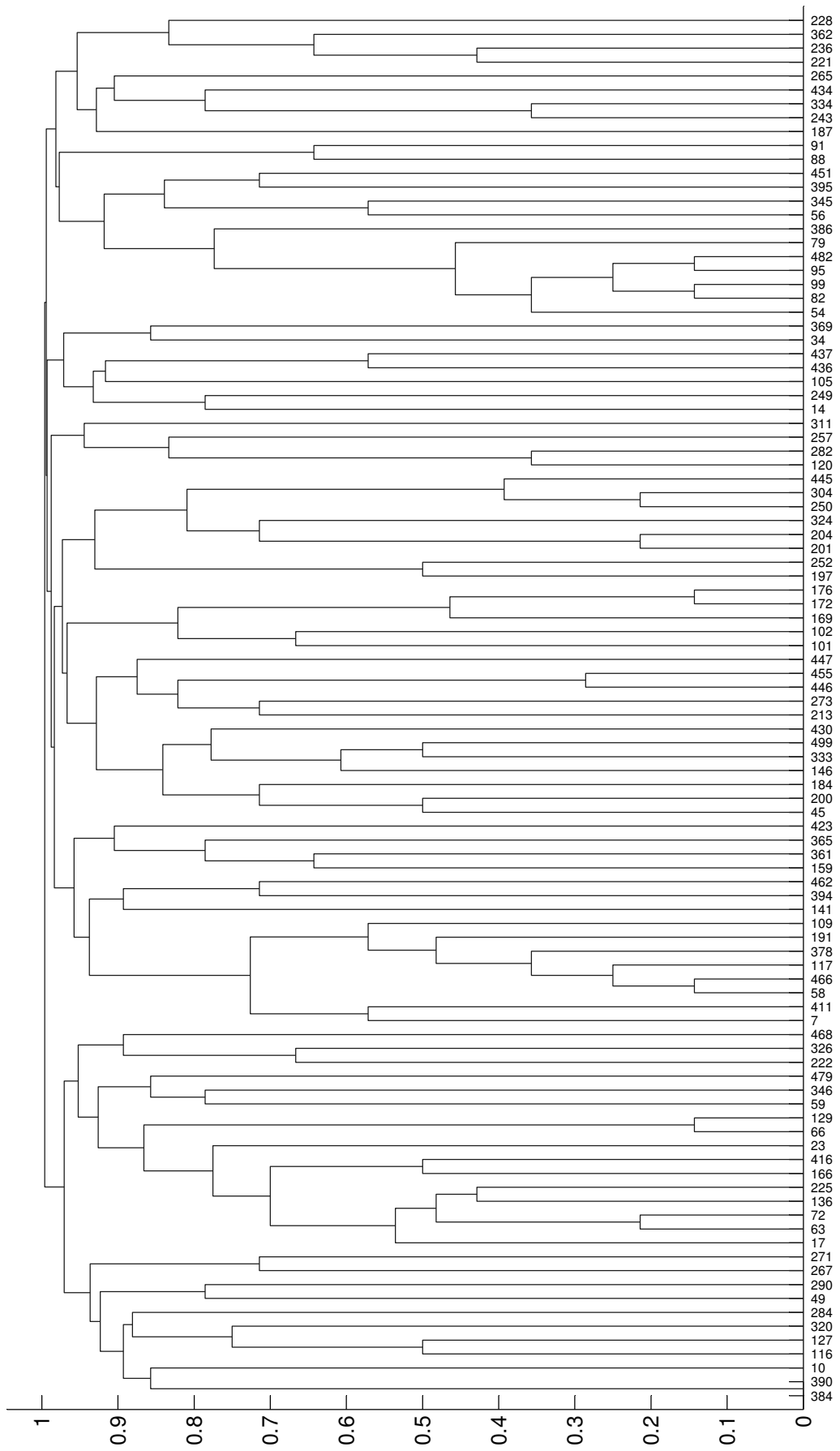


Figure 10.2: Dendrogram sub-sampled from *Tex500* Grouping Experiment

Textures	New Dendrogram	Original Dendrogram
66, 129	Pair at height zero	Pair at low merge height
384, 390	Pair at low merge height	Pair at height zero
172, 176	Pair at low merge height	Pair at low merge height
58, 117, 378, 466	Cluster at low merge height	Cluster at low merge height
250, 304, 445	Cluster at low merge height	Cluster at low merge height

Table 10.1: Comparison of *Tex500[100]* and sub-sampled *Tex500* Dendrograms

10.3 Design Data Augmentation Experiment

10.3.1 Stimuli Selection

The stimuli for the data augmentation experiment is the four-hundred textures formed by the compliment of the *Tex500[100]* subset used in our new grouping experiment. This is labelled *Tex500[400]*. As we want 10 observations per stimulus we generate ten separate random permutations of these textures which we partition into two-hundred experimental sessions, each consisting of 20 stimuli. This ensures that no stimuli set contains two identical stimuli while maintaining random presentation and separation.

10.3.2 Design SOM_{R2} Browsing Environment

In section 6.2.5, we concluded that the SOM_G browsing environment was the most efficient of those tested. We will therefore follow the design described in section 5.4 to build a SOM browsing environment using the similarity matrix produced by our new grouping experiment on *Tex500[100]*. This browsing environment is known as SOM_{R2}.

10.3.3 Task Specification

For each stimulus, observers will be asked to use the SOM_{R2} browsing environment to select those textures they consider to be most similar to the stimulus. They must select between two and four textures before continuing to the next trial. They record their response by adding similar textures to a palette from which textures may also be removed in the event they discover a more suitable selection. Once they have submitted their response to a trial, no changes can be made. Figure 10.3 shows a screen capture of the interface presented to observers.

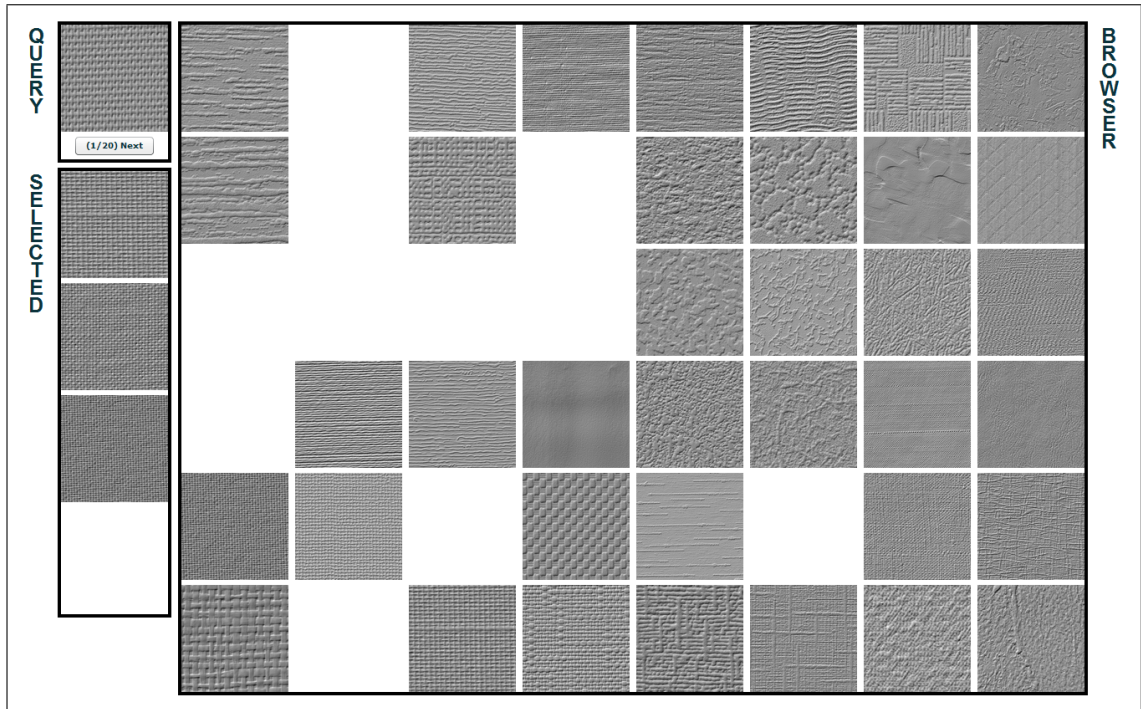


Figure 10.3: SOM_{R2} Browsing Environment Screen Capture

10.3.4 Sample Size

In *MTurk*, observer sessions are known as Human Interface Tasks (HITs). Two hundred of these will be offered, each one representing a similarity experiment on twenty randomly ordered stimuli. Owing to the large user base of a crowdsourcing platform like *MTurk* these HITs are consumed rapidly leaving no opportunity for the same observer to carry out large numbers of sessions. For this reason, and because there are no negative learning effects in this experiment, we decided not to restrict the number of SOM_{R2} HITs account holders were permitted to engage in. Where an experiment is not completed or where the response does not qualify for payment, the stimuli set is made available to future users until a complete and satisfactory response is received for each.

10.3.5 Payment Criteria

It is well known that some users in crowdsourcing networks will give ill considered responses to tasks in order to proceed as quickly as possible to payment, and thereby earn participation fees for little or no effort. In order to protect our valuable data from corruption and to prevent wasting our limited research budget, we employed a two tier payment authorisation algorithm. The criteria for payment were supplied to users at the start of each experimental session and all session data tested before

claims for payment authorised. As we had the benefit of our original perceptual similarity matrix we were able to use this to assess whether or not the crowdsourcing responses were plausible. Of course this would not be available to us had we not already conducted the grouping experiments detailed in chapter 3. In subsection 10.3.5 we discuss possible solutions where previous perceptual data is not available.

It was made clear to observers that with reasonable consideration of the task, meeting the minimum criteria AND the bonus criterion should be easily achievable.

Minimum Payment

The minimum payment was set at 0.75 USD and to qualify,

- the mean similarity value of all selected textures must be higher than the mean similarity value of all non-zero similarity textures in SOM_{R2} for each stimulus presented, and
- no more than 20% of all selected textures may have a zero similarity with their respective stimulus.

Bonus Payment

A bonus payment of 0.25 USD will be earned if an observer makes an average of three or more selections per query texture. The minimum is two and the maximum if four.

No Previous Data Held

If we were not in possession of previous perceptual data captured in chapter 3, we might have considered using either of the following two approaches to enforce quality:

- increase the size of our bootstrapping grouping experiment, randomly select a number of these additional textures for inclusion in each HIT, and check that all of these ‘test’ stimuli obtain a plausible response before authorising payment, or
- if we assume that the majority of observers wish to give a genuine response (we found this was the case), then we could test for sessions that represent high deviation from typical response and exclude these from payment. This approach necessitates completing the whole experiment before processing claims for payment.

10.3.6 Observer Selection

Volunteers were invited from the membership of the Amazon Mechanical Turk (MTurk) [4] crowd-sourcing community. No age or other restrictions were placed on volunteers. Observers were informed they would have to use their mouse and VDU and that they should make sure if they needed to use corrective eyewear that they wore this for the experiment. No age or other qualifying data was collected from observers.

10.3.7 Instructions to Observers

Prior to taking part in the experiment each observer was asked to indicate implied consent by clicking ‘OK’ on a series of instruction/information screens built in to the Adobe Flex application deployed to run the experiment. The first of these explained the task the observer would be required to undertake:

Instructions - please read very carefully

You will be presented with 20 query textures (labelled QUERY). You need to use the BROWSER to find the most similar textures to the query. You will need to select between 2 and 4 similar textures to continue to the next trial.

In the browser (labelled BROWSER), each texture image represents a group of 1 to 8 similar textures. Click on a texture to see the entire group in a pop up window. To select a texture from the pop up window, simply click on the texture. It will be added to your selection (labelled SELECTED).

You can remove a texture from your selection by clicking on the texture. After you have selected at least two textures, a ‘NEXT’ button will appear, but you should continue adding similar textures if they can be found in the browser. Once you hit ‘NEXT’ your selection will be logged and your next query texture shown.

DO NOT USE BROWSER NAVIGATION OR REFRESH AS YOUR DATA WILL BE LOST!

To impress upon observers the importance of giving reliable observations of perceptual similarity throughout the experiment, an additional message about payment criteria was displayed. It was hoped that the strength of the wording here would deter those observers who intentionally click on any stimuli to try and gain payment. If we were

not in a position to check the data for reliability (i.e. we had no previous perceptual data to test against) it may well be worth still making this statement as a deterrent.

IMPORTANT: Payment Criteria

We already hold high resolution similarity data on the dataset you are about to be tested on. We will check that your similarity judgements meet a minimum threshold before authorising payment. Do not attempt this experiment if you intend to rely on random chance as the threshold is much higher.

The following criteria must be met for payment of 0.75 USD to be authorised:

1. Your mean similarity value must be higher than the mean similarity value of all possible non-zero similarity selections in the browser for each stimuli presented
2. No more than 20% of your selections can have a zero similarity with the stimulus

A bonus payment of 0.25 USD will be applied if you make an average of 2.5 selections per stimulus rather than the minimum of 2 (but the above minimum criteria still apply).

If you intend to take the task seriously you will find that meeting the minimum criteria AND the bonus criterion are easily achievable, indeed we expect most observers to do so.

We also encourage you to return and participate in our experiment again as we have a very large number of stimuli to experiment with.

At the end of your experiment you will be given a token which you must send to us via MTurk to claim payment. We will endeavour to authorise payment for good quality submissions on the next working day after taking part.

Thank you for your interest and enjoy the experiment!

Failing to click 'OK' on either of the above messages would stop the experiment and return the observer to their account page. After completion of the experiment the observer was presented with the following message and the payment token code could be copied and pasted into their claim for payment.

Session completed

Thank you for taking part.

Please copy and paste this token to claim your payment through mturk:
<TOKEN CODE>

Your payment will be authorised once the quality of your submission has been checked by our processing system.

10.3.8 Completion Time

An initial pilot experiment consisting of 10 HITs were launched to test the experimental apparatus and result logging server. This took minutes to complete. We then offered the remaining 190 experimental sessions to the community and found that this resolved in 2-3 hours. After removing spoiled or incomplete HITs or HITs which did not achieve the minimum criteria for payment, a final offering was made to conclude the experiment. Again this small number of HITs were completed in a few minutes.

The time taken to complete this experiment was a fraction of the time taken to run experiments where observers are required to attend the research lab. It also represents a deal of time saving on the part of the researcher as, even with the overhead of producing a robust online experiment environment, the experiment can be run without any significant further intervention.

10.4 Augmentation Experiment Data Analysis

10.4.1 Votes

The outcome of the similarity perception experiment is an array of similar texture votes for each stimulus. Selected results from the crowdsourcing experiment can be found in appendix B.5.1. An example of an observer response that did not meet the minimum criteria for payment can be found in appendix B.5.2.

10.4.2 Build SOM_{A2} Browsing Environment

Figure 10.4 shows the dendrogram after augmenting the dataset using the crowdsourcing experiment. We can clearly see that the structure contrasts with the structure of the dendrogram from our original perceptual grouping experiment (figure 4.1). This may be explained by the projection of a large (*Tex500[400]*) texture set onto the relatively compact texture space representing the *Tex500[100]* stimuli.

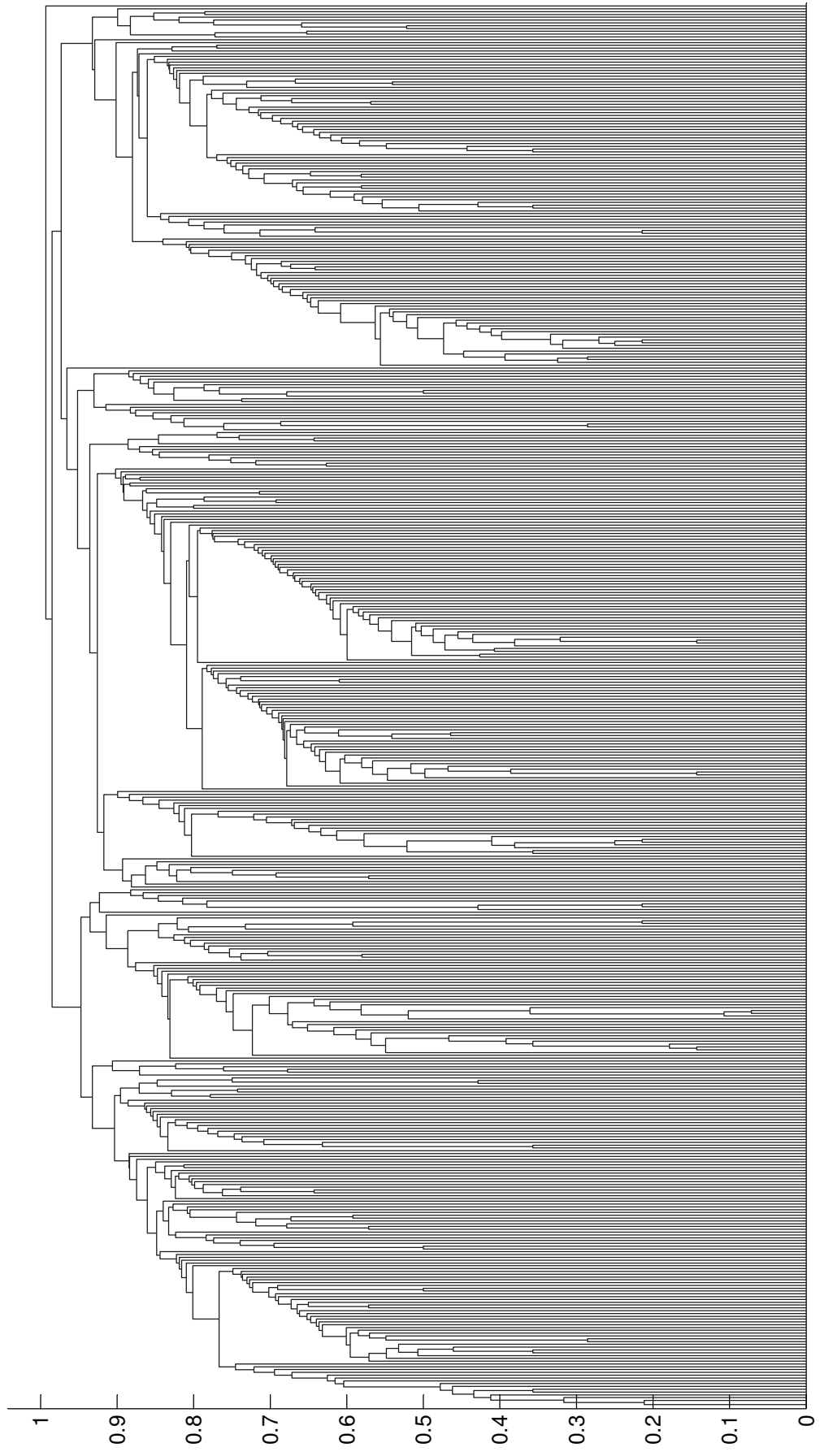


Figure 10.4: Crowdsourcing Augmented *Tex500[100]* Dendrogram

By applying our algorithm (see equation 7.4) to the result of the similarity perception experiment we obtain a new similarity matrix which we use as the data organisation for a new browsing environment. The browsing environment built on this augmented data organisation is referred to as SOM_{A2} .

10.5 Conclusions

In this chapter we described a scalable dataset augmentation experiment based on our pilot experiment in chapter 7. We built on our proof of concept by significantly adjusting the ratio of bootstrap dataset size to augmentation stimuli set size from 476:24 to 100:400. Again, we used the new data organisation to build an augmented browsing environment SOM_{A2} .

By using the SOM_{A2} browsing environment to browse the dataset, the author formed the opinion that the data organisation was logical and did not represent an inferior experience to the SOM_G browsing environment. He therefore formed the hypothesis shown in equation 10.1.

$$\text{Hypothesis: } SOM_{A2} \approx SOM_G \quad (10.1)$$

This hypothesis is tested in a browsing environment comparison experiment in chapter 11.

Chapter 11

SOM_G & SOM_{A2} Browsing Environments Comparison

11.1 Introduction

In this chapter we describe the experiment used to test the performance of the SOM_G and SOM_{A2} browsing environments introduced in chapters 5 and 10 respectively. Table 11.1 gives a summary of the data organisations on which these are based.

Having extensively used each interface to browse the dataset, the author found that the data organisation in SOM_{A2} was largely as plausible as the data organisation in SOM_G. This informed the hypothesis shown in equation 11.1 (with respect to mean task time).

$$\text{Hypothesis: } \text{SOM}_{A2} \approx \text{SOM}_G \quad (11.1)$$

Browser	Underlying Data Organisation
SOM _G	Perceptual similarity data for <i>Tex500</i> obtained using our initial grouping experiment (chapter 3)
SOM _{A2}	Perceptual similarity data for <i>Tex500[100]</i> obtained using our bootstrapping grouping experiment (section 10.2) augmented using the scalable dataset augmentation experiment (chapter 10)

Table 11.1: Summary of SOM_G & SOM_{A2} Underlying Data Organisations

11.2 Experimental Design

The purpose of our experiment is to assess the efficiency of each of the browsing environments in question and to discover which, if any, is statistically the most efficient. We must therefore consider an appropriate task, stimuli and experimental approach to make that assessment.

11.2.1 Task Specification

As we wish to assess the efficiency of various browsing environments, we must model the kind of task that would be a core reason for using such a browsing environment in the first place. It would therefore seem appropriate to ask users to find a particular texture from the dataset. The task stimulus could be a description of the surface to be found but this would be open to the interpretation of individual observers and it would be difficult to assess whether the observer had arrived at the correct, or at least a plausible selection. We therefore decided that the task stimulus, or query, should be an actual surface from the dataset that is present in the browsing environment.

11.2.2 Interface Presentation

As observers would have a range of cognitive abilities in relation to the task we decided it best to ask all observers test both interfaces. To eliminate any learning effects from the experiment the order of presentation would be randomised.

11.2.3 Stimuli Selection

For the reasons detailed in section 6.2.3 of chapter 6, and to maintain consistency between browsing environment comparison experiments, the same stimuli set, *Tex500[024]*, will be used here. As this experiment compares two browsing environments rather than three, each observer will experience twelve trials for each browsing environment rather than the eight in chapters 6 and 9.

11.2.4 Sample Size

As we have already discussed, twelve observers would each carry out twelve trials on each browsing interface, giving a total of one hundred and forty-four trials per

browsing environment. Although we can never be certain from the outset that this is sufficient to reach a statistically significant conclusion, it does seem like a fair sample size.

11.2.5 Performance Measurement

For the same reasons given in the previous browsing environment comparison experiment (see section 9.2.5), we again only measure mean task time.

11.2.6 Observer Selection

Volunteers were invited from the student population of the School of Mathematical and Computer Sciences at Heriot Watt University. No age or other restrictions were placed on volunteers. Observers were informed they would have to use a mouse and VDU and that they should make sure if they needed to use corrective eyewear that they wore this for the experiment. No age data was collected from observers but in the opinion of the author they all fell within the 18–35 years age group.

11.2.7 Instructions to Observers

Prior to taking part in the experiment each observer was asked to complete a consent form with their name, address and email address and to sign an agreement in the following terms.

I have agreed to take part in an experiment on texture browsing environments. The procedure has been explained to me and I understand that I am free to leave the experiment at any time.

The same form continued with the following instructions which were read to the observer by the author.

1. You will be presented with twenty-four query textures (left screen)
2. Your task is to find an exact match using a grid browser (right screen)
3. Each texture in the grid browser represents a group of one or more similar textures. Click on a texture in the grid to see the group of similar textures. Click on the ‘up’ icon to return to the grid
4. Double click a texture at the group level to select your answer. It will appear in the answer box along with whether your selection is correct or not. If the

selection is incorrect, continue browsing until you find the correct texture

5. After a correct selection, once you are ready, click ‘continue’ to see your next query
6. Note that for each query texture, the grid browser layout may change
7. After clicking ‘continue’ to see the next query texture, you should try to locate it in the grid browser as quickly as possible

The following description and an author-lead demonstration was given of each of the interfaces at point of use.

This interface has a top level map where each grid square represents a collection of one or more textures. You should notice that similar collections of textures will be close to each other on this grid. Click on any grid square to see the member textures [author clicks top left texture]. To return to the grid, click on the arrow [author clicks arrow]. Double click a texture to select it in the detail level [author demonstrates].

11.3 Analysis of Result

11.3.1 Mean Task Time

Figure 11.1 shows the mean task time for each of the two data organisations. These were 65 and 82 seconds respectively. Standard error bars are also indicated on the plot, the standard error ($\sigma_{\bar{x}}$) being calculated as shown in equation 6.2 where s is the sample standard deviation and N is the sample size. The intervals are plotted as shown in equations 6.3 and 6.4.

11.3.2 Statistical Significance and Experimental Effect

As with the previous browsing environment comparison experiments, we cannot draw any conclusions from the differences in the mean task time until we test this for statistical significance. The statistical test we use is called the *dependent t-test*, which is calculated from the mean difference between our samples (\bar{D}) and the standard error of differences ($\sigma_{\bar{D}}$) as shown in equation 6.5.

We estimate the standard error of differences from the standard deviation of differences obtained within the sample (s_D) and the sample size (N) as shown in equation 6.6.

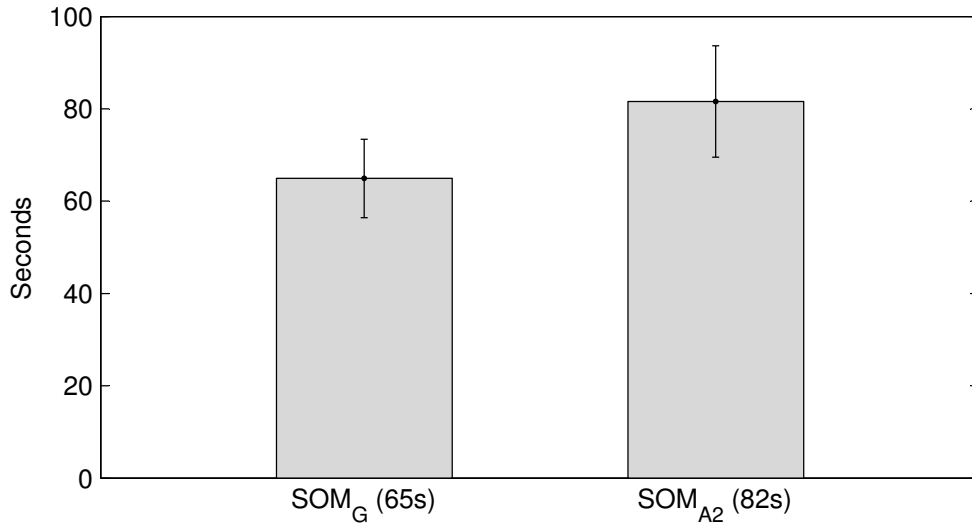


Figure 11.1: Mean Task Time (with Standard Error Bars) from SOM_G & SOM_{A2} Browsing Environments Comparison Experiment

Interfaces Compared	t-test	p-value	Stat. Sig.	r-value	df
SOM _G vs. SOM _{A2}	1.04	0.32	N	0.30	11

Table 11.2: Statistical Analysis of SOM_G & SOM_{A2} Mean Task Time

Once we have a value for t we can use this to calculate the effect size, r , as shown in equation 6.7. df denotes degrees of freedom, which is derived from the sample size N as shown in equation 6.8.

The analysis of our experimental results is summarised in table 11.2. As we are interested in an overall result reflecting 95% confidence, and we are comparing one pair of browsing environments, we must achieve a p value for the comparison of $p < 1 - 0.95$ ($p < 0.05$). As this p condition is not achieved for SOM_G over SOM_{A2} we can say there is no statistically significant difference in efficiency between them. In terms of the r-values, by cross referencing with table 11.3, we can also see that there is a medium effect size, indicating that the differences in mean task time was due largely to the differences in the respective data organisations, rather than variance within the sample.

For detailed experimental results please see appendix B.6.

r-value	Effect
0.1	Small
0.3	Medium
0.5	Large

Table 11.3: Experimental Effect Sizes

11.4 Conclusions

Our experiment revealed there was no statistical significance to the differences between mean task time in SOM_G and SOM_{A2} . This suggests that the scalable augmentation approach proposed in chapter 10 has merit and produced a data organisation for a SOM browsing environment that is as intuitive to users as the costly perceptual data obtained from the initial grouping experiments in chapter 3.

Chapter 12

Summary and Future Work

12.1 Summary of Thesis Argument

In section 1.2 we set out the goals of this research project. In addition to the objectives of producing a comprehensive literature review and capturing human similarity judgements and identifying structure, we also listed the following items, elements of which we argue in this chapter contribute to the state of the art in texture research:

1. Source or develop a large surface texture database
2. Develop and evaluate browsing environments
3. Investigate scalable methods of augmenting similarity matrices

Figure 12.1 illustrates the argument flow of this thesis. The solid rectangles represent chapters and feature itemised lists summarising the outcomes associated with the chapter. The chapters are linked by arrows representing the argument flow. These arrows are annotated with a summary of argument leading from one chapter to another. Chapters which have a strong association with other chapters are grouped within a dotted outline. The remainder of this chapter summarises the key products of this research and highlights important outcomes from the thesis argument.

Scalable Grouping Experiment [Ch. 3]

Observers who took part in the pilot grouping experiment on the *Tex500* dataset took on average three hours to complete the experiment. Due to mental and physical fatigue, risk to quality of result and difficulties recruiting observers for such an extended experiment, we had to alter the traditional grouping approach and design

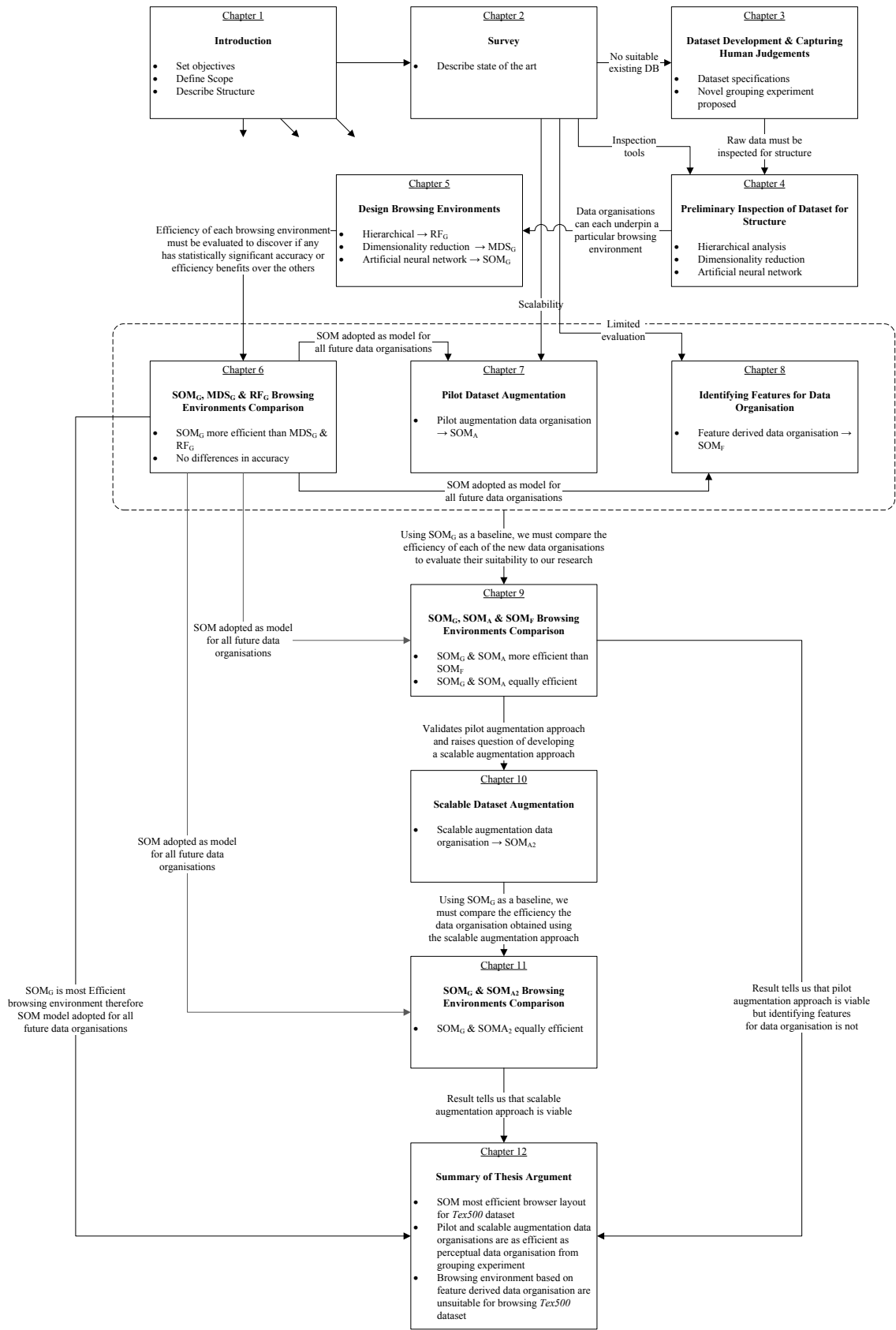


Figure 12.1: Summary of Thesis Argument

a more appropriate experiment. Our novel solution involved partitioning the dataset into three clusters using the perceptual data collected in the pilot experiment (six observers) and limiting each new observer to one of these three groups of textures. To combine the result two matrices were recorded. The opportunity matrix was incremented for each texture pair that an observer has an opportunity of pairing. The occurrence matrix was incremented for each texture pair that an observer placed in a group together. After the experiments were complete the occurrence matrix was divided by the opportunity matrix to produce a normalised similarity matrix in the range 0-1. We labelled the approach *pseudo*-scalable in recognition of the partitioning stage requiring a set of grouping results from experiments using the whole dataset. In this respect the approach is not strictly scalable but in our case as we had access to this data anyway, we were able to take advantage of it. Had the dataset been larger then we might have found it impossible for any observer to finish the grouping task.

An Efficient Browsing Model [Ch. 5 & 6]

Three browsing models were developed that organised the textures in 3D space, hierarchically and on a self organising map (SOM). These were tested for efficiency and accuracy using a psychophysical experiment where observers were asked to use each of the browsing environments to find a set of query textures. The mean task time was analysed for each browsing model and subjected to a *t-test*. We found that the SOM browsing model was statistically more efficient than the other two models but that the other two models did not differ with statistical significance so could not be ranked. In terms of accuracy we found no statistical significance in accuracy difference between models. The SOM browsing model was a novel approach to image/texture browsing by the author.

Novel Database Augmentation Technique [Ch. 7, 9, 10 & 11]

Given the difficulty in capturing human judgements about a large image/texture database using a traditional grouping experiment we had incentive to develop an approach where we could capture perceptual information about a subset of a database then use an alternative technique augmenting the database with the remaining textures. We approached this in two stages. Proof of concept was established with a pilot experiment where a database of 476 textures was augmented with 24 textures. A scalable experiment established that the technique was scalable when the balance was changed to augmenting a dataset of 100 textures with 400 textures. Furthermore,

we were able to establish that this approach could be successfully carried out using *crowdsourcing* at low cost and with a short completion time.

Trace Transform Features not Effective for Browsing Environments [Ch. 8 & 9]

We were fortunate to have access to a further means of generating a data organisation for us in browsing environments in the form of a feature bank and a feature selection approach. These were developed by Emrith in his PhD thesis [25]. We applied his technique to our dataset then compared a browsing environment based on computation features with one based on our perceptual data. We found that the difference in mean task time between the two data organisations was statistically significant and that the feature based data organisation was much less efficient than the perceptual organisation. Although we introduced no new research about computational features and feature selection, we were able to demonstrate that using computational features to model the data organisation for use in browsing environments gave poor results.

Contributions to the State of the Art [Ch. 3, 6, 9, 10 & 11]

Broadly speaking, the novelties of this research that we regard as contributing to the state of the art are detailed below.

Database Development: In our survey we identified the absence of any projects specifically concerned in browsing texture. In order to identify candidate databases for use in this project we searched other areas of texture research but failed to find any sufficiently large dataset that satisfied our selection criteria. Our development of a large dataset of surface textures (five hundred), together with perceptual similarity data, can be regarded as novel in the field of texture research as a dataset of this size, quality and consistency of capture conditions is unprecedented and can be of use to future texture research projects.

Rigorous Browser Comparison: Our rigorous browsing model comparison experiments identified that the SOM browsing model was the most efficient (in terms of mean task time) for browsing the *Tex500* dataset. Not only was the SOM browser itself somewhat novel, but to the knowledge of the author no other research has tested a variety of browsing environments in order to identify the most efficient.

Scalable Dataset Augmentation: Given the high cost of obtaining perceptual

similarity data about a large database of stimuli using lab based experiments, we believe the development of a scalable method of augmenting similarity matrices for browsing using crowdsourcing was a significant contribution to the state of the art in texture research.

12.2 Future Work

Here we detail possible future work that was deemed to be beyond the scope of this research project but which may be of interest to the community or represent potential additional contribution to the state of the art. Some of these points were known in advance of the research contained in this volume whilst others became evident as a result of conducting our research.

Dataset Size

The result we reported in respect of the most efficient model on which to base texture browsing environments was obtained using our *Tex500* database of 500 textures. Whilst it is inviting to believe this result might hold for significantly larger databases we feel that further experiments would need to be carried out to verify this. We tested our dataset augmentation techniques at two specific ratios. The first was 476 : 24 (476 known textures to 24 new textures) in a lab based experiment and 100 : 400 where similarity judgements were captured from members of a crowdsourcing community. Our tests revealed there were no statistically significant differences between browsing environments built using these data compared with our baseline browsing environment built using perceptual grouping data. However, no assumptions can be made about whether these results would hold for the same ratios applied to a significantly larger dataset, or indeed different ratios. Again additional experiments would be required to verify the efficacy of our approach as the dataset size increases.

Stimuli Type

Our browsing environment research was entirely based on images of surface texture as stimuli and we see no obvious reason why the results obtained should not hold for other forms of visual stimuli such as photographs, non-grayscale texture, images of 3D components, etc. Recently, Padilla [76], also in the Texture Lab, had success in repeating the author's techniques with a database of 500 thumbnails of abstract colour artwork. There is scope for additional investigation to be made into the

applicability of these techniques to alternative visual stimuli such as video clips, moving pictures and animations as well as non-visual stimuli such as sound, or any other stimuli where it is costly to obtain perceptual similarity data from human judgements.

Capture of Human Judgements

Our initial capture of human similarity judgements was through the use of perceptual grouping experiments. We also relied on this method to bootstrap our scalable augmentation (via crowdsourcing) approach. However, other methods, such as pairwise comparison, perceptual ordering, ratio scaling, etc. are available. Use of these methods may result in different findings and would need further investigation in the context of browsing environments. Indeed, some of these approaches may be suitable for deployment of the bootstrapping task to crowdsourcing environments, allowing the whole process to be completed without the need for any observers to attend lab-based experiments.

Applications

We used and tested our augmented data only in the context of perceptual browsing and navigation but it could conceivably be used and tested in other applications such as retrieval.

Observer Demographic

Further investigation would be required to verify that our experimental results held for alternative observer demographics. For example, the use of participants from other age groups, cultural backgrounds, cognitive abilities, etc.

Computational Features

Our use of the computational trace transform features proved to be unsuccessful for browsing and navigation. However, we may well have enjoyed greater success in applying this approach to other applications such as retrieval, similarity estimation, appearance measurement, etc. Likewise, other classes of computational features may well be more suitable to the application of browsing. Further investigation is required.

Application Specificity

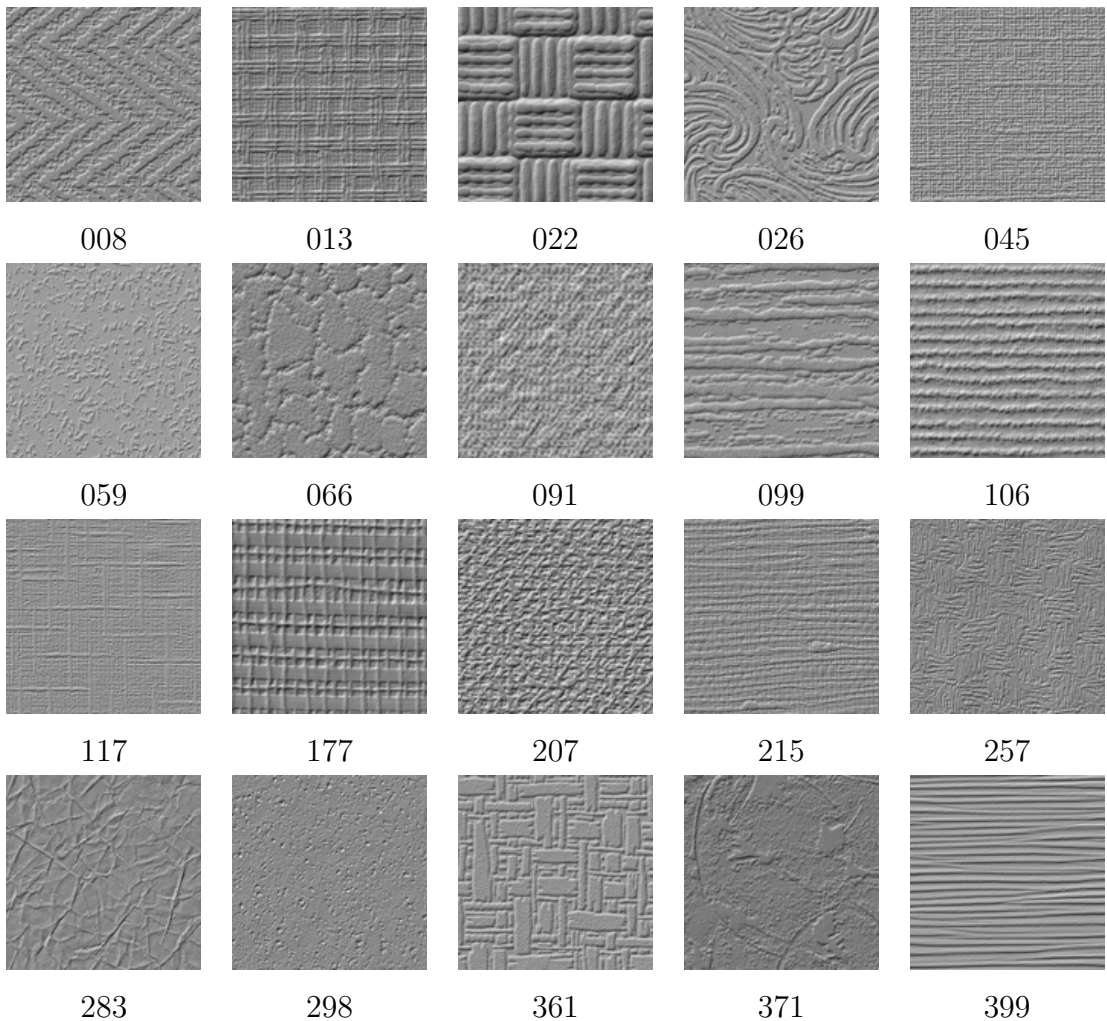
While we adopted an application independent approach in this project, the capture of similarity judgements and use of browsing environments may well be affected when constrained to particular applications or texture types, e.g. food production, wood laminate manufacture, etc. where quality control or defect detection is carried out by visual inspection by experts. Domain specific applications may be an area of future interest.

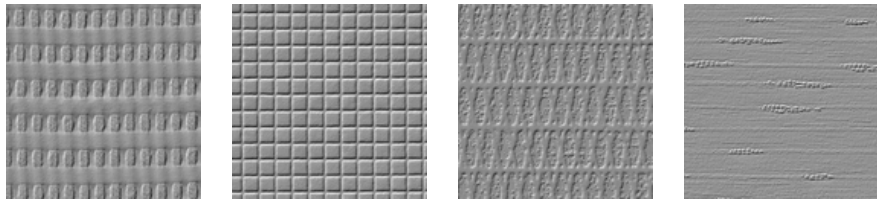
Appendix A

Experimental Stimuli

A.1 *Tex500[024]*

These 24 textures are referred to as *Tex500[024]* in the text. The remaining 476 textures are known as *Tex500[476]*.





425

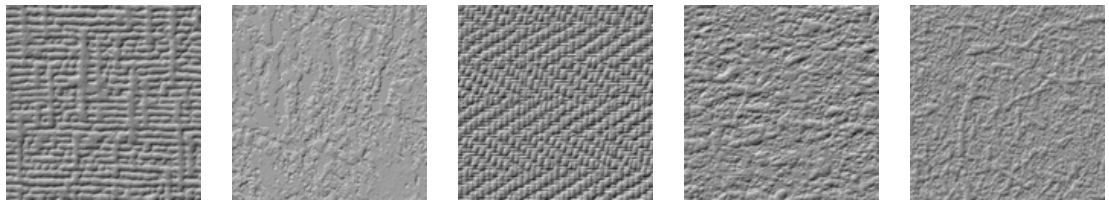
444

448

485

A.2 *Tex500[100]*

These 100 textures are referred to as *Tex500[100]* in the text. The remaining 400 textures are known as *Tex500[400]*.



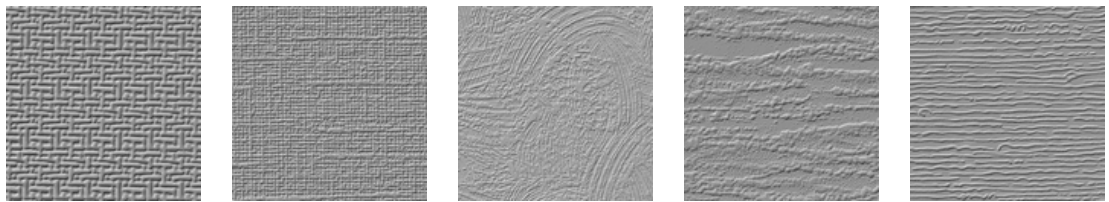
007

010

014

017

023



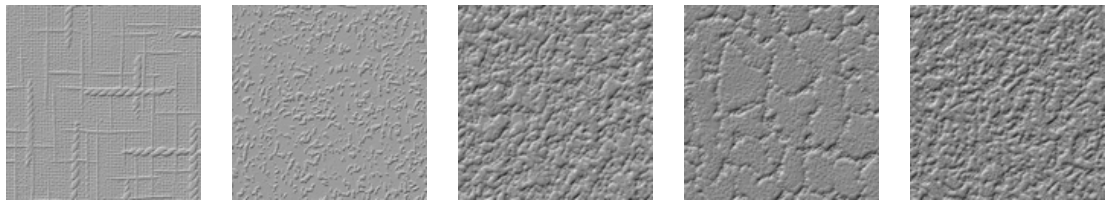
034

045

049

054

056



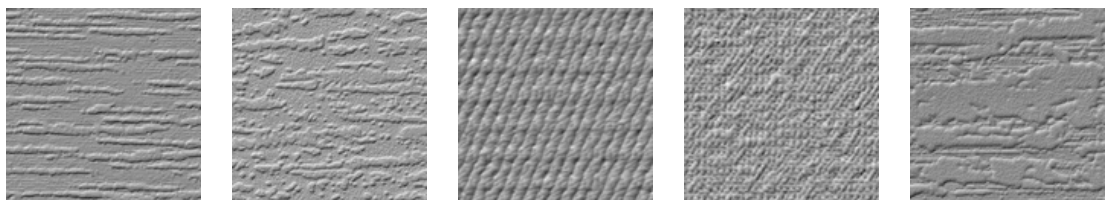
058

059

063

066

072



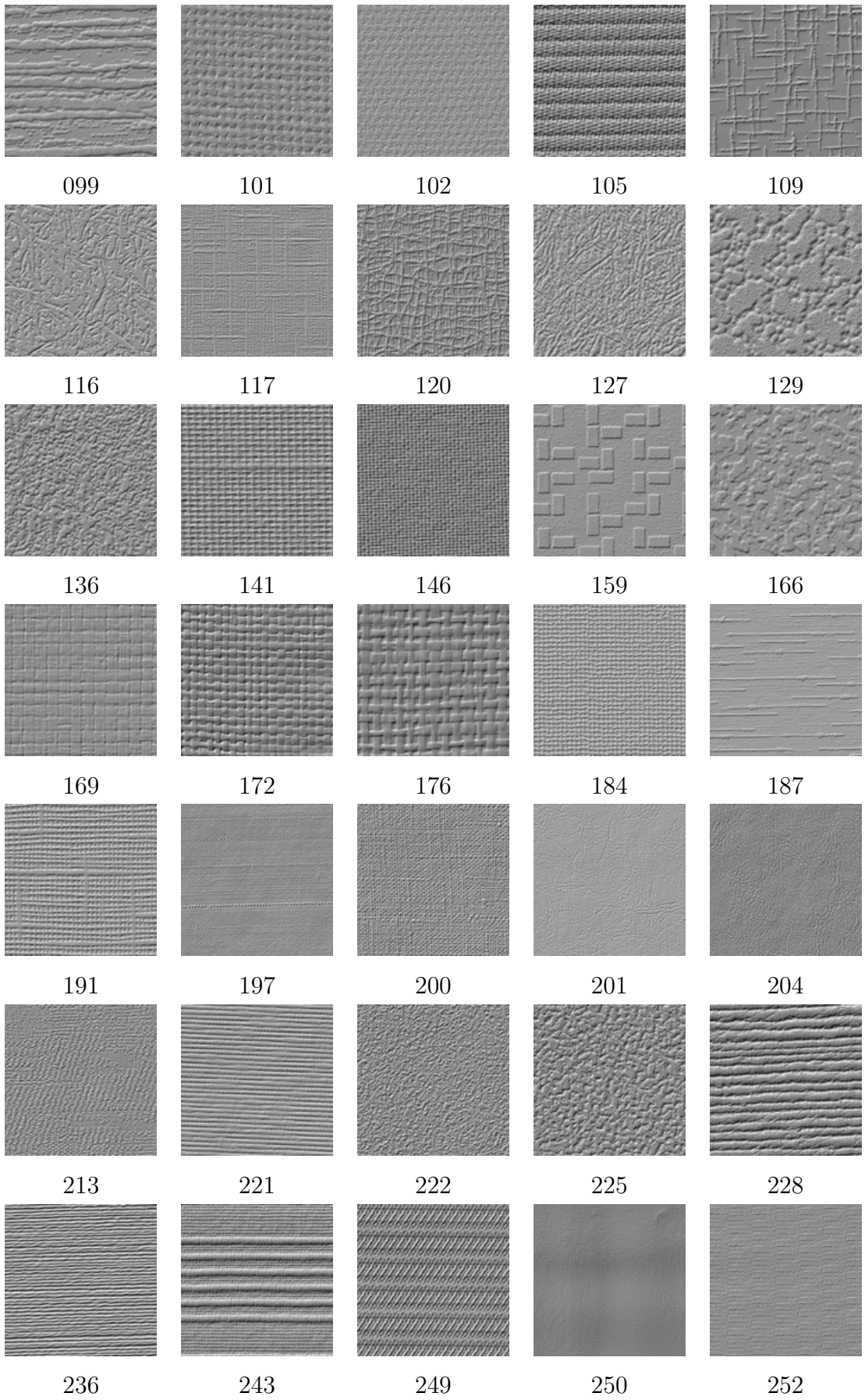
079

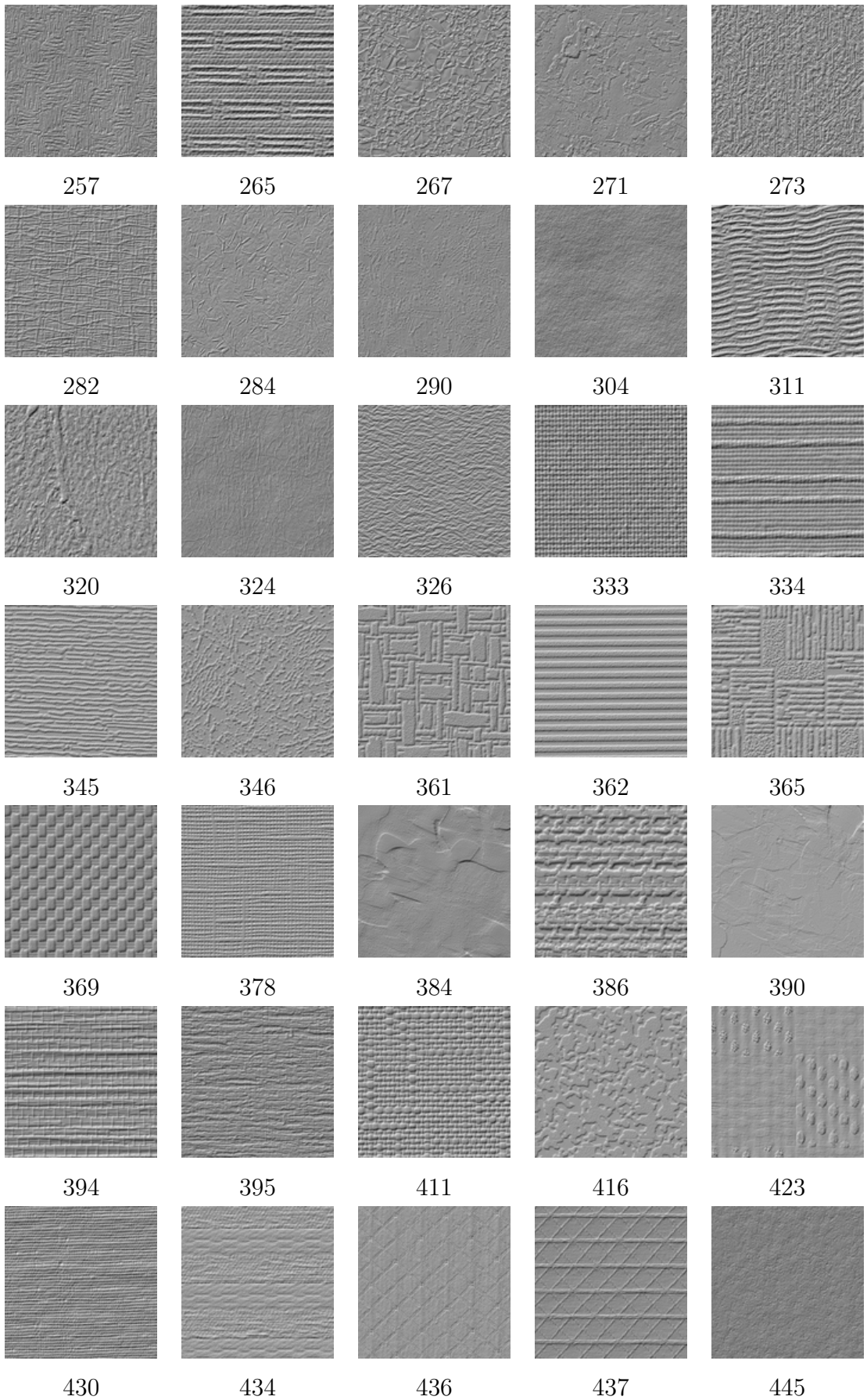
082

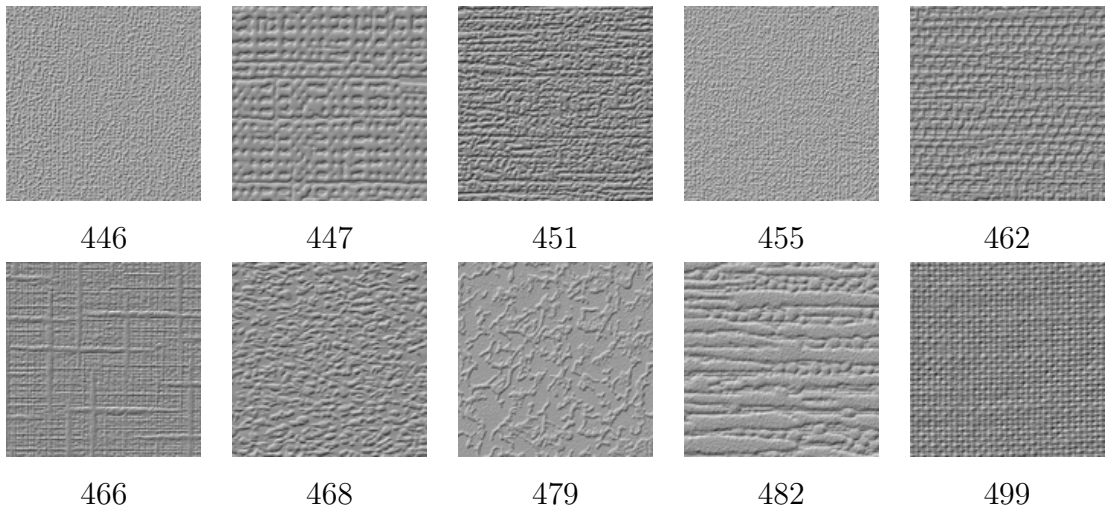
088

091

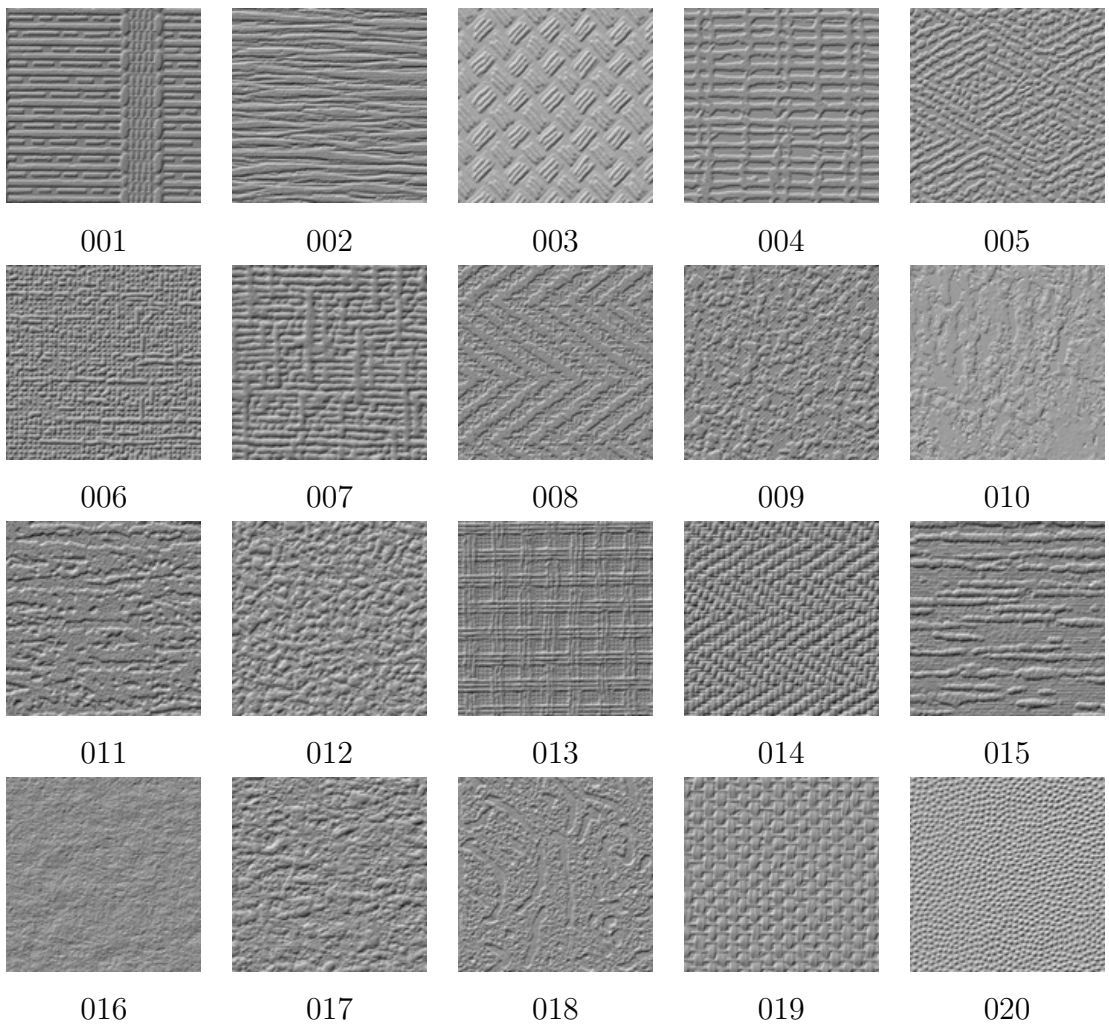
095

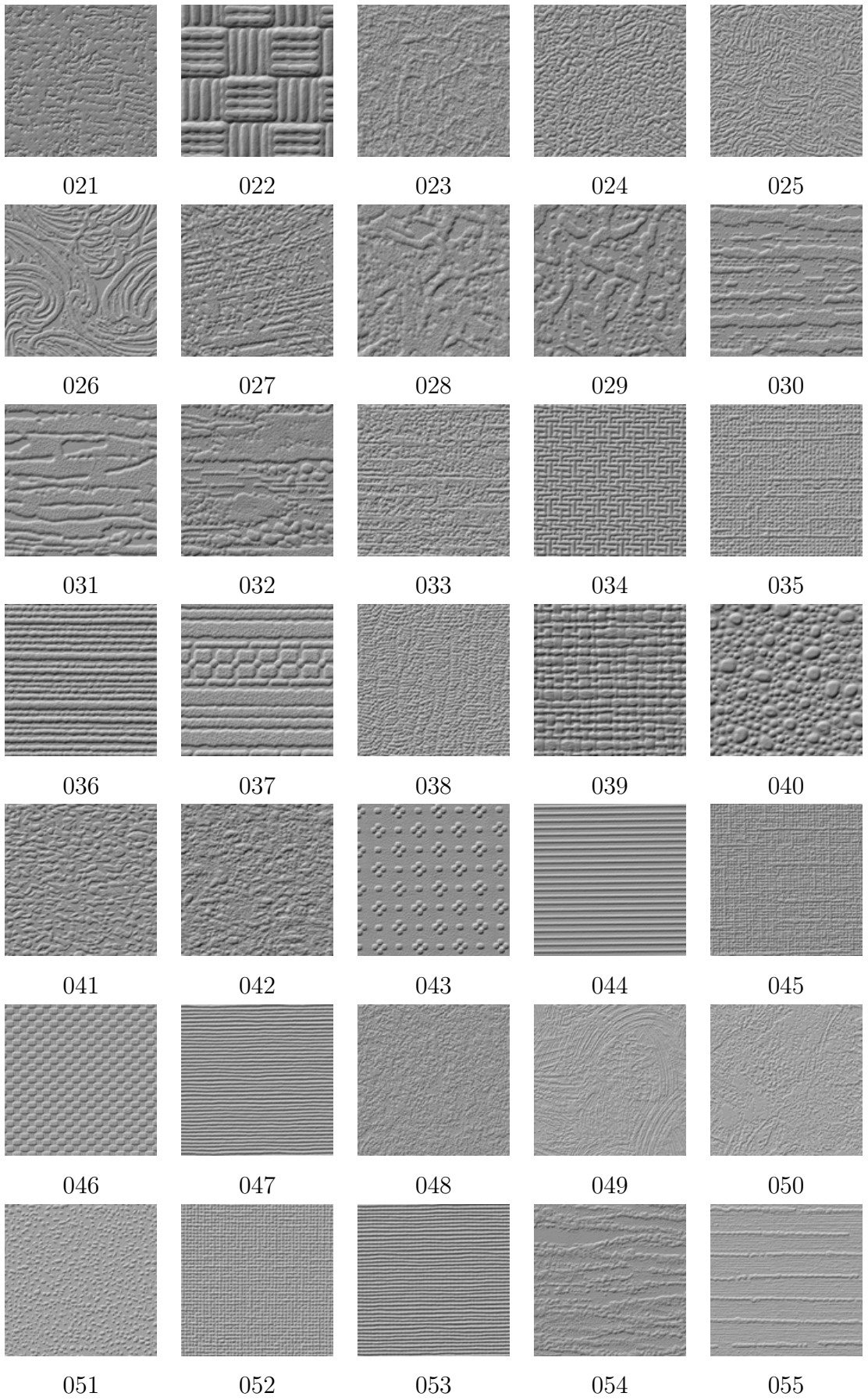


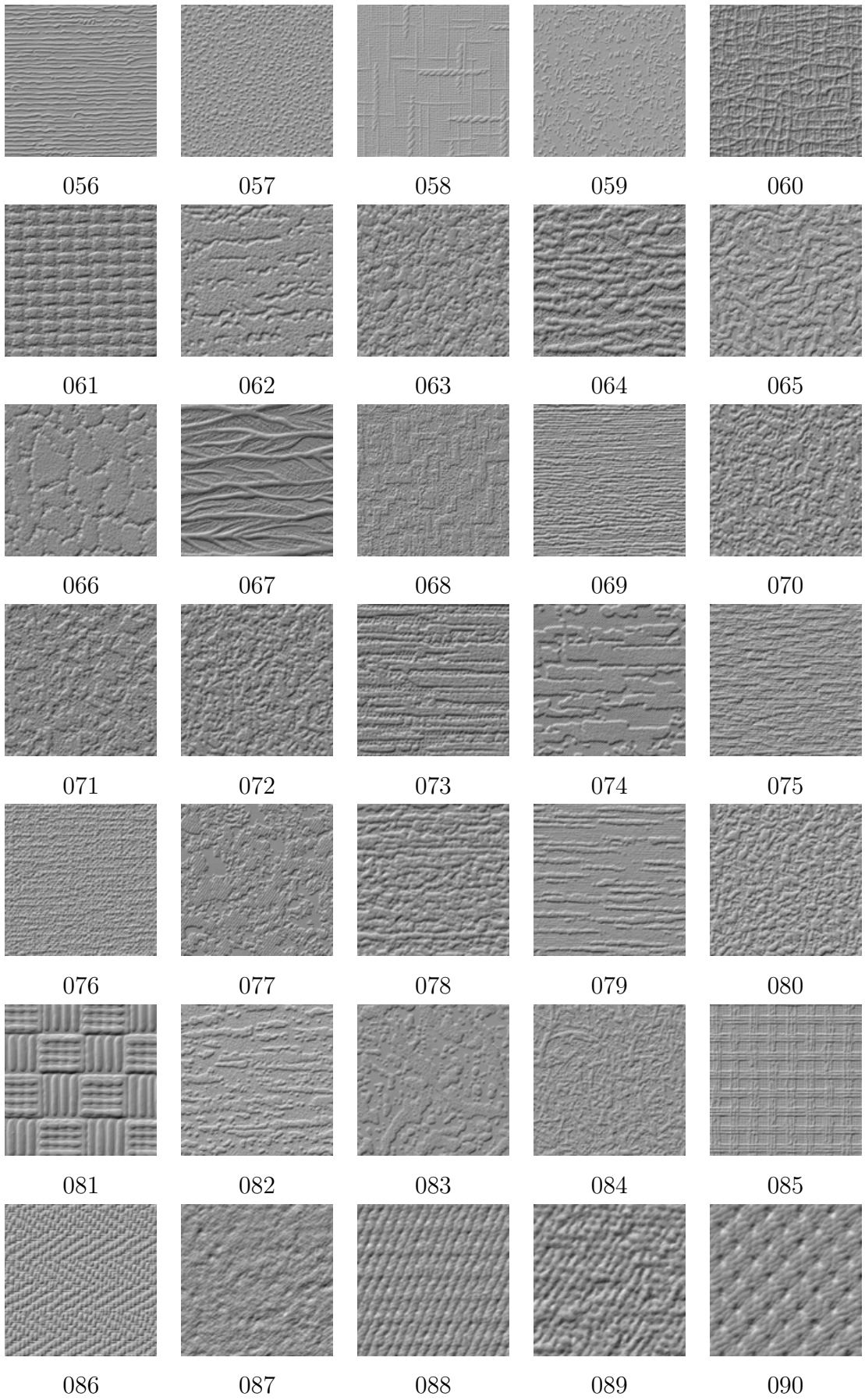


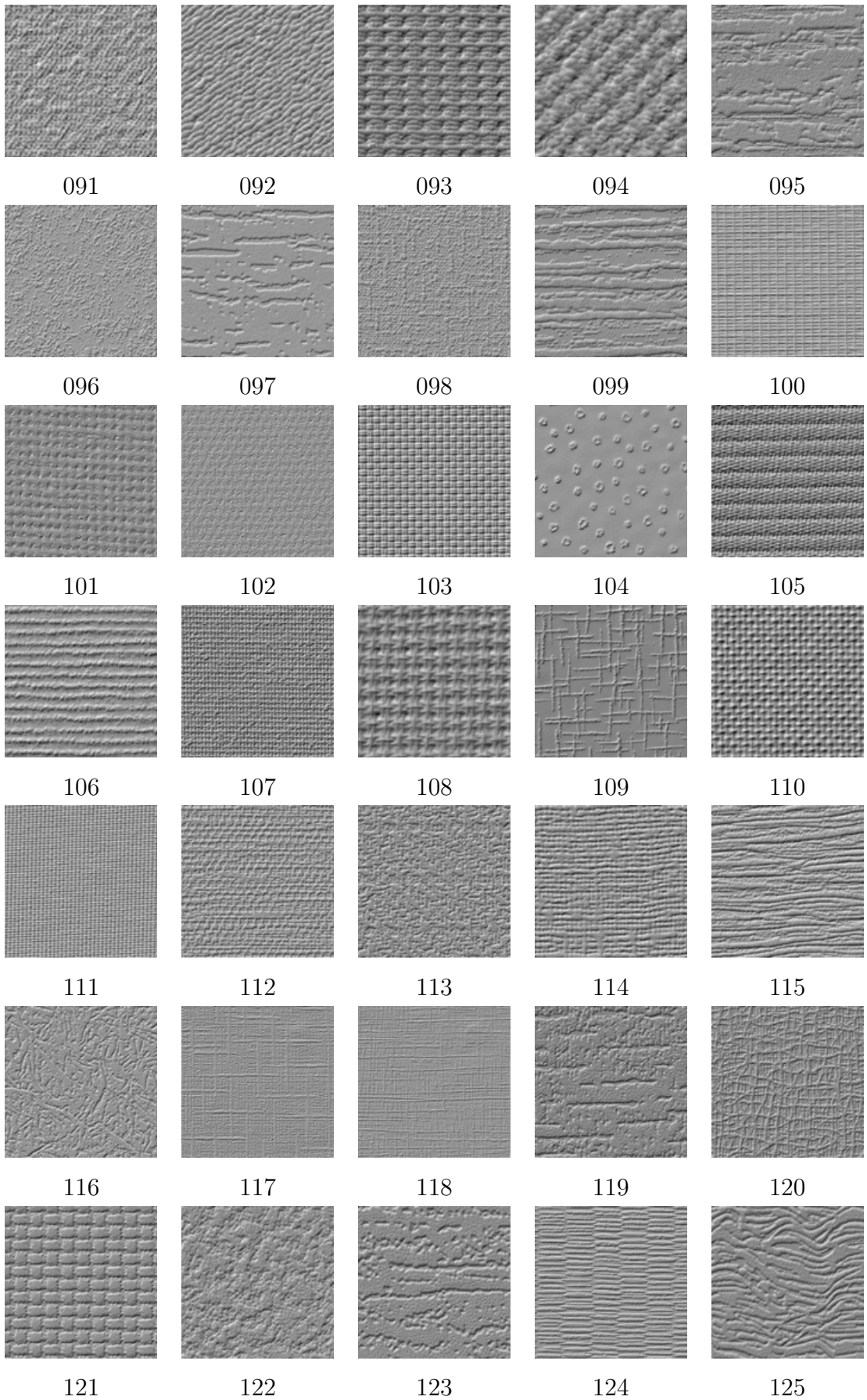


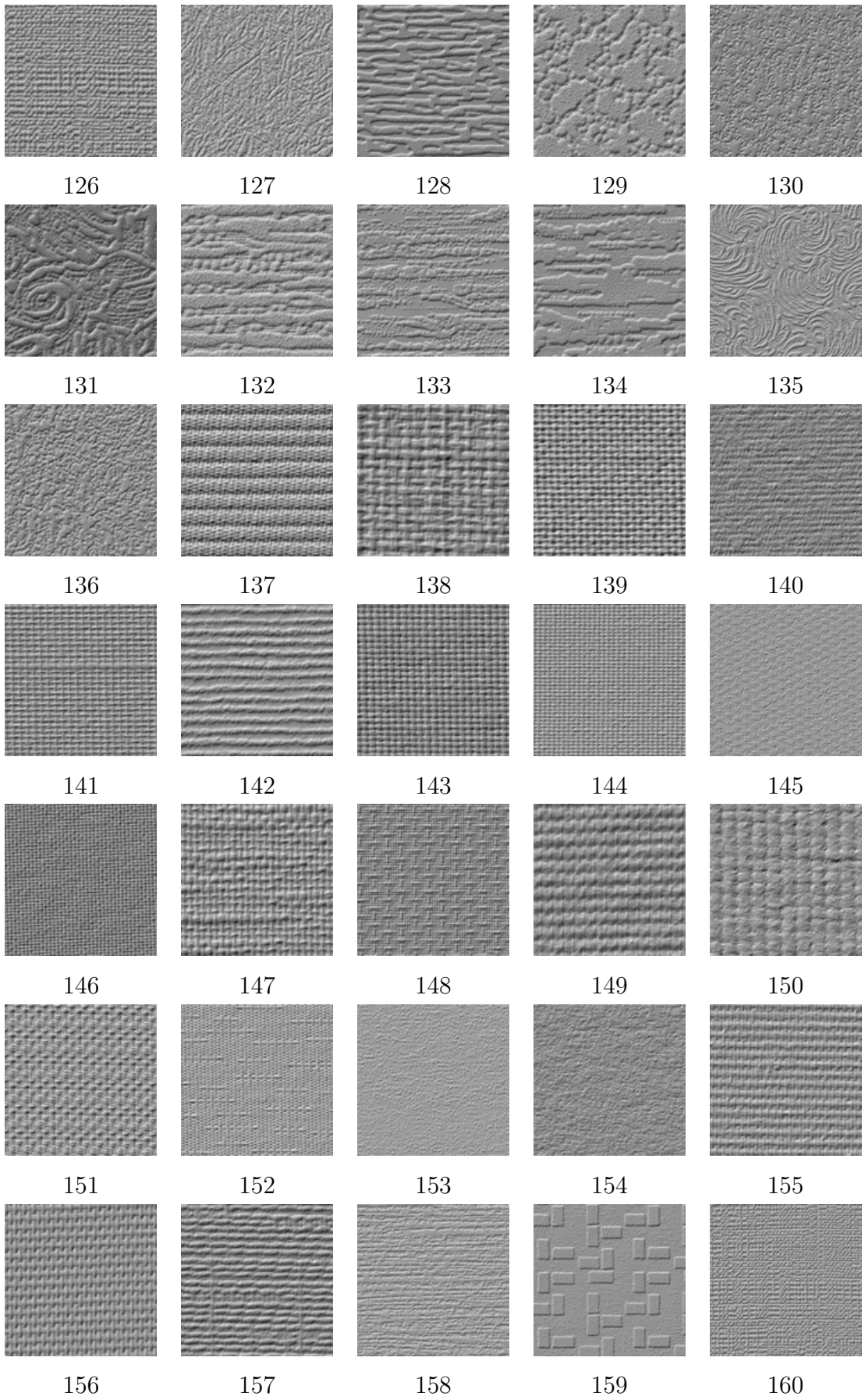
A.3 *Tex500*

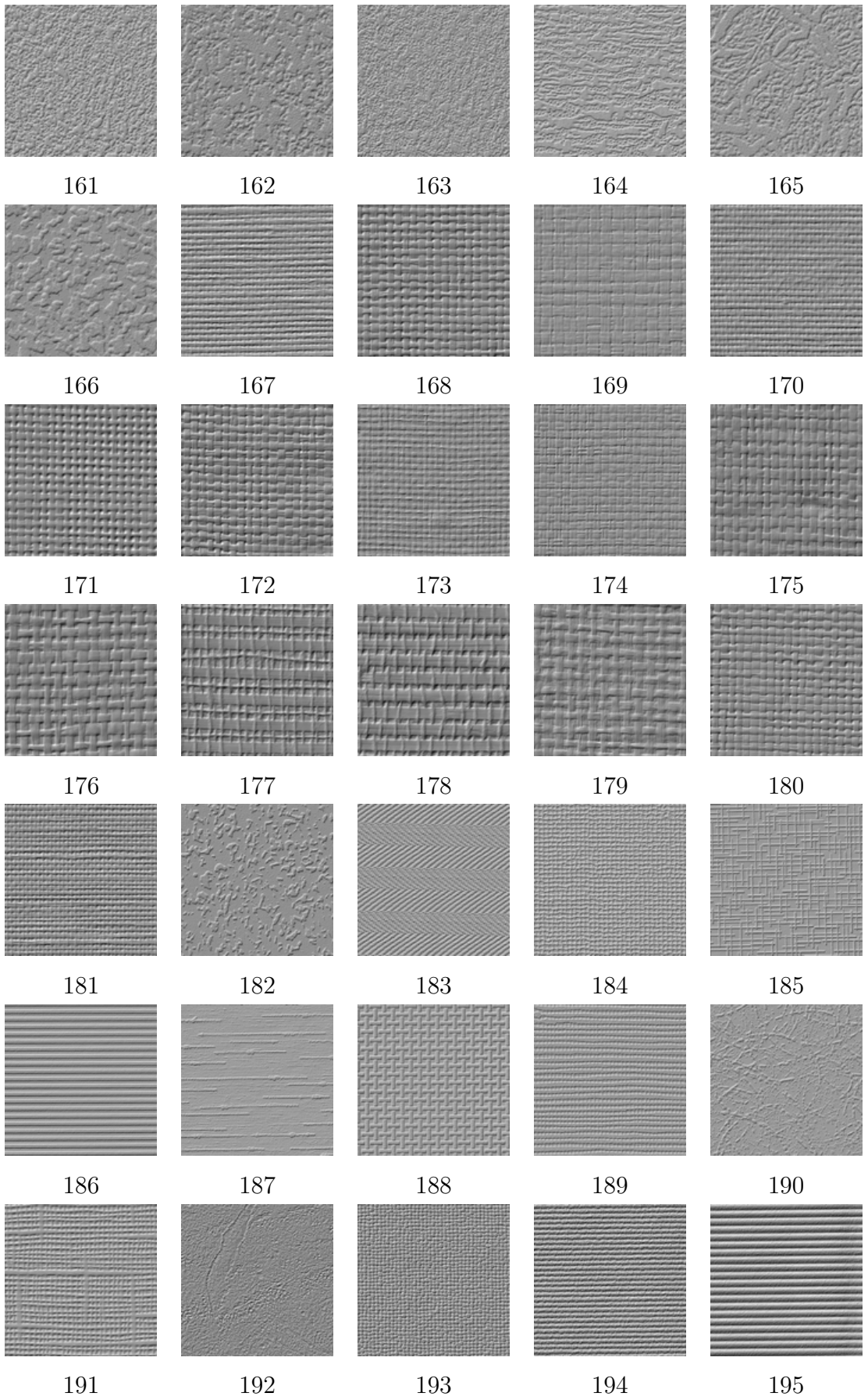


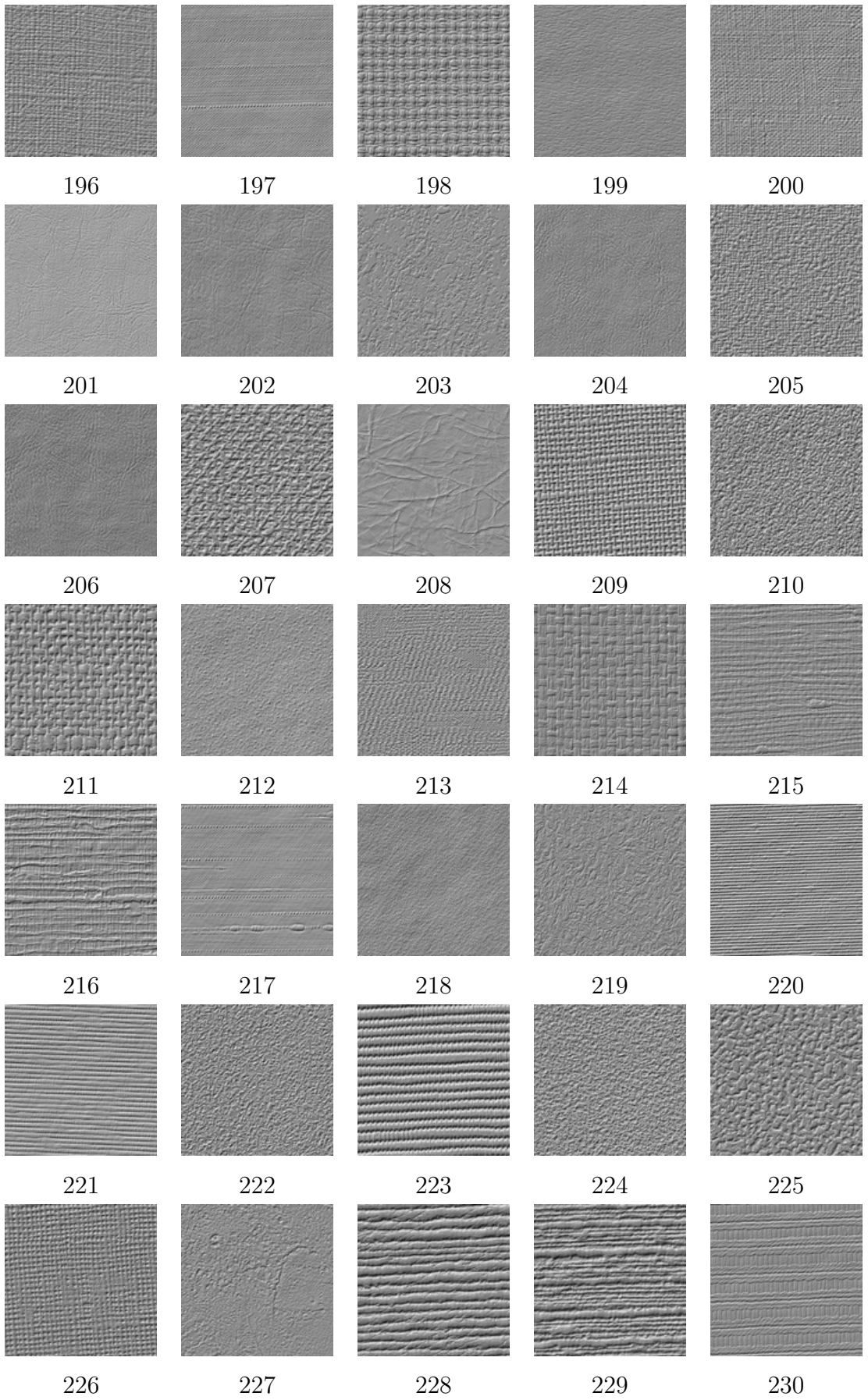


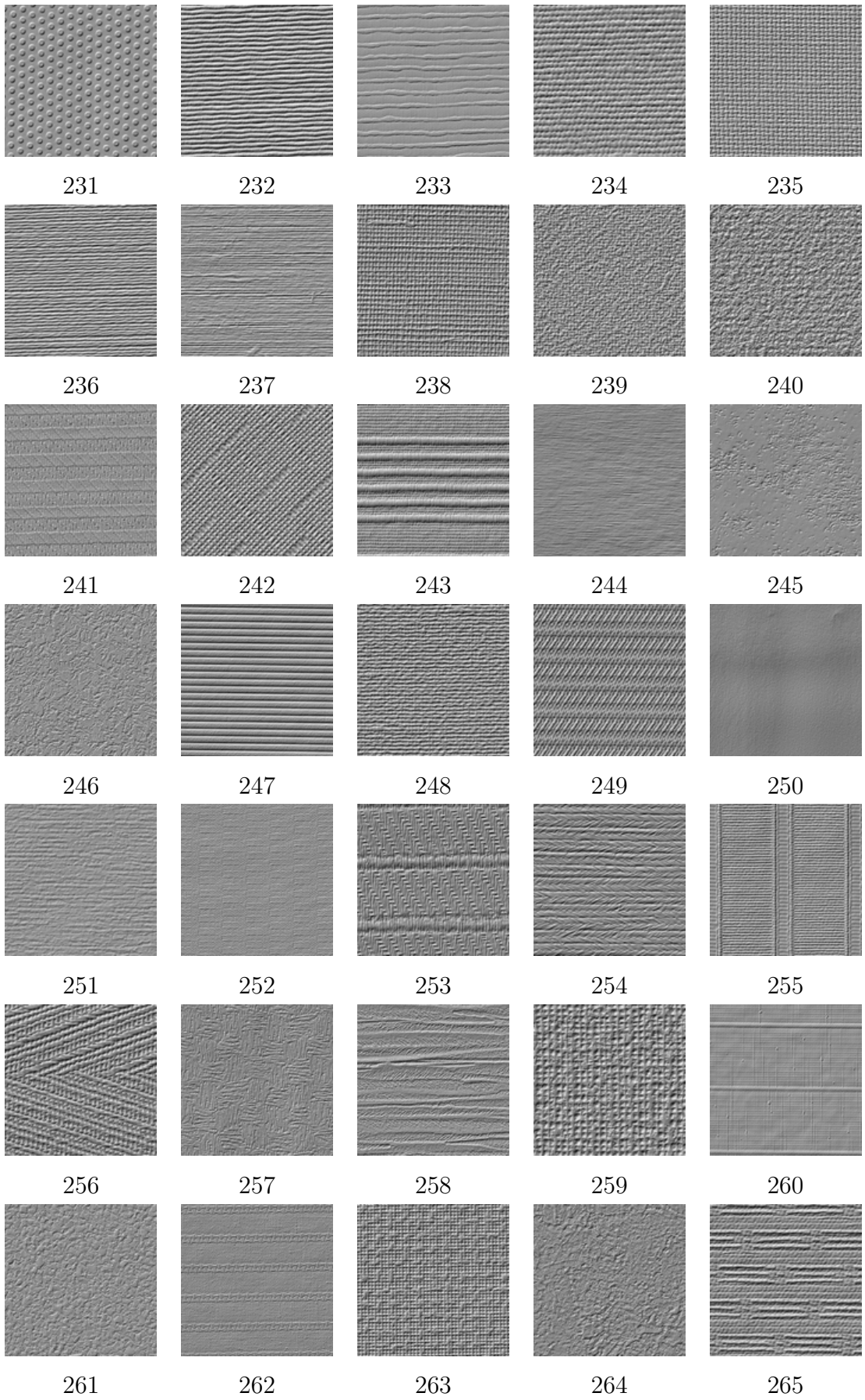


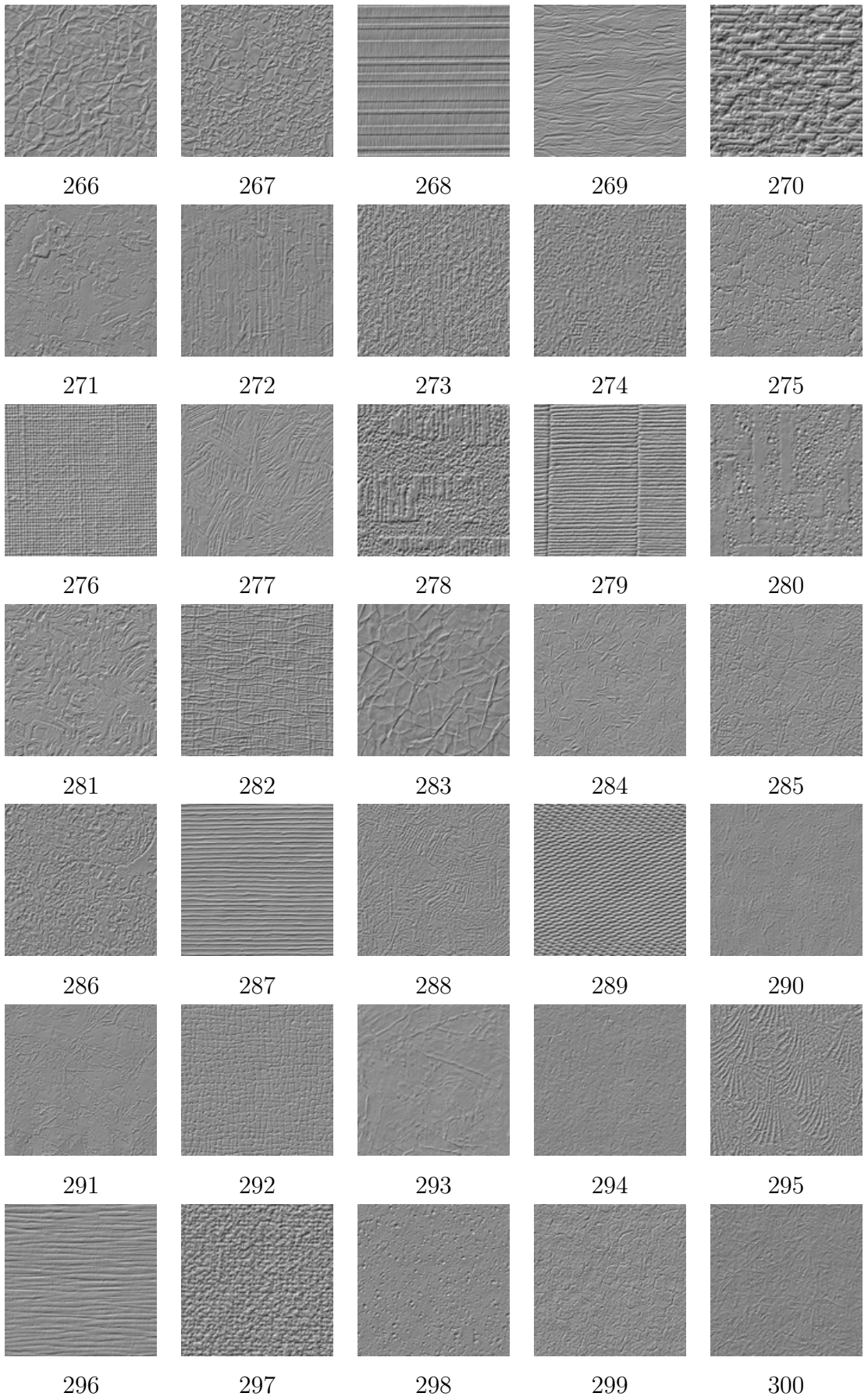


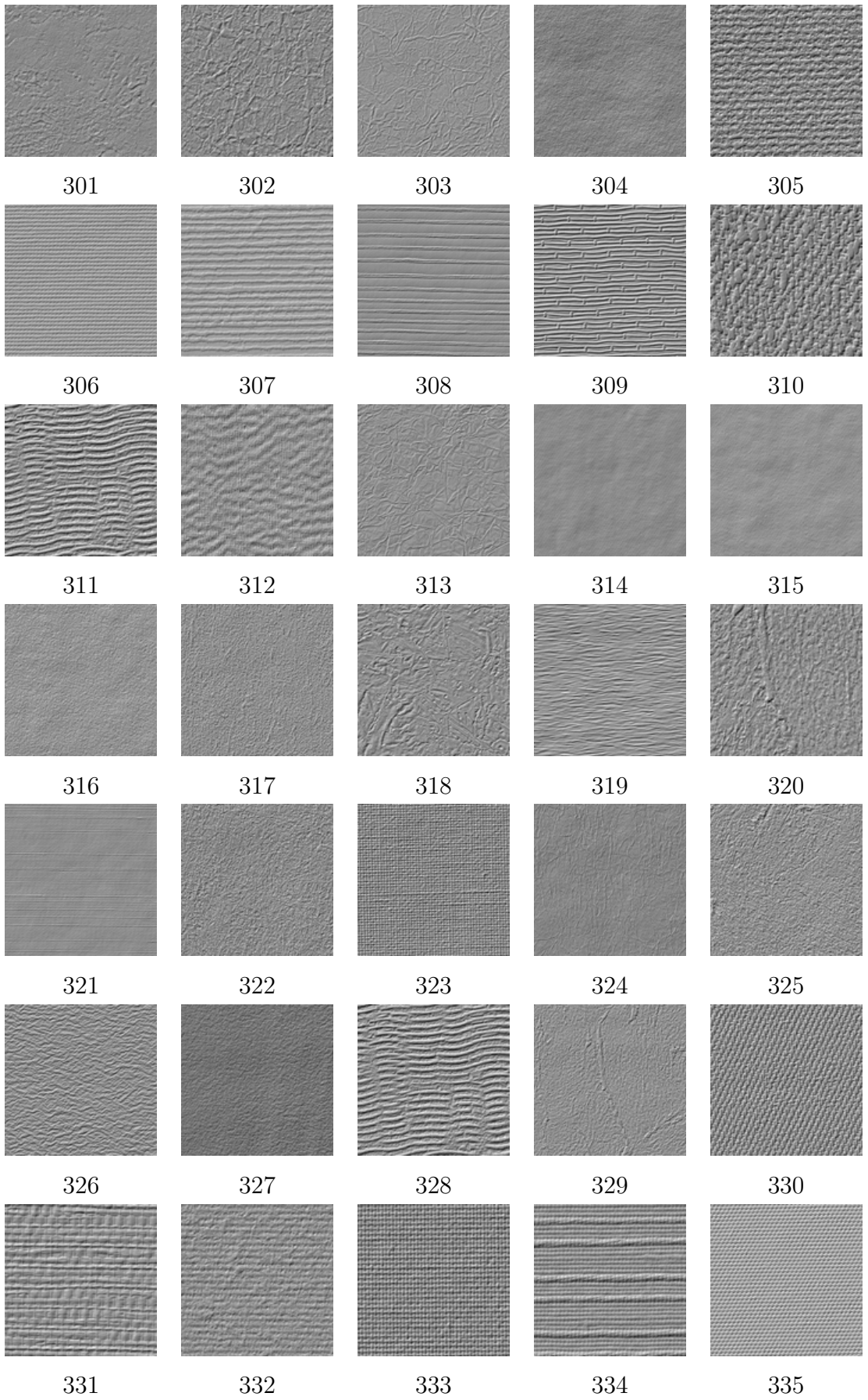


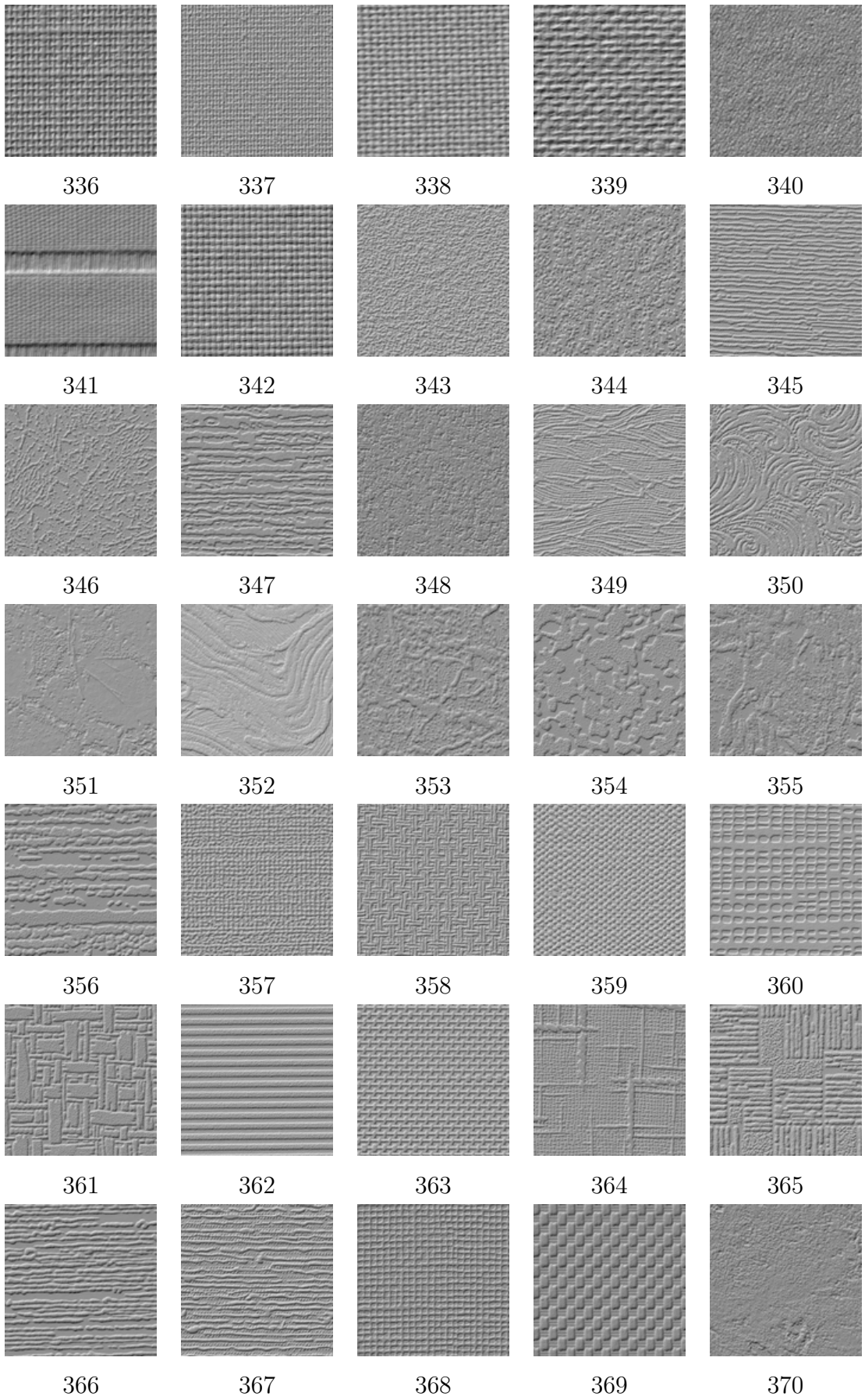


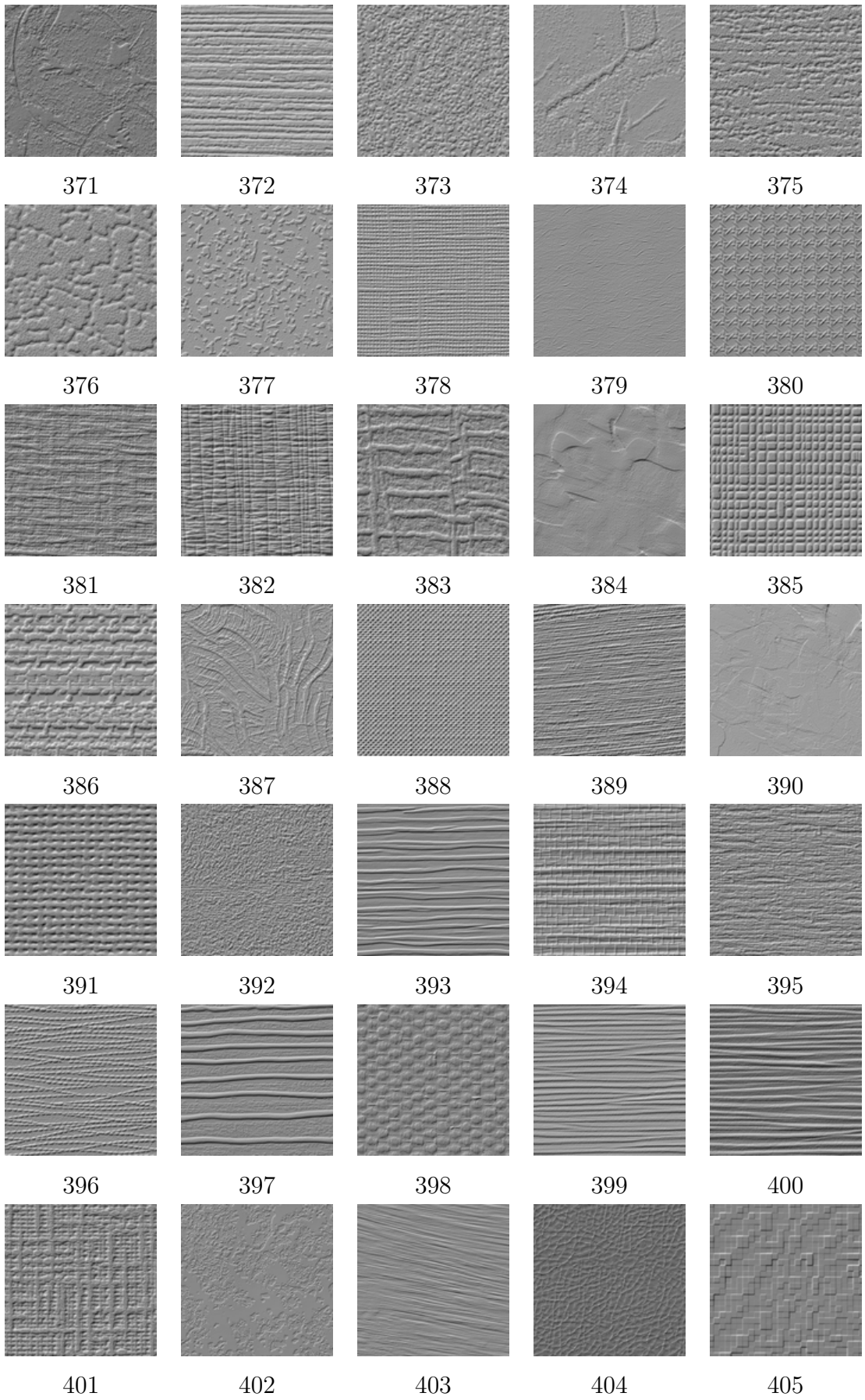


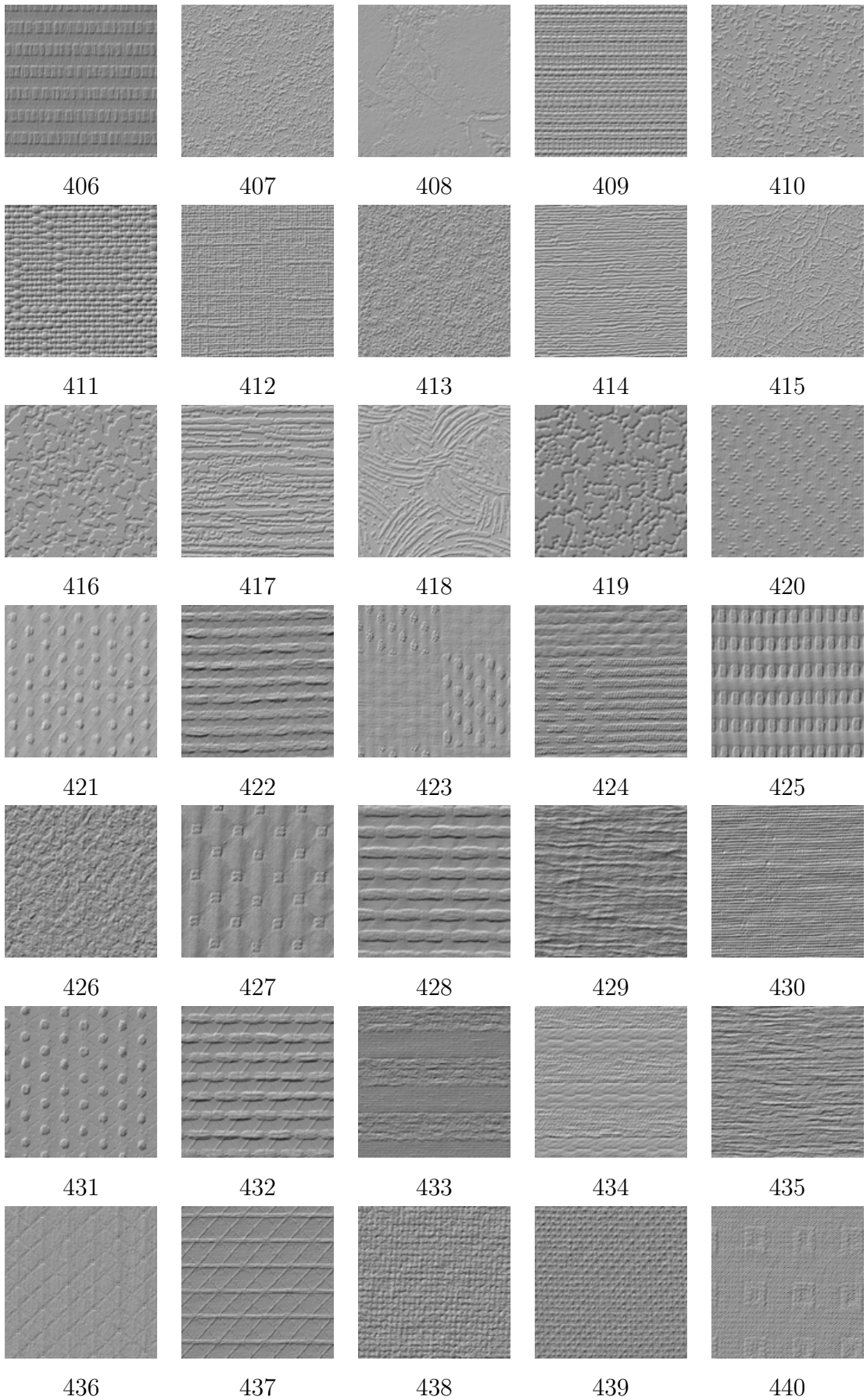


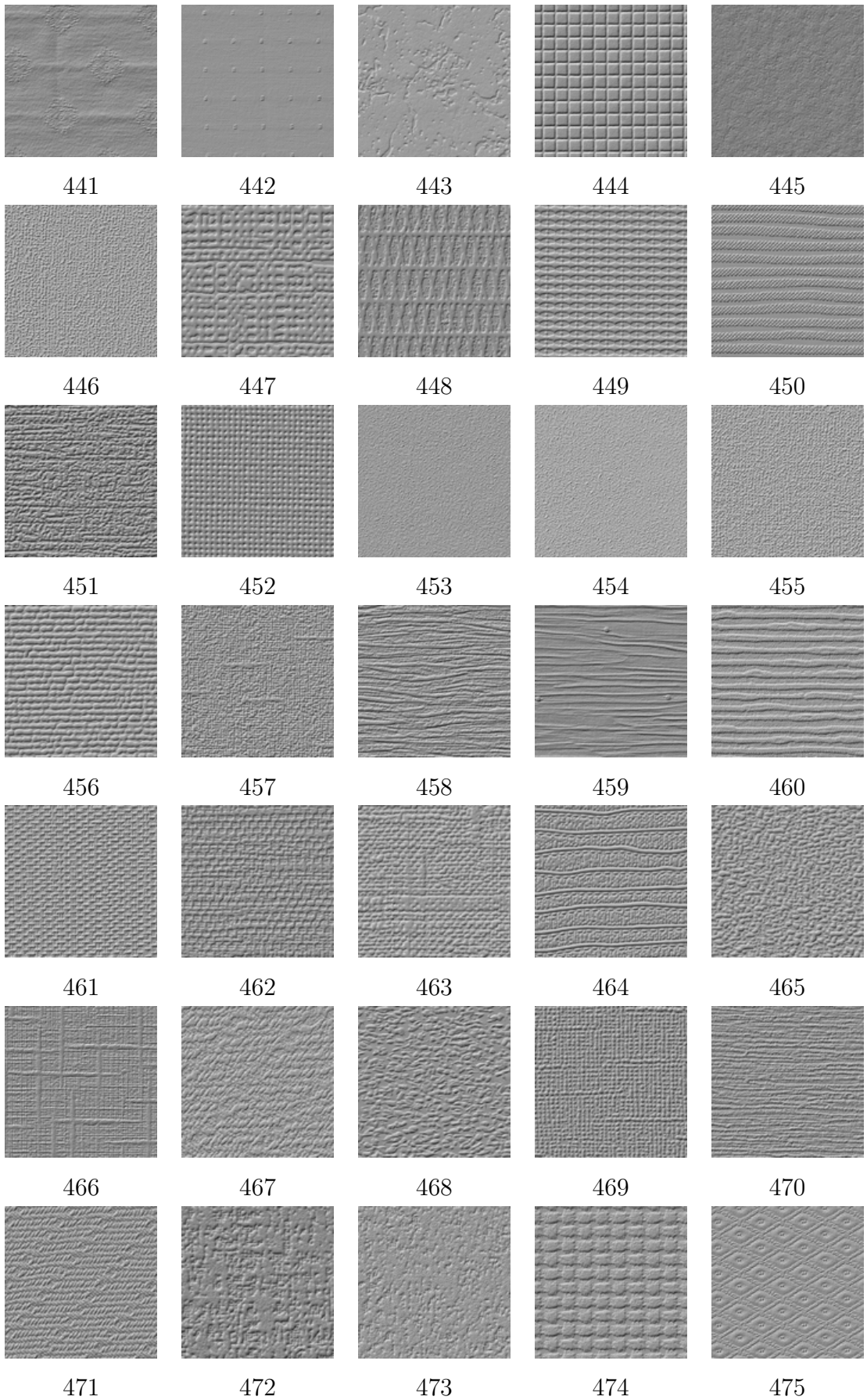


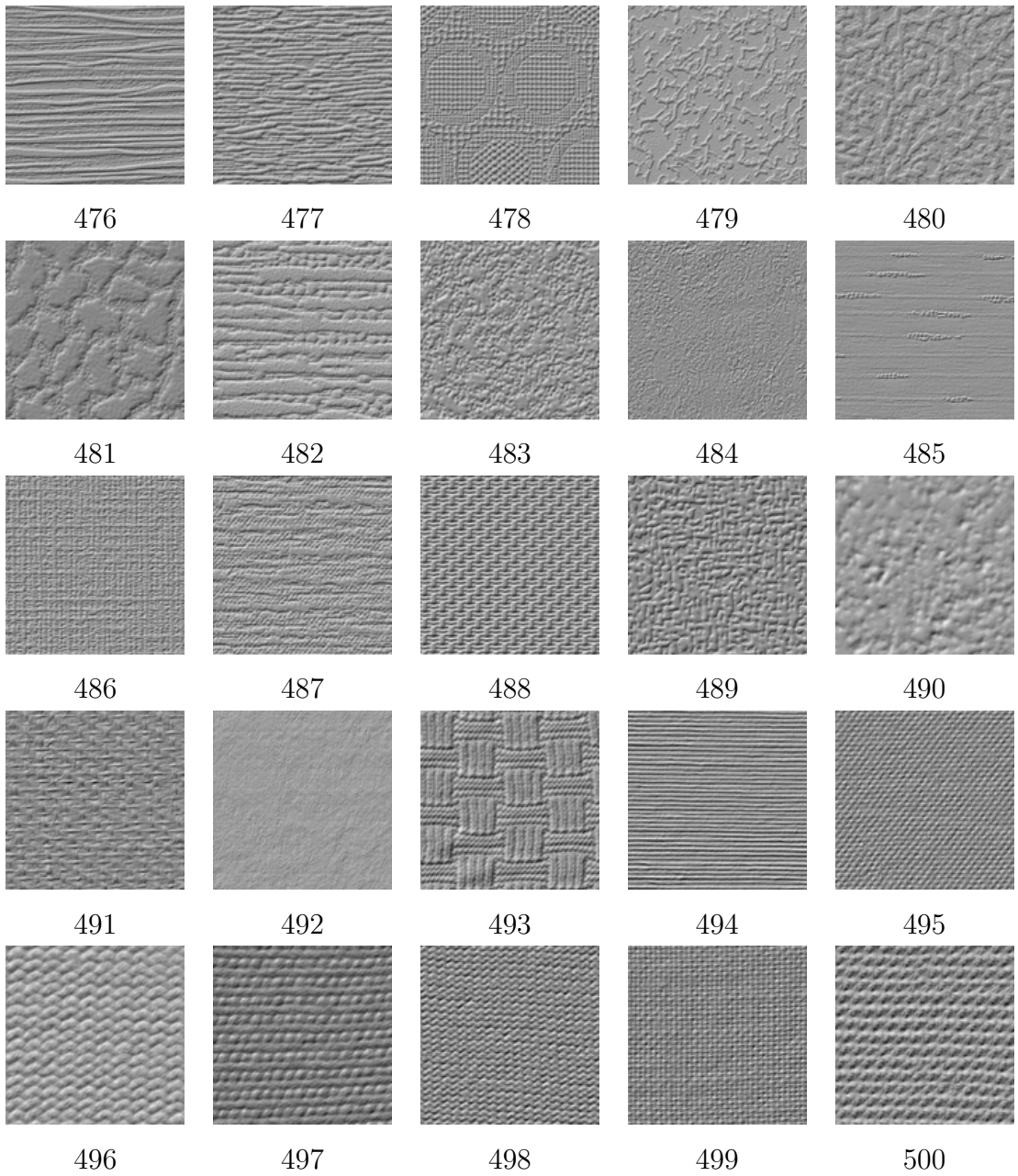












Appendix B

Detailed Experimental Results

B.1 SOM_G, MDS_G & RF_G Browsing Environments Comparison

B.1.1 Mean Task Time

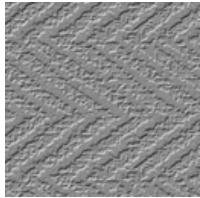
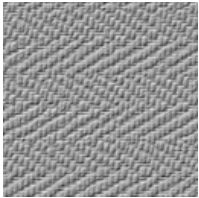
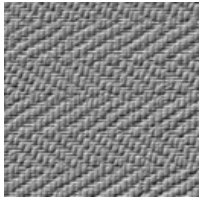
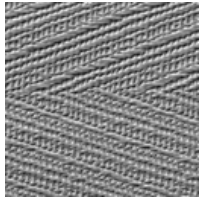
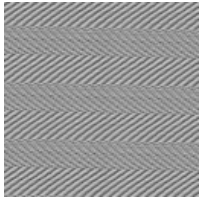
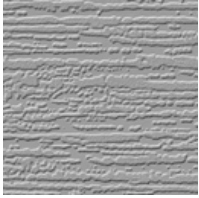
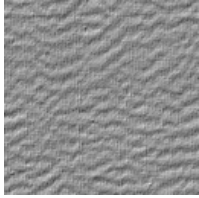
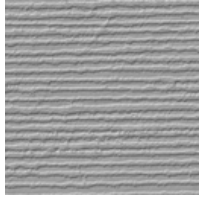
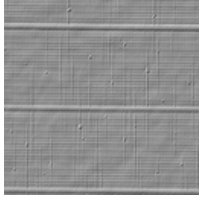
Observer	Mean Task Time			Preferred
	SOM _G	MDS _G	RF _G	
1	62.500	62.000	77.875	MDS _G
2	52.125	120.250	215.125	SOM _G
3	67.875	54.375	116.875	MDS _G
4	38.125	122.375	173.375	SOM _G
5	26.625	226.875	122.625	SOM _G
6	38.125	174.000	165.500	SOM _G
7	53.875	37.750	66.625	SOM _G
8	44.125	81.125	129.875	SOM _G
9	90.500	59.375	212.875	MDS _G
10	47.625	91.625	78.500	SOM _G
11	36.000	187.000	239.500	SOM _G
12	53.625	128.375	123.125	SOM _G
Means & Most Preferred	50.927	112.094	143.490	SOM _G

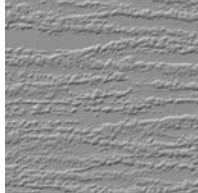
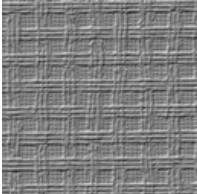
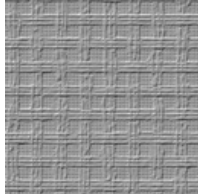
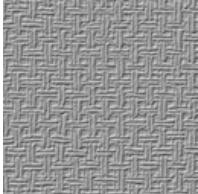
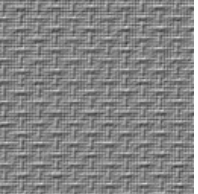
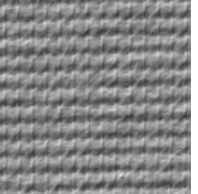
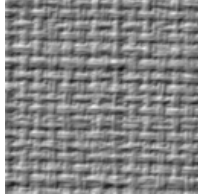
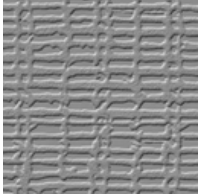
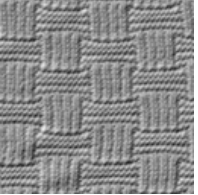
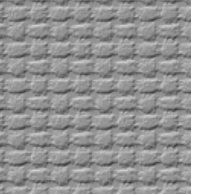
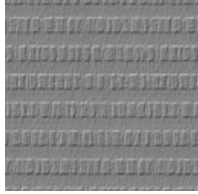
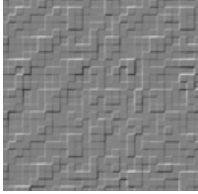
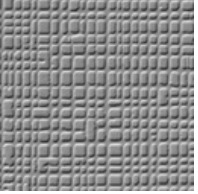
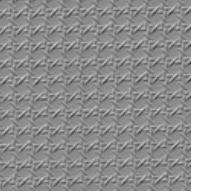
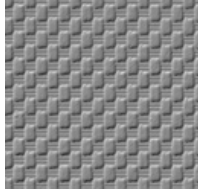
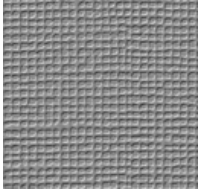
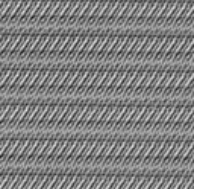
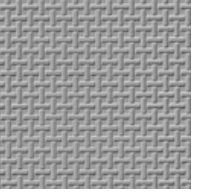
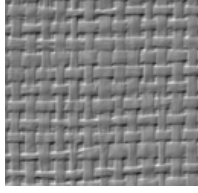
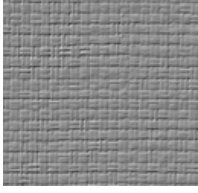
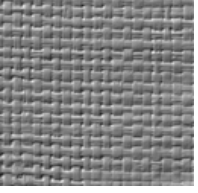
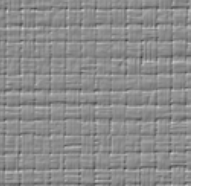
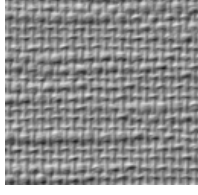
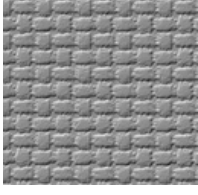
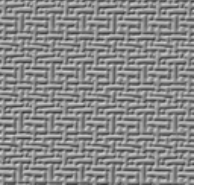
B.1.2 Accuracy

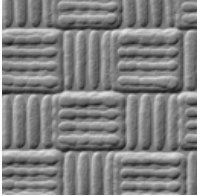
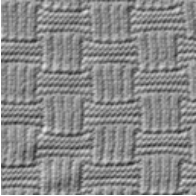
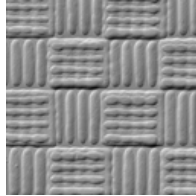
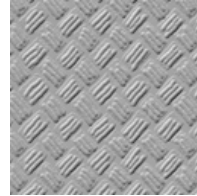
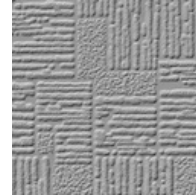
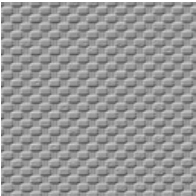
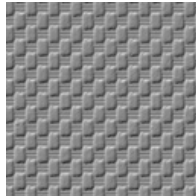
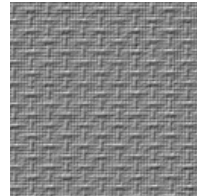
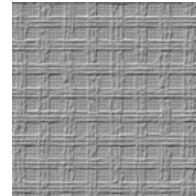
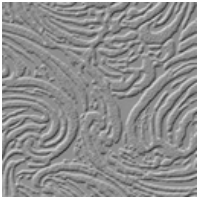
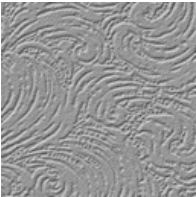
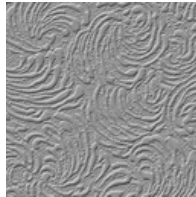

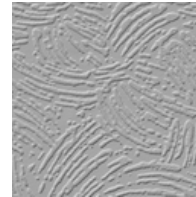
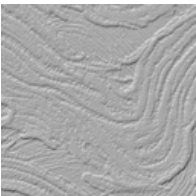
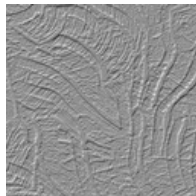
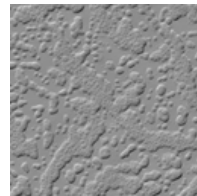
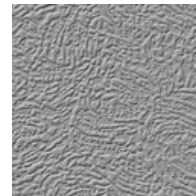
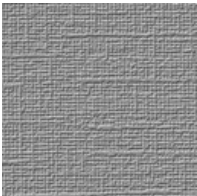
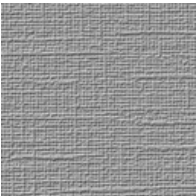
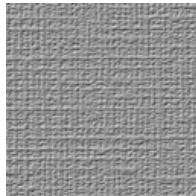
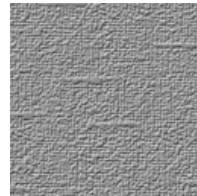
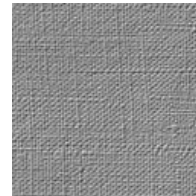
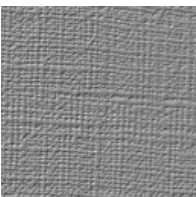
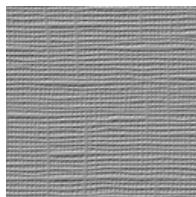
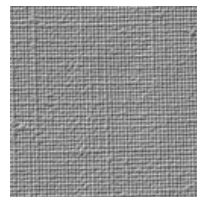
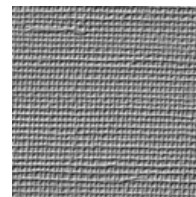
Observer	Accuracy (%)		
	SOM _G	MDS _G	RF _G
1	100.000	100.000	100.000
2	100.000	100.000	100.000
3	97.375	100.000	100.000
4	100.000	100.000	100.000
5	100.000	97.375	100.000
6	100.000	97.375	100.000
7	100.000	100.000	100.000
8	100.000	100.000	93.750
9	100.000	100.000	100.000
10	93.750	97.375	100.000
11	100.000	97.375	100.000
12	100.000	100.000	100.000
Means	99.260	99.125	99.480

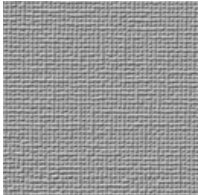
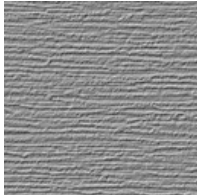
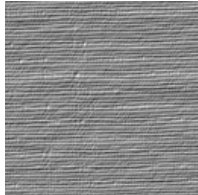
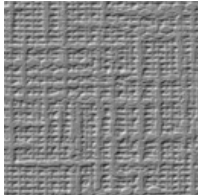
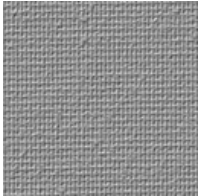
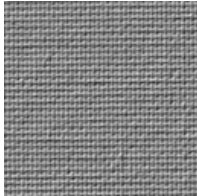
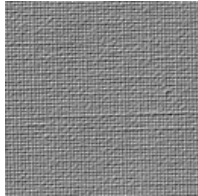
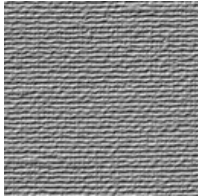
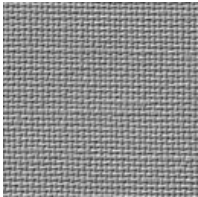
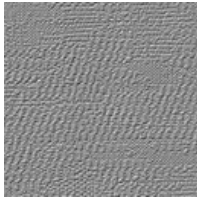
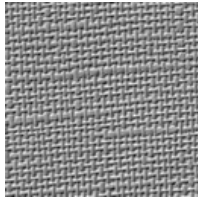
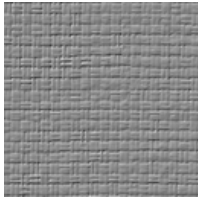
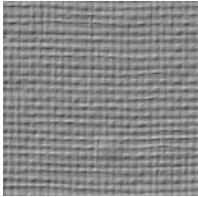
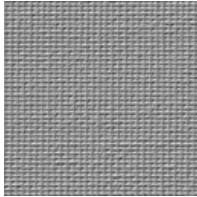
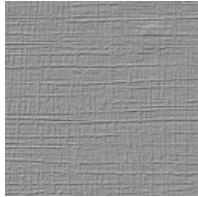
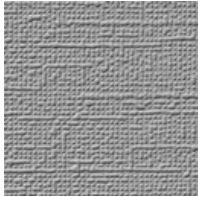
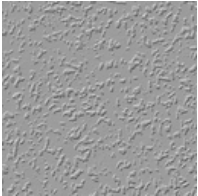
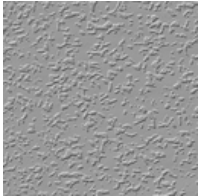
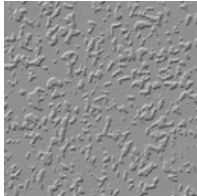
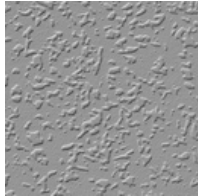
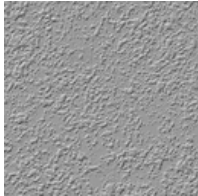
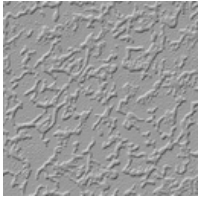
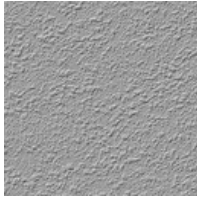
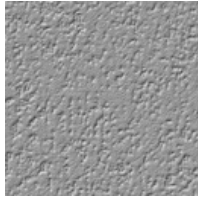
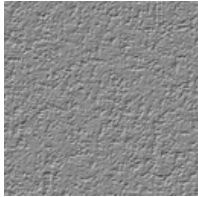
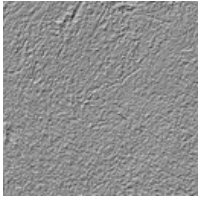
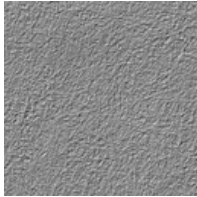
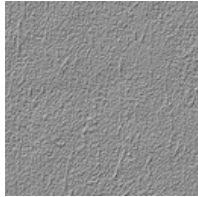
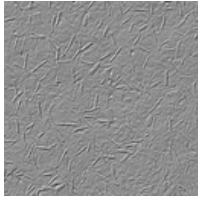
B.2 Pilot Dataset Augmentation

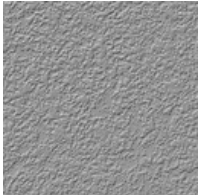
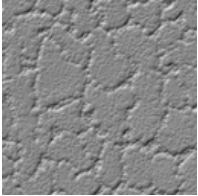
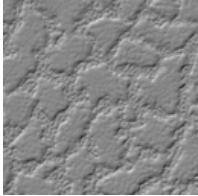
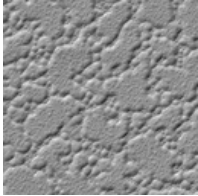
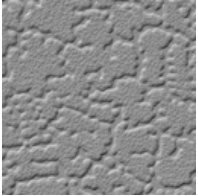
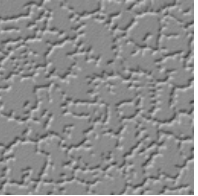
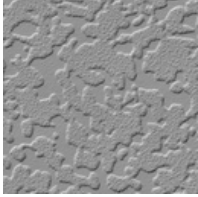
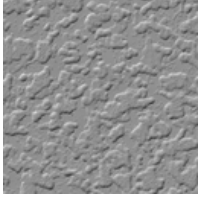
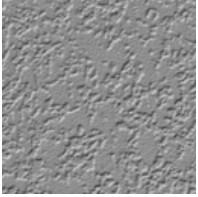
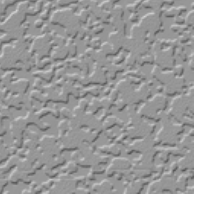
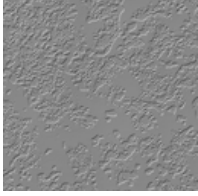
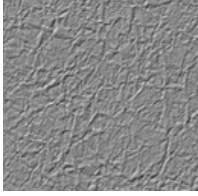
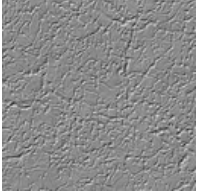
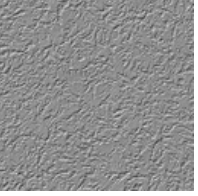
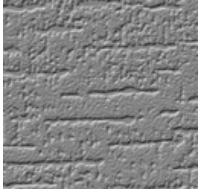
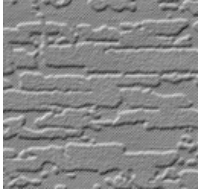
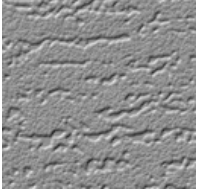
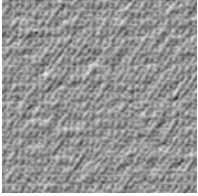
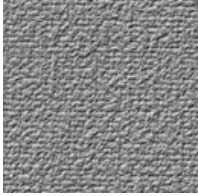
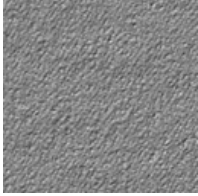
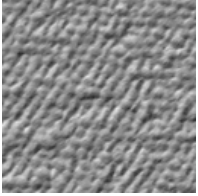
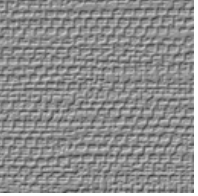
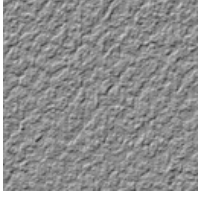
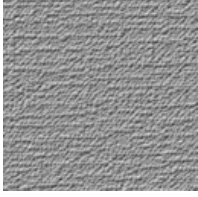
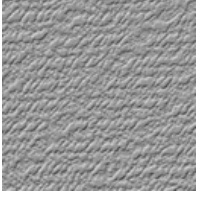
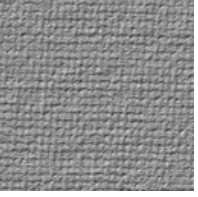
Note: Figure in brackets denotes number of votes received.

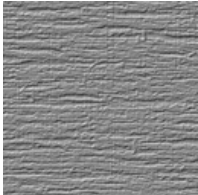
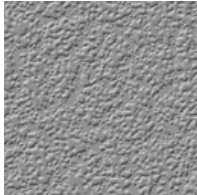
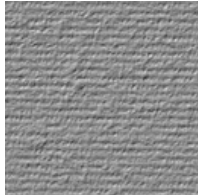
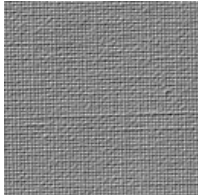
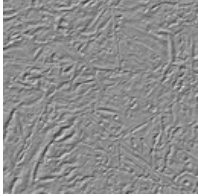
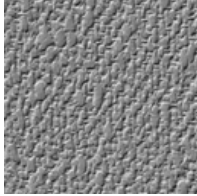
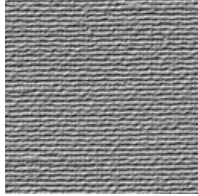
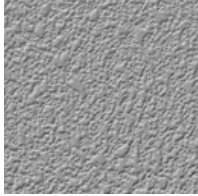
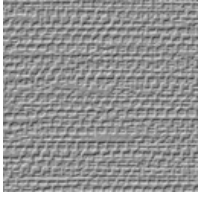
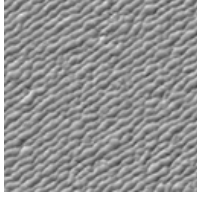
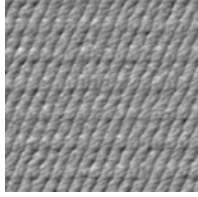
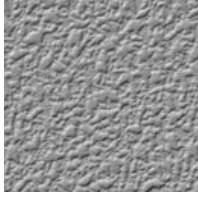
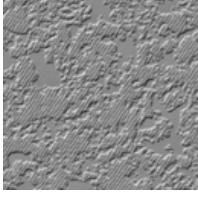
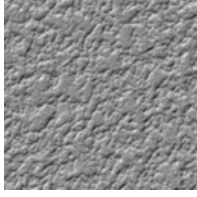
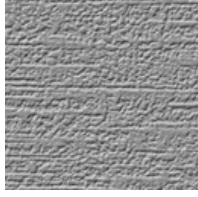
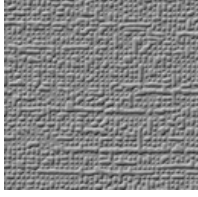
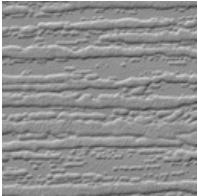
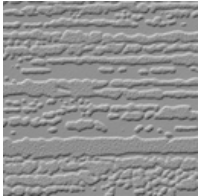
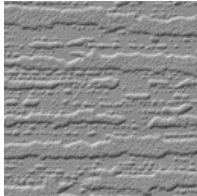
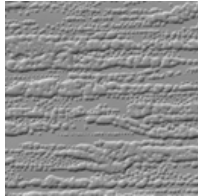
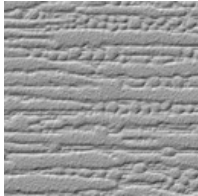
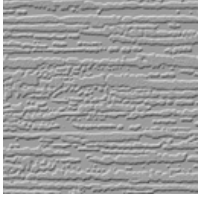
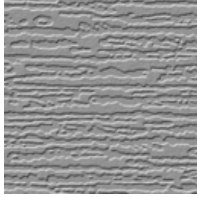
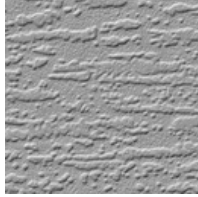
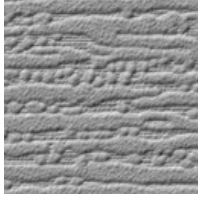
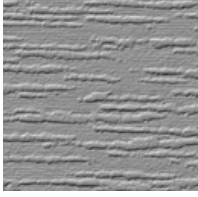
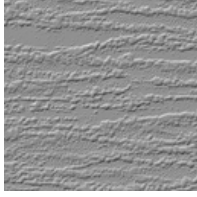
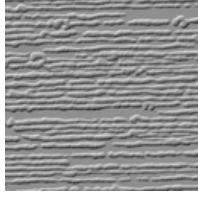
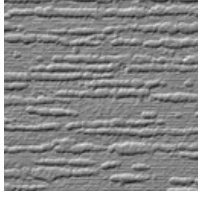
Stimuli	Responses			
				
008	086 (7)	014 (7)	256 (6)	183 (2)
				
	417 (1)	312 (1)	307 (1)	260 (1)

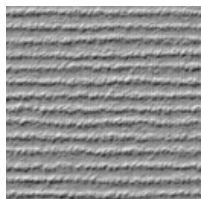
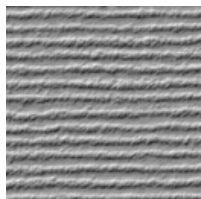
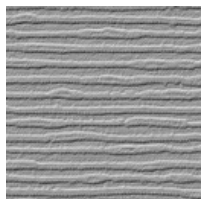
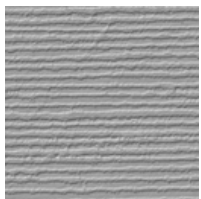
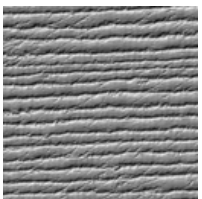
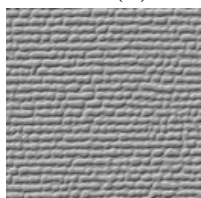
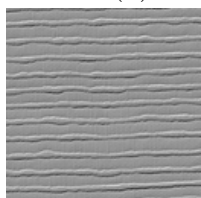
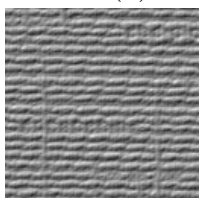
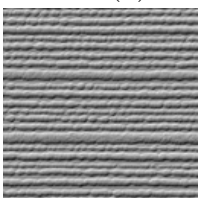
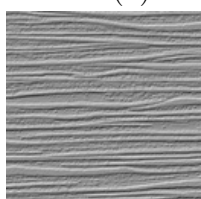
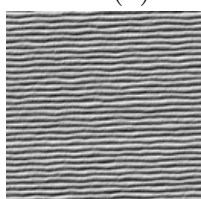
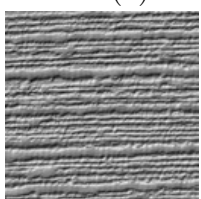
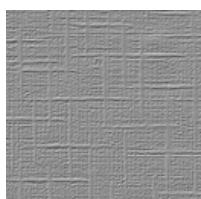
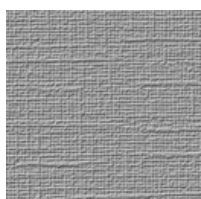
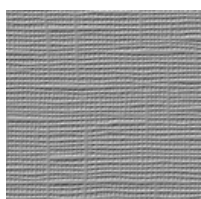
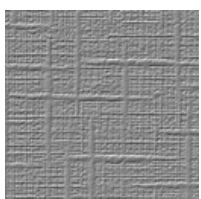
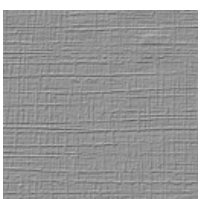
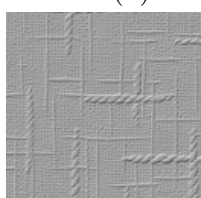
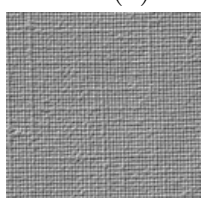
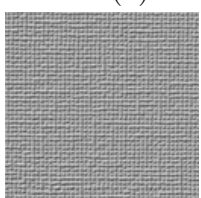
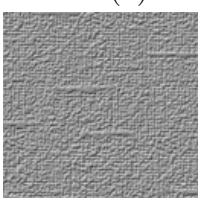
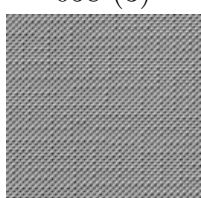
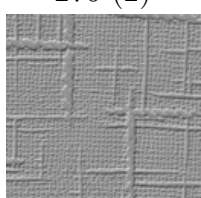
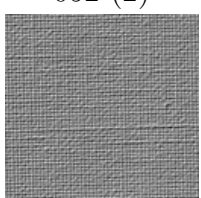
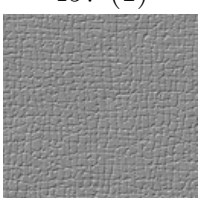
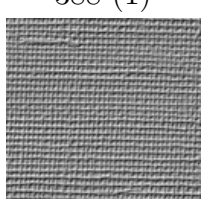
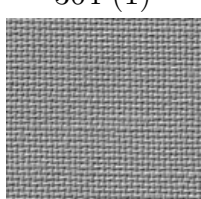
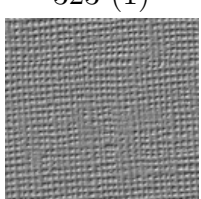
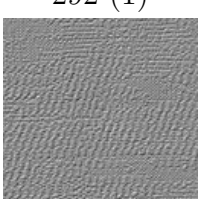
				
	054 (1)			
				
013	085 (6)	358 (3)	148 (3)	149 (2)
				
	138 (2)	004 (2)	493 (1)	474 (1)
				
	406 (1)	405 (1)	385 (1)	380 (1)
				
	369 (1)	368 (1)	249 (1)	188 (1)
				
	176 (1)	174 (1)	172 (1)	169 (1)
				
	147 (1)	121 (1)	034 (1)	

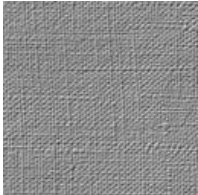
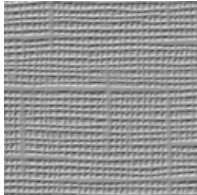
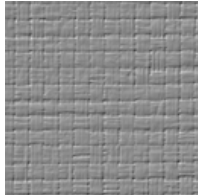
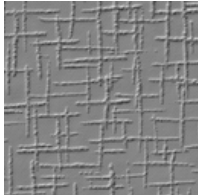
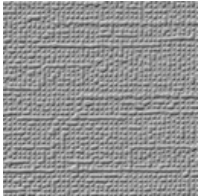
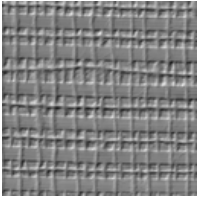
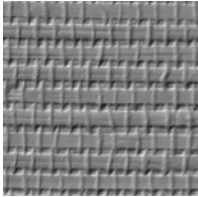
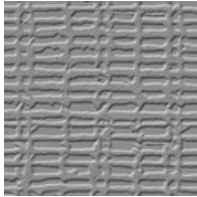
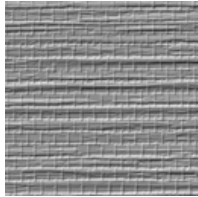
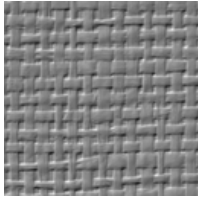
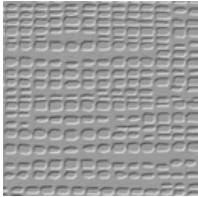
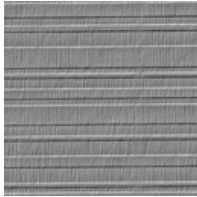
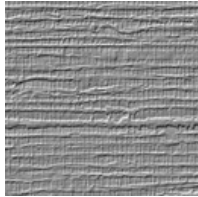
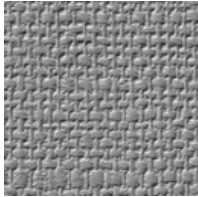
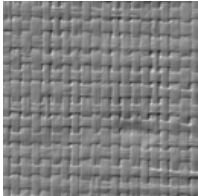
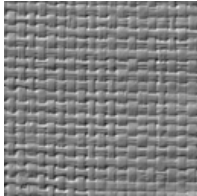
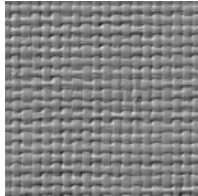
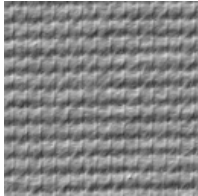
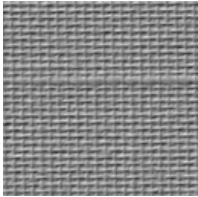
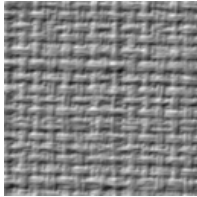
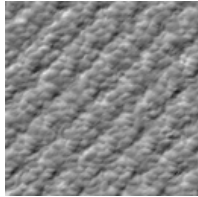
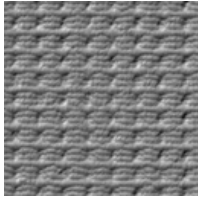
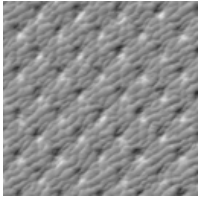
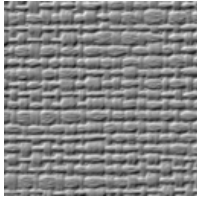
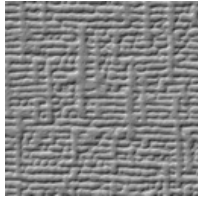
 022	 493 (10)	 081 (10)	 003 (3)	 365 (2)
	 046 (2)	 369 (1)	 148 (1)	 085 (1)
 026	 350 (9)	 135 (9)	 131 (7)	 418 (5)
	 352 (3)	 387 (2)	 083 (1)	 025 (1)
 045	 412 (8)	 486 (5)	 457 (4)	 200 (4)
	 196 (4)	 378 (3)	 276 (3)	 238 (2)

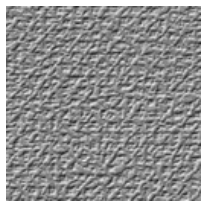
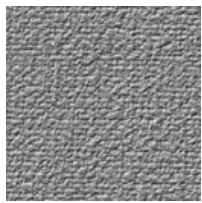
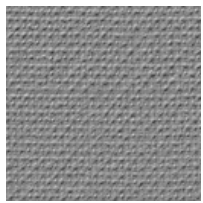
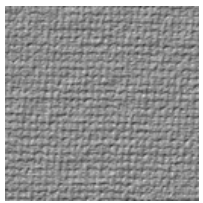
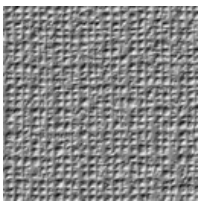
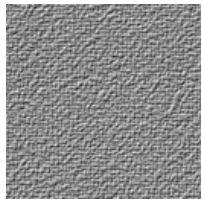
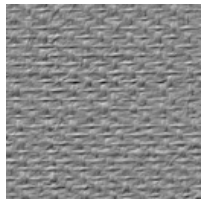
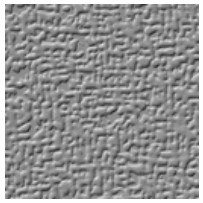
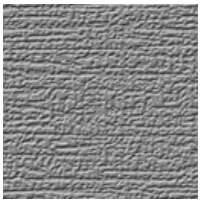
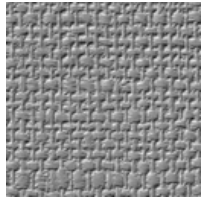
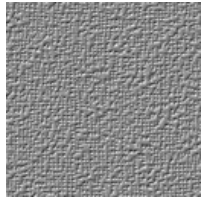
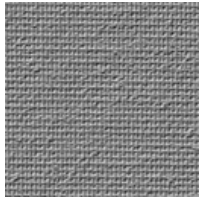
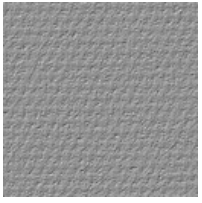
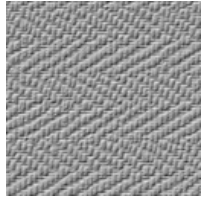
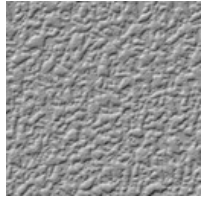
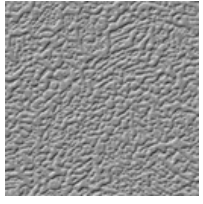
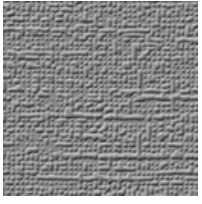
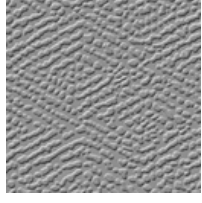
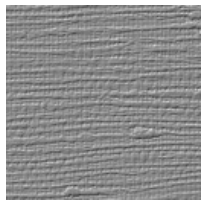
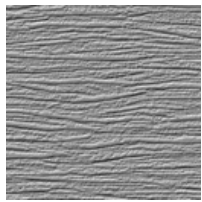
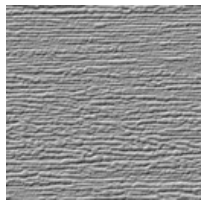
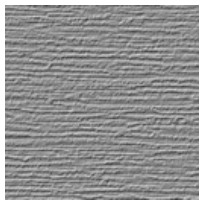
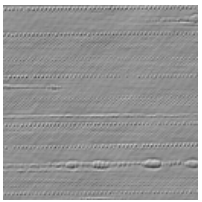
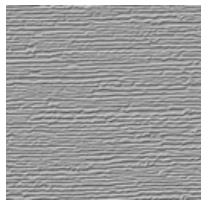


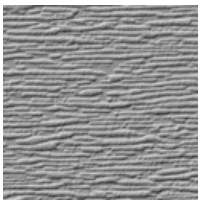
				
	052 (2)	470 (1)	430 (1)	401 (1)
				
	337 (1)	333 (1)	323 (1)	248 (1)
				
	235 (1)	213 (1)	209 (1)	174 (1)
				
	173 (1)	144 (1)	118 (1)	035 (1)
				
059	410 (9)	182 (8)	377 (7)	096 (5)
				
	479 (3)	407 (3)	473 (1)	348 (1)
				
	325 (1)	322 (1)	317 (1)	284 (1)

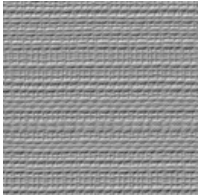
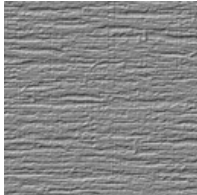
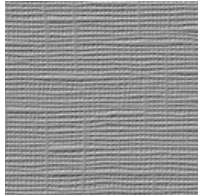
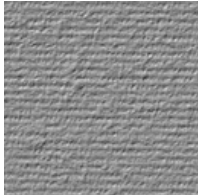
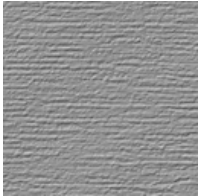
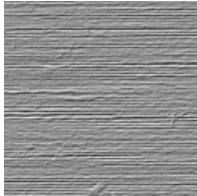
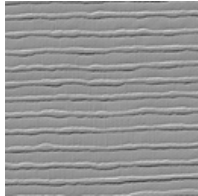
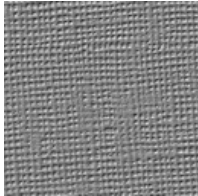
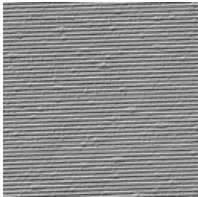
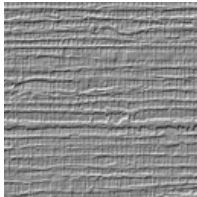
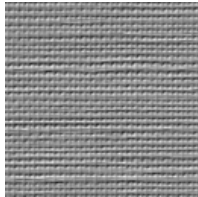
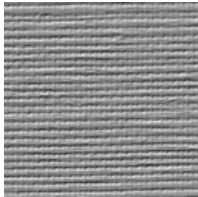
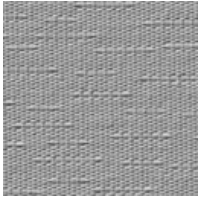
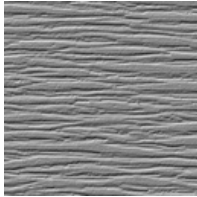
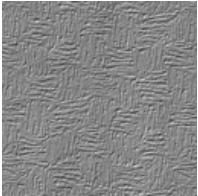
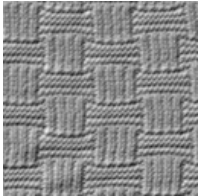
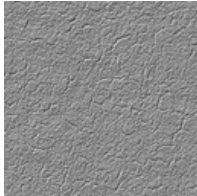
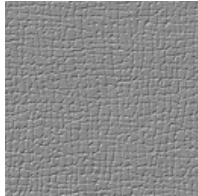
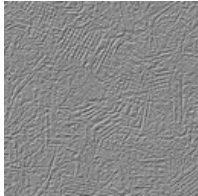
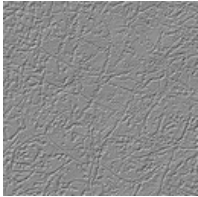
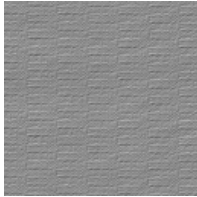
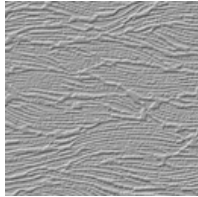
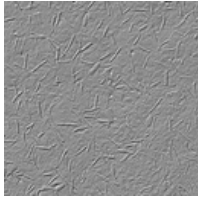
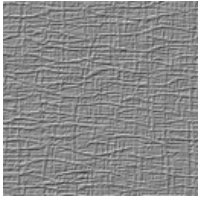
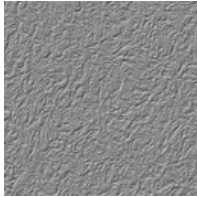
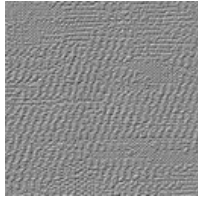
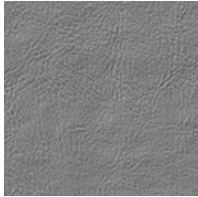
				
	048 (1)			
				
066	481 (8)	129 (8)	376 (6)	419 (4)
				
	354 (2)	166 (2)	162 (2)	416 (1)
				
	402 (1)	302 (1)	275 (1)	246 (1)
				
	119 (1)	074 (1)	062 (1)	
				
091	297 (5)	340 (3)	089 (3)	462 (2)
				
	426 (2)	076 (2)	467 (1)	438 (1)

				
	395 (1)	373 (1)	332 (1)	323 (1)
				
	318 (1)	310 (1)	248 (1)	161 (1)
				
	112 (1)	092 (1)	088 (1)	080 (1)
				
	077 (1)	063 (1)	033 (1)	006 (1)
				
099	356 (10)	030 (10)	133 (6)	482 (4)
				
	417 (4)	347 (4)	082 (3)	132 (2)
				
	079 (2)	054 (2)	366 (1)	015 (1)

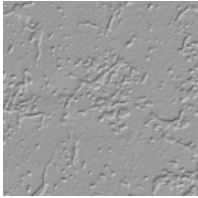
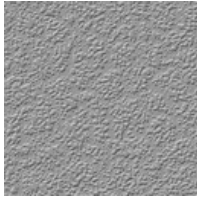
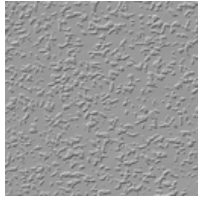
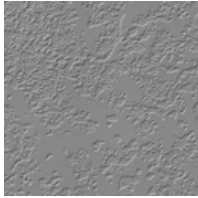
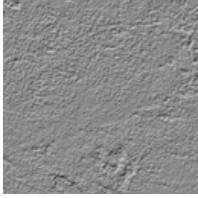
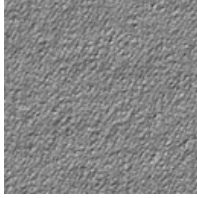
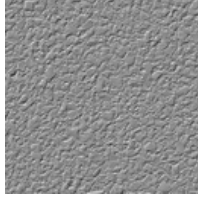
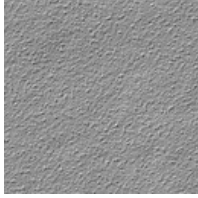
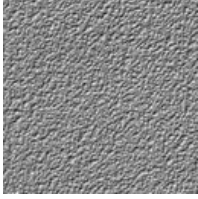
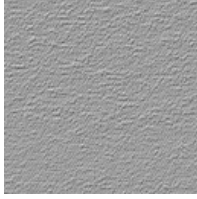
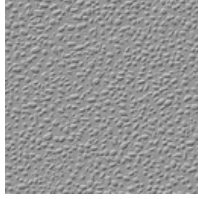
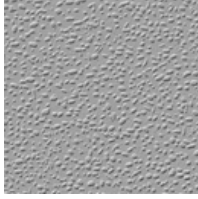
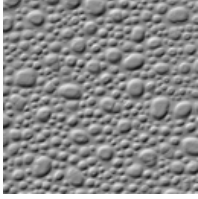
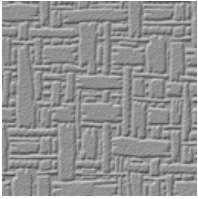
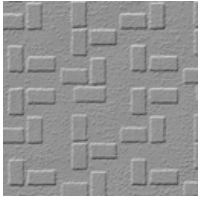
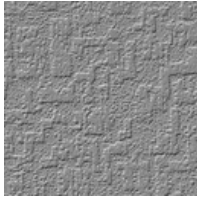
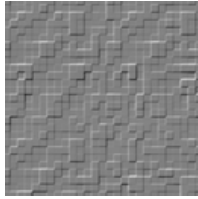
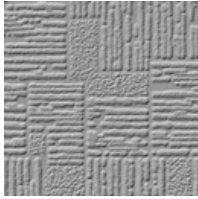
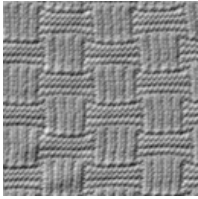
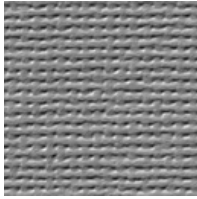
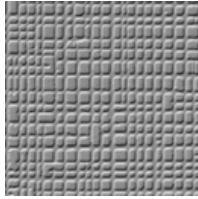
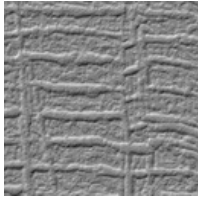
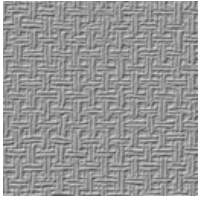
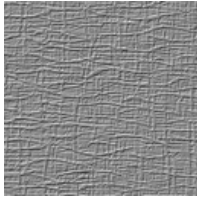
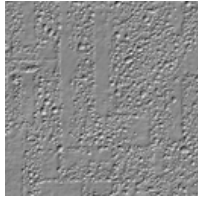
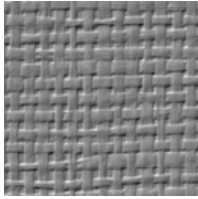
				
106	142 (9)	460 (3)	307 (3)	228 (3)
				
	456 (2)	233 (2)	157 (2)	036 (2)
				
	476 (1)	232 (1)	229 (1)	
				
117	412 (6)	378 (6)	466 (4)	118 (4)
				
	058 (3)	276 (2)	052 (2)	457 (1)
				
	388 (1)	364 (1)	323 (1)	292 (1)
				
	238 (1)	235 (1)	226 (1)	213 (1)

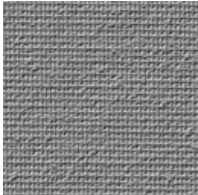
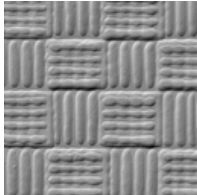
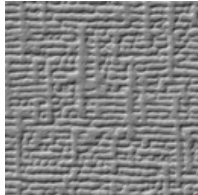

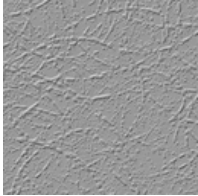
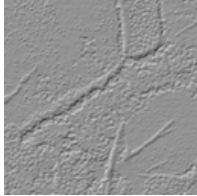
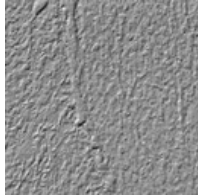
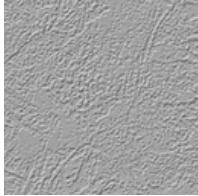
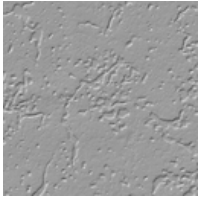
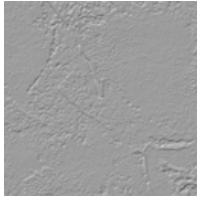
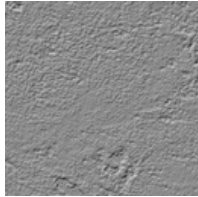
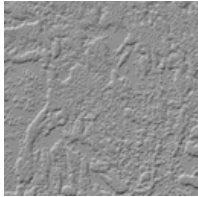
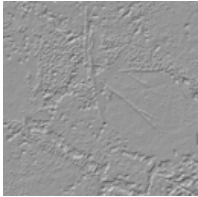
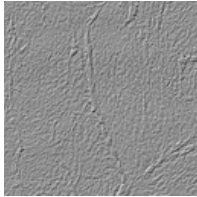
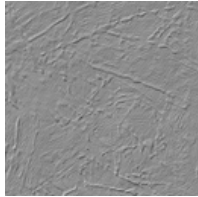
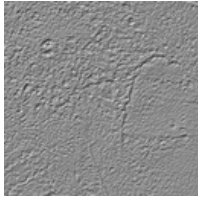
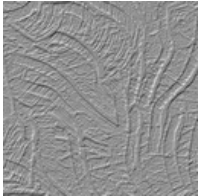
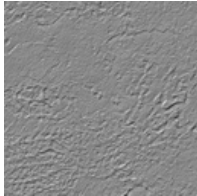
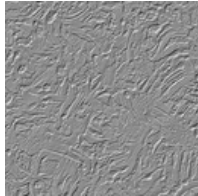
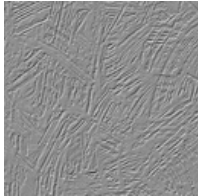
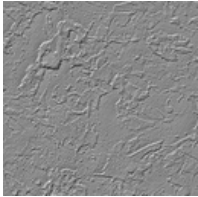
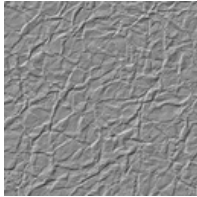
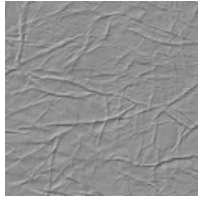
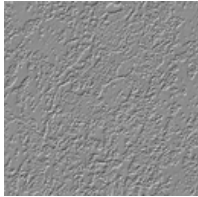
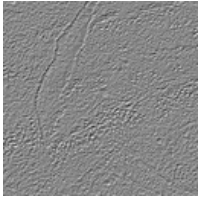
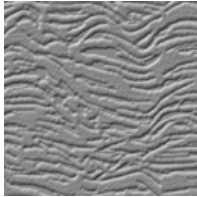
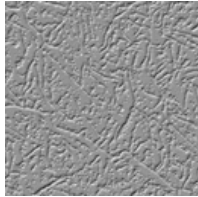
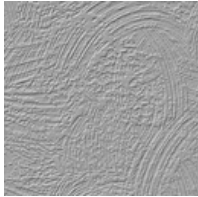
				
	200 (1)	191 (1)	169 (1)	109 (1)
				
	035 (1)			
				
177	178 (5)	004 (4)	394 (3)	176 (3)
				
	360 (1)	268 (1)	216 (1)	211 (1)
				
	175 (1)	172 (1)	168 (1)	149 (1)
				
	141 (1)	138 (1)	094 (1)	093 (1)
				
	090 (1)	039 (1)	007 (1)	

				
207	297 (7)	439 (2)	438 (2)	259 (2)
				
	239 (2)	491 (1)	489 (1)	451 (1)
				
	211 (1)	205 (1)	107 (1)	102 (1)
				
	086 (1)	080 (1)	024 (1)	006 (1)
				
	005 (1)			
				
215	458 (4)	069 (4)	470 (3)	217 (3)
				
	414 (2)	296 (2)	158 (2)	477 (1)

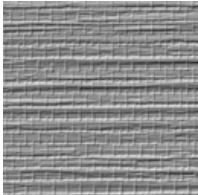
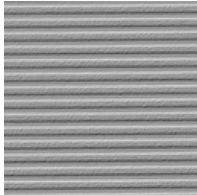
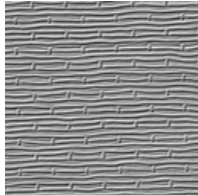
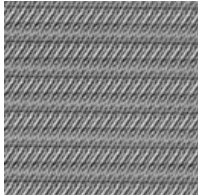
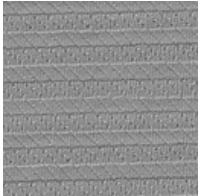
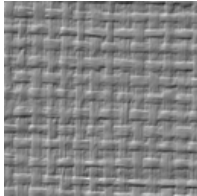
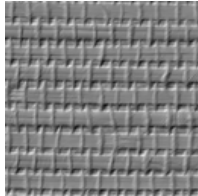
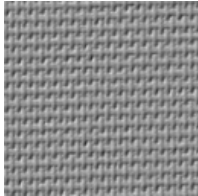
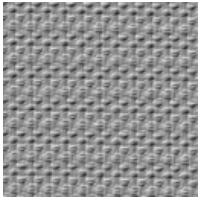
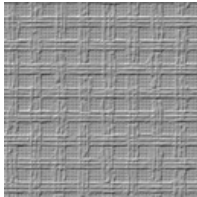
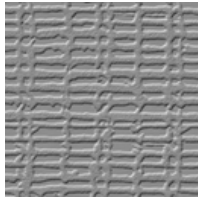
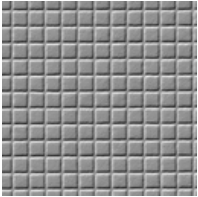
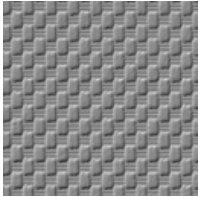
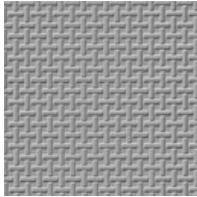
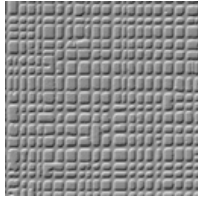
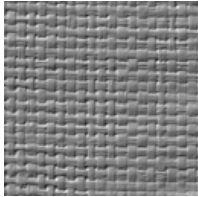
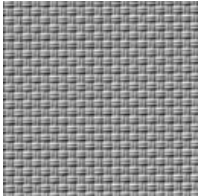
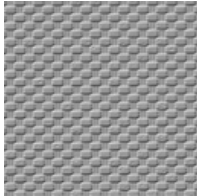
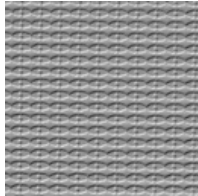
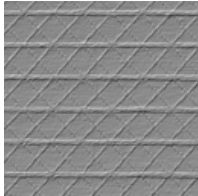
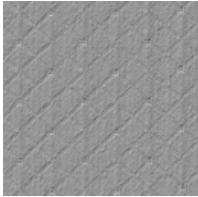
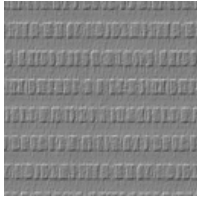
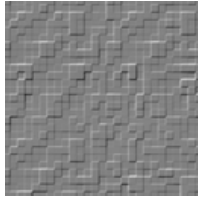
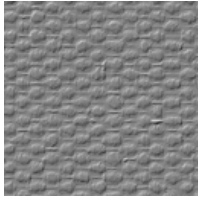
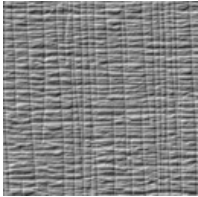
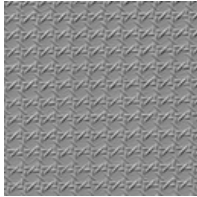
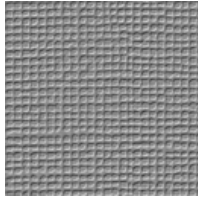
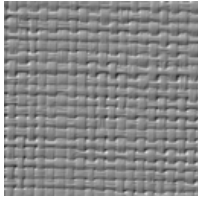
				
	409 (1)	395 (1)	378 (1)	332 (1)
				
	251 (1)	237 (1)	233 (1)	226 (1)
				
	220 (1)	216 (1)	181 (1)	167 (1)
				
	152 (1)	002 (1)		
				
257	493 (3)	299 (3)	292 (2)	288 (2)
				
	285 (2)	252 (2)	349 (1)	284 (1)
				
	282 (1)	219 (1)	213 (1)	206 (1)

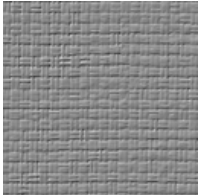
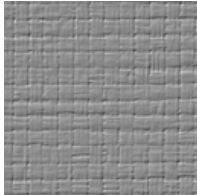
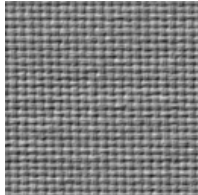
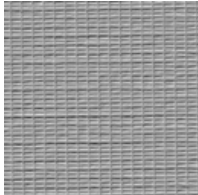
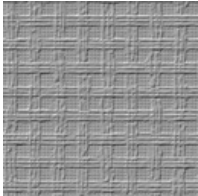
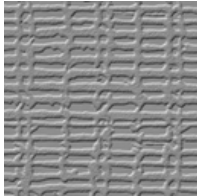
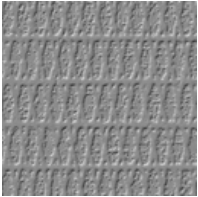
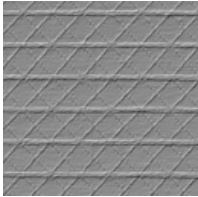
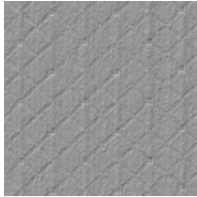
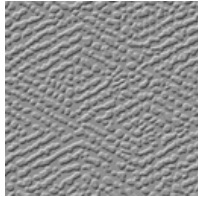
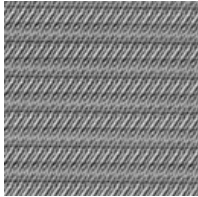
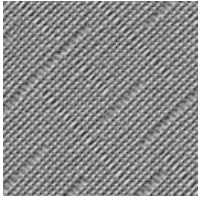
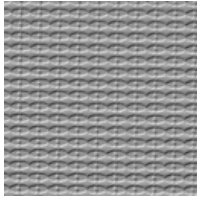
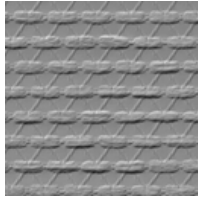
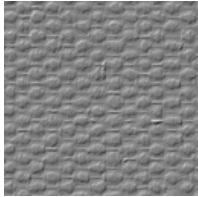
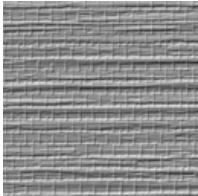
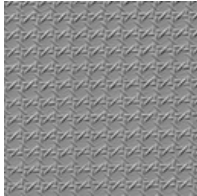
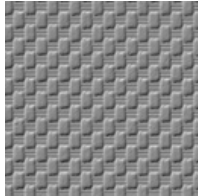
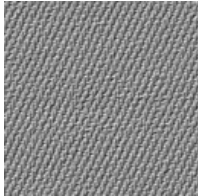
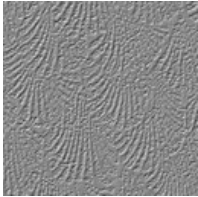
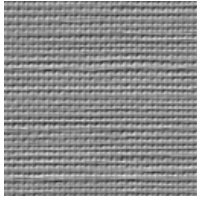
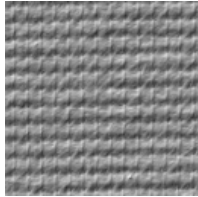
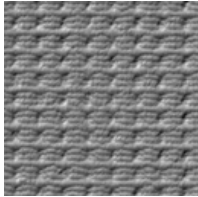
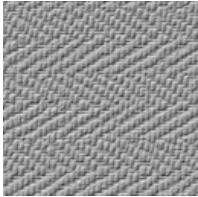
	127 (1)	120 (1)	116 (1)	102 (1)
	085 (1)			
283	266 (9)	208 (8)	313 (6)	302 (5)
	303 (3)	116 (2)	415 (1)	318 (1)
	282 (1)	267 (1)	127 (1)	
298	227 (6)	294 (4)	245 (3)	379 (2)
	301 (2)	192 (2)	484 (1)	473 (1)

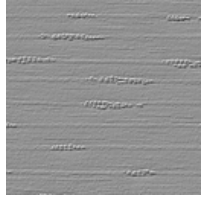
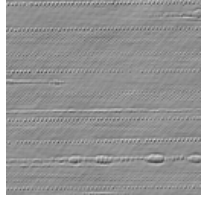
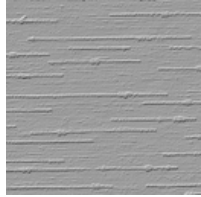
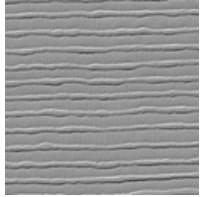
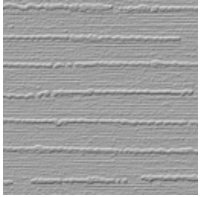
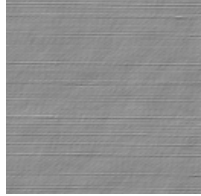
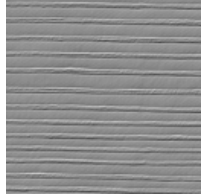
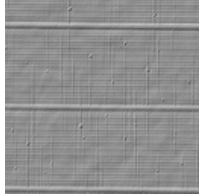
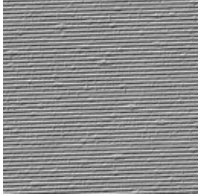
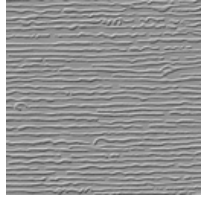
				
	443 (1)	413 (1)	410 (1)	402 (1)
				
	370 (1)	340 (1)	261 (1)	212 (1)
				
	210 (1)	153 (1)	057 (1)	051 (1)
				
	040 (1)			
				
361	159 (7)	068 (6)	405 (3)	365 (2)
				
	493 (1)	391 (1)	385 (1)	383 (1)
				
	358 (1)	282 (1)	280 (1)	176 (1)

				
	109 (1)	081 (1)	007 (1)	
				
371	190 (4)	374 (3)	320 (3)	050 (3)
				
	443 (2)	408 (2)	370 (2)	355 (2)
				
	351 (2)	329 (2)	293 (2)	227 (2)
				
	387 (1)	301 (1)	281 (1)	277 (1)
				
	271 (1)	266 (1)	208 (1)	203 (1)
				
	192 (1)	125 (1)	116 (1)	049 (1)

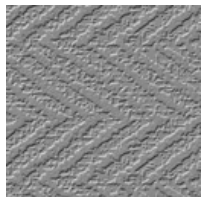
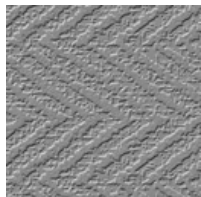
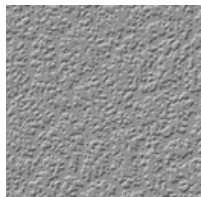
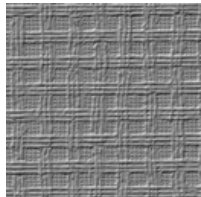
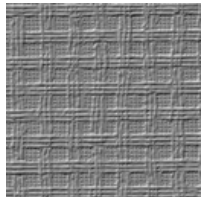
	018 (1)	010 (1)			
399	400 (8)	393 (6)	476 (4)	459 (3)	
	296 (2)	232 (2)	002 (2)	460 (1)	
	458 (1)	362 (1)	186 (1)		
425	369 (3)	432 (2)	406 (2)	149 (2)	
	493 (1)	442 (1)	440 (1)	428 (1)	
	427 (1)	422 (1)	421 (1)	398 (1)	

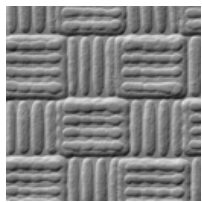
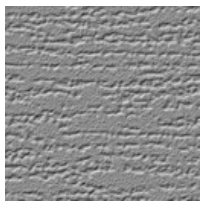
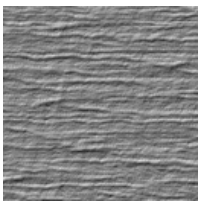
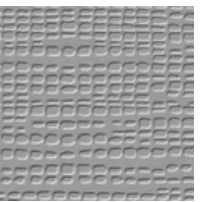
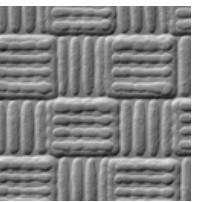
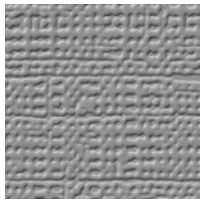
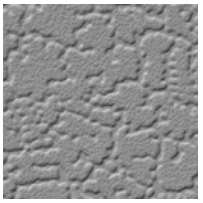
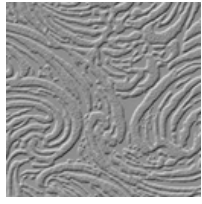
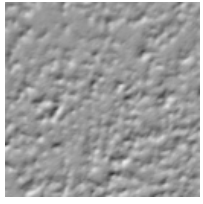
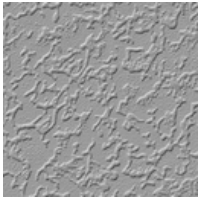
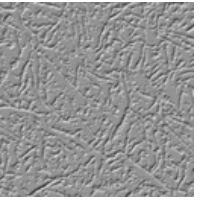
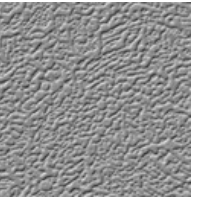
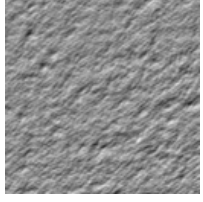
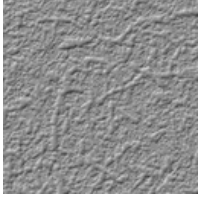
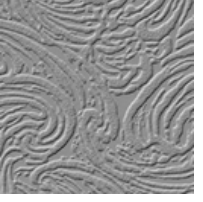
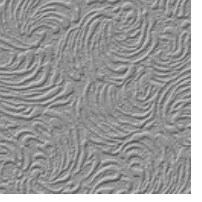
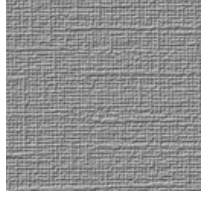
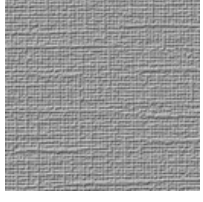
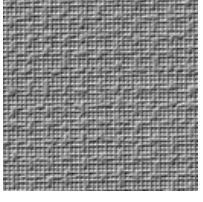
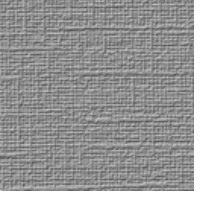
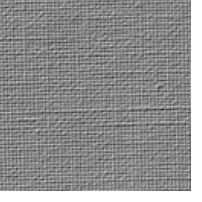
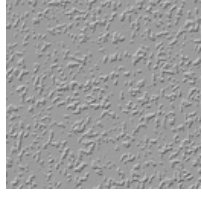
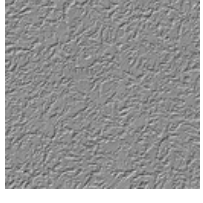
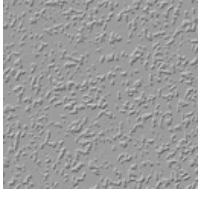
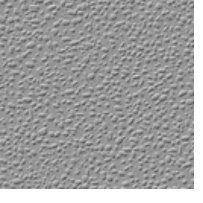
				
	394 (1)	362 (1)	309 (1)	249 (1)
				
	241 (1)	179 (1)	178 (1)	156 (1)
				
	151 (1)	085 (1)	004 (1)	
				
444	369 (6)	188 (3)	385 (2)	172 (2)
				
	103 (2)	046 (2)	449 (1)	437 (1)
				
	436 (1)	406 (1)	405 (1)	398 (1)
				
	382 (1)	380 (1)	368 (1)	180 (1)

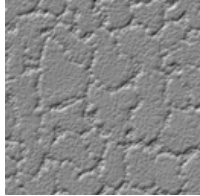
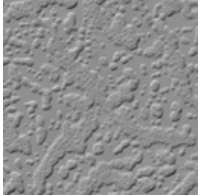
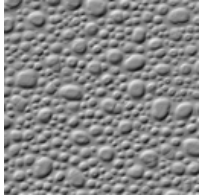
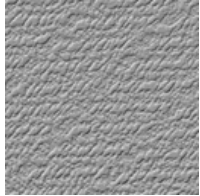
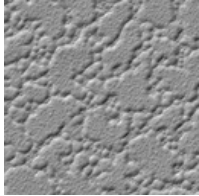
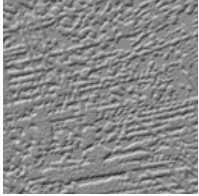
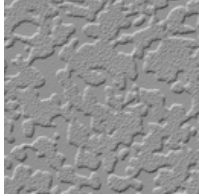
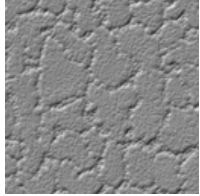
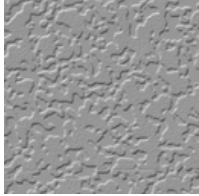
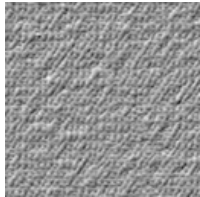
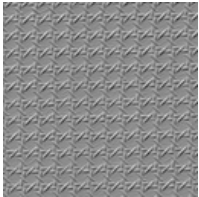
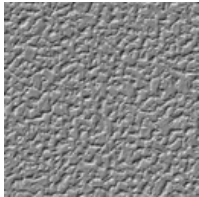
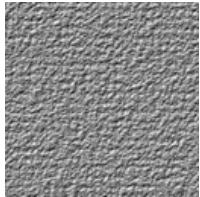
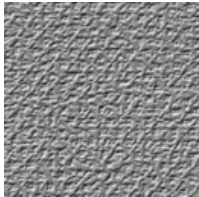
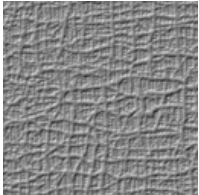
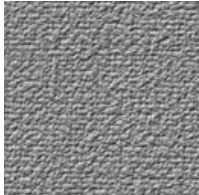
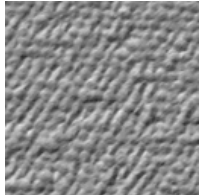
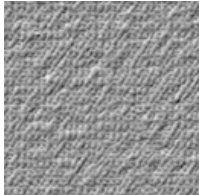
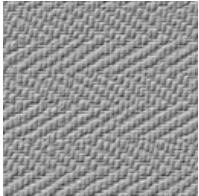
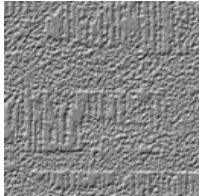
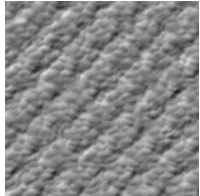
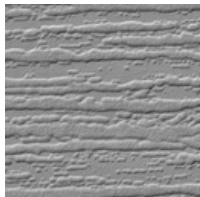
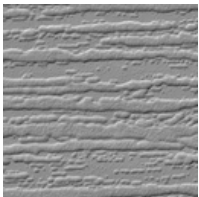
				
	174 (1)	169 (1)	143 (1)	100 (1)
				
	085 (1)	004 (1)		
				
448	437 (7)	436 (3)	005 (3)	249 (2)
				
	242 (2)	449 (1)	432 (1)	398 (1)
				
	394 (1)	380 (1)	369 (1)	330 (1)
				
	295 (1)	181 (1)	149 (1)	093 (1)
				
	086 (1)			

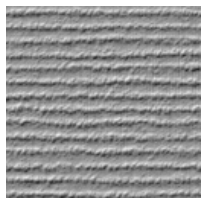
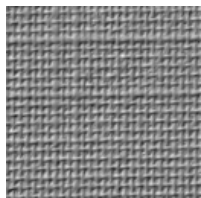
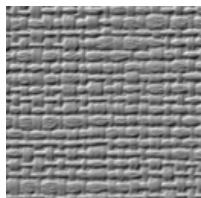
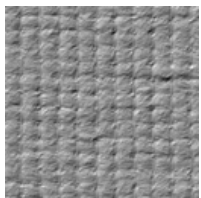
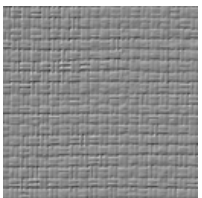
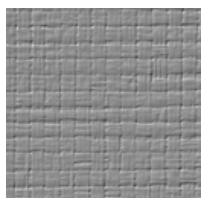
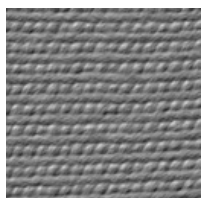
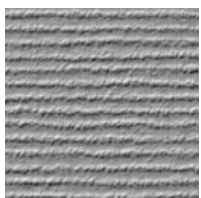
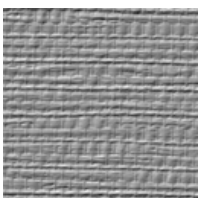
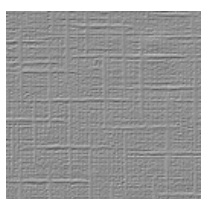
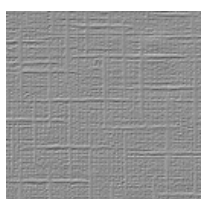
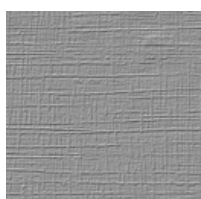
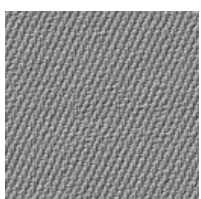
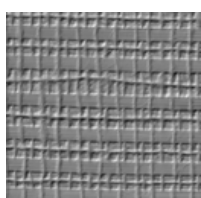
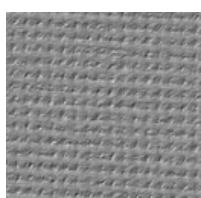
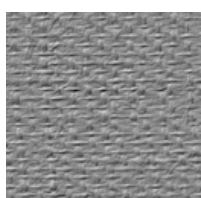
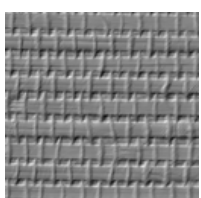
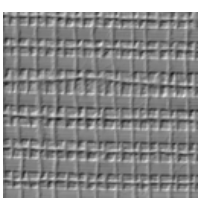
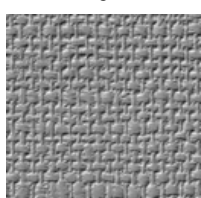
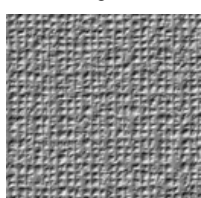
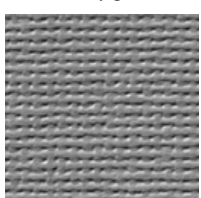
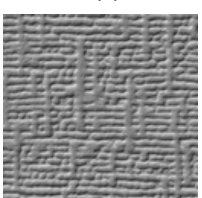
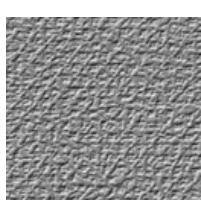
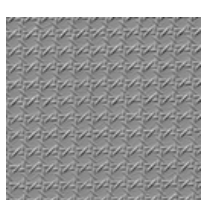
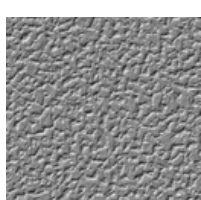
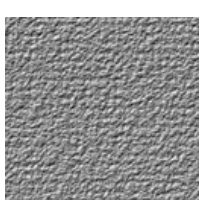
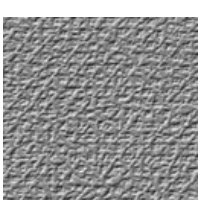
				
485	217 (9)	187 (8)	233 (3)	055 (3)
				
	321 (2)	308 (1)	260 (1)	220 (1)
				
	056 (1)			

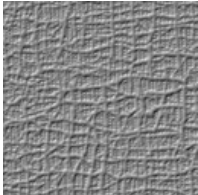
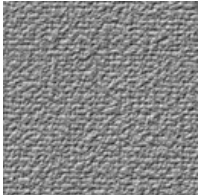
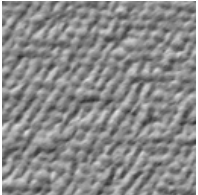
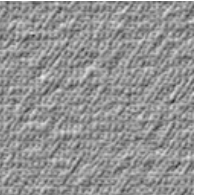
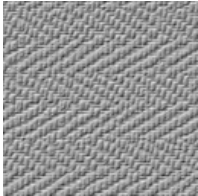
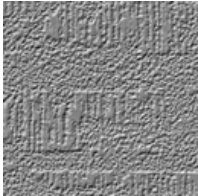
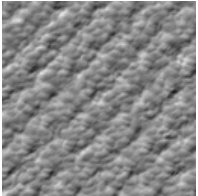
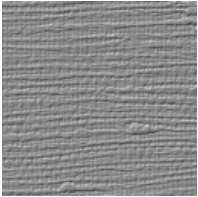
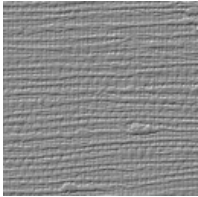
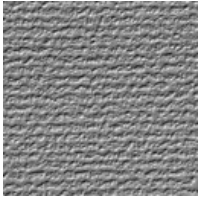
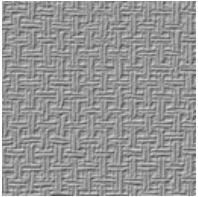
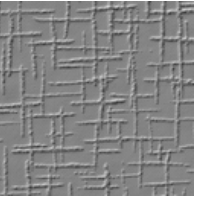
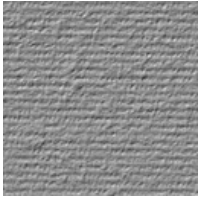
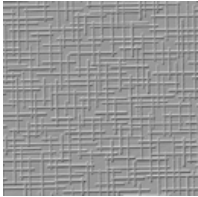
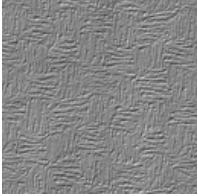
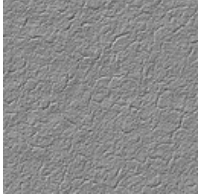
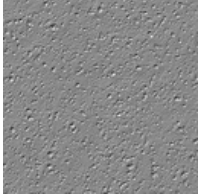
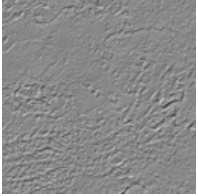
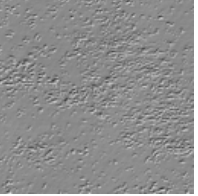
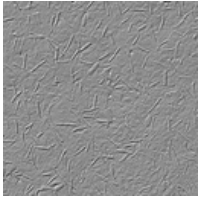
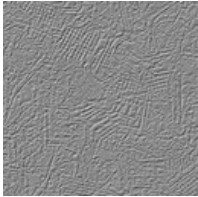
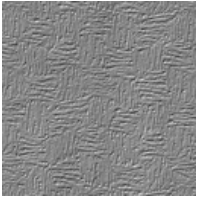
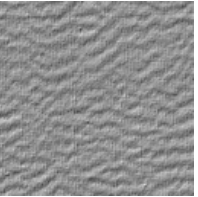
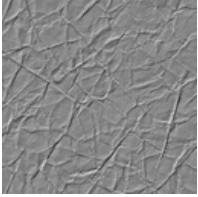
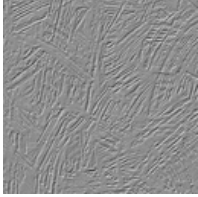
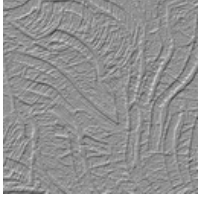
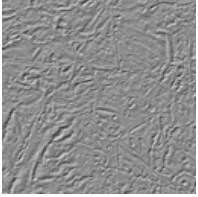
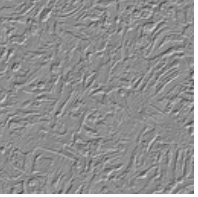
B.3 Neurons Containing *Tex500[024]* Stimuli in SOM_F Browsing Environment

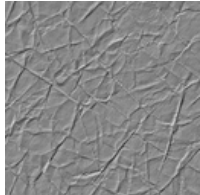
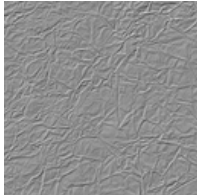
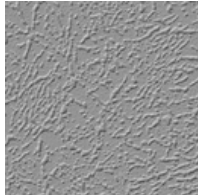
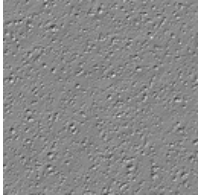
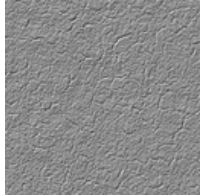
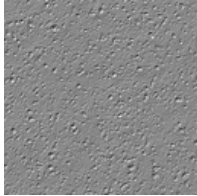
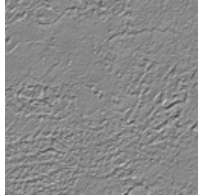
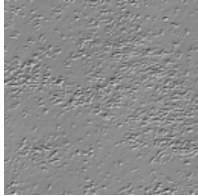
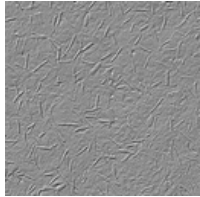
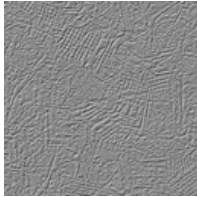
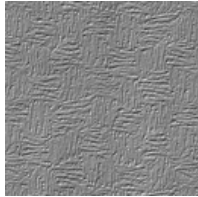
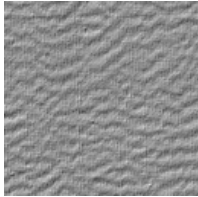
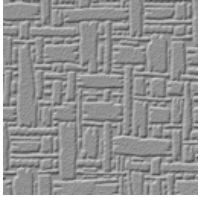
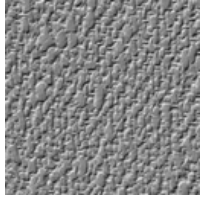
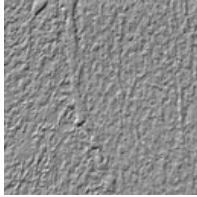
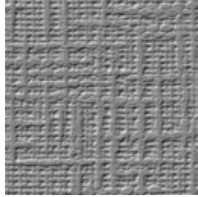
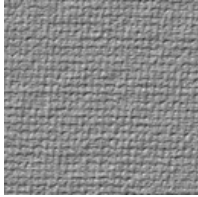
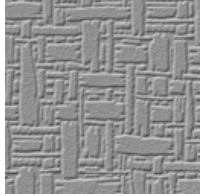
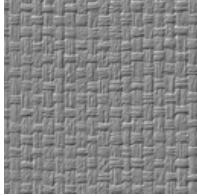
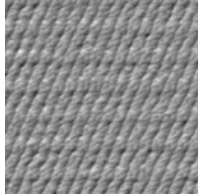
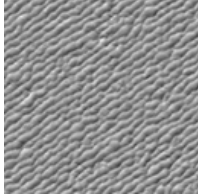
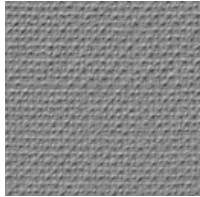
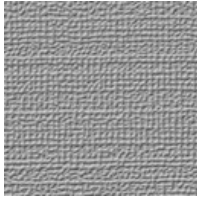
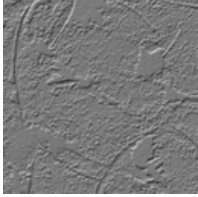
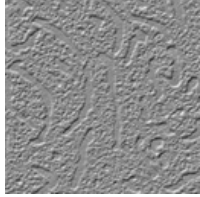
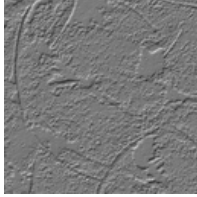
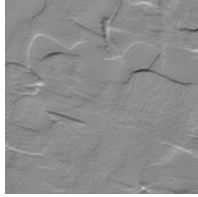
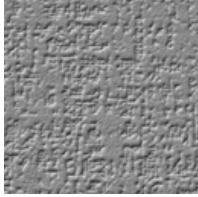
Stimuli	Neuron Contents (Starting with Neuron Exemplar)	
		
008	008	344
		
013	013	

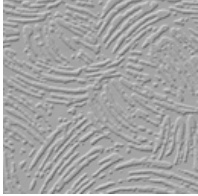
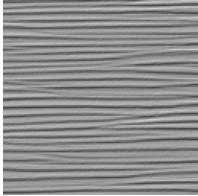
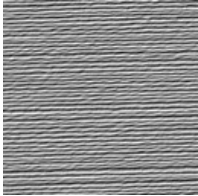
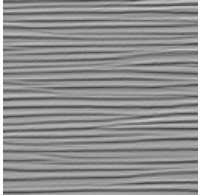
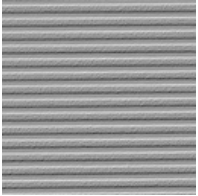
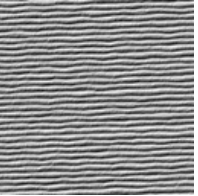
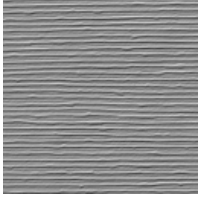
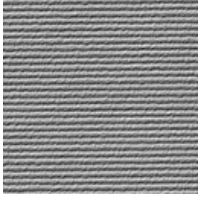
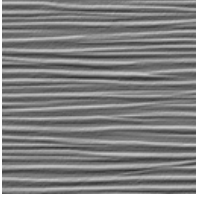
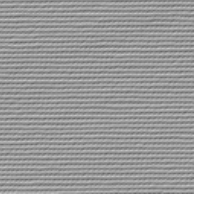
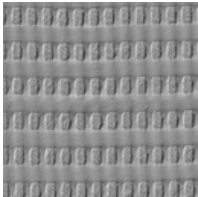
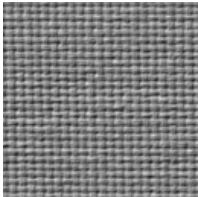
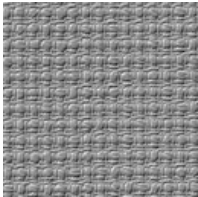
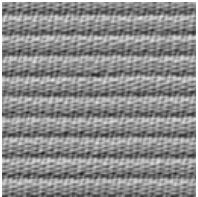
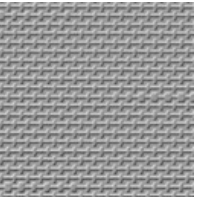
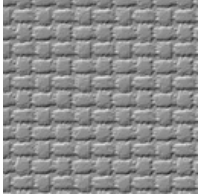
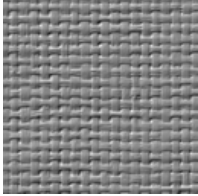
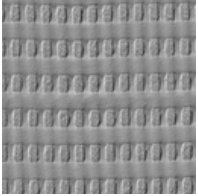
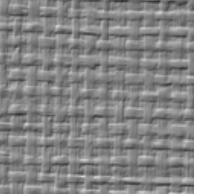
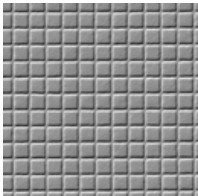
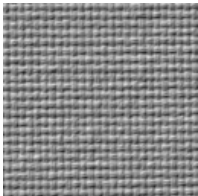
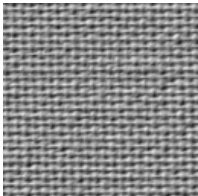
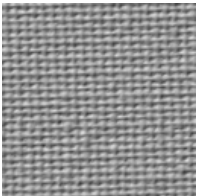
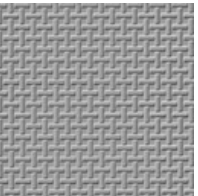
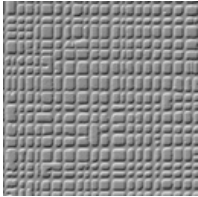
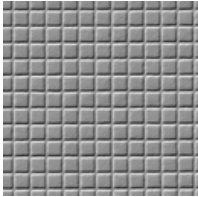
				
022	375	429	360	022
				
	447	376		
				
026	490	479	116	024
				
	087	023	026	135
				
045	412	263	045	323
				
059	246	059	057	

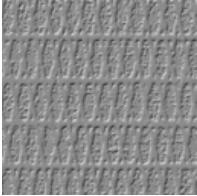
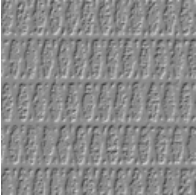
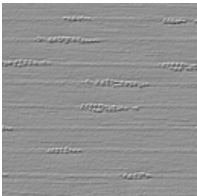
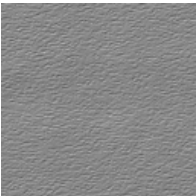
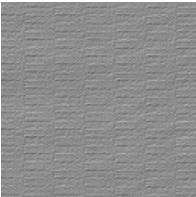
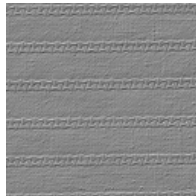
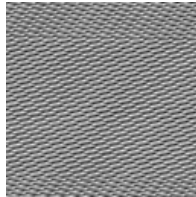
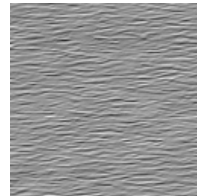
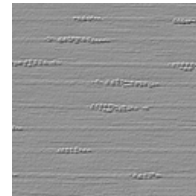
 066	 083	 040	 467	 129
	 027	 354	 066	 416
 091	 380	 225	 240	 207
	 060	 297	 089	 091
	 086	 278	 094	
 099	 099			

				
106	336	039	150	174
				
	169	497	106	331
				
117	117	118	330	
				
177	101	491	178	177
				
	211	259	391	007
				
207	380	225	240	207

				
	060	297	089	091
				
	086	278	094	
				
215	215	305	358	109
				
	332	185		
				
257	299	298	301	245
				
	284	288	257	312
				
283	277	387	318	281

				
	283	313	346	
				
298	299	298	301	245
				
	284	288	257	312
				
361	310	320	401	438
				
	361	214	088	092
				
	439	357		
				
371	018	371	384	472

					
	418				
					
399	236	399	362	232	
					
	287	194	400	306	
					
425	143	198	137	363	
					
	121	168	425	179	
					
444	342	139	338	188	
					
	385	444			

 <p>448</p>	 <p>448</p>			
 <p>485</p>	 <p>199</p>  <p>252</p>	 <p>262</p>  <p>289</p>	 <p>319</p>	 <p>485</p>

B.4 SOM_G, SOM_A & SOM_F Browsing Environments Comparison

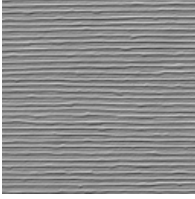
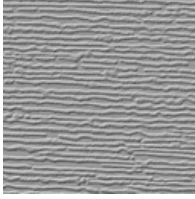
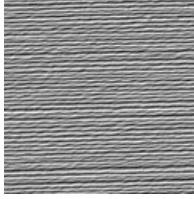
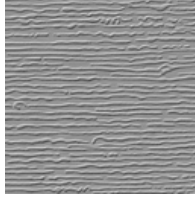
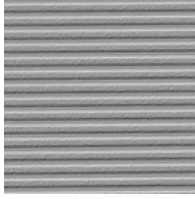
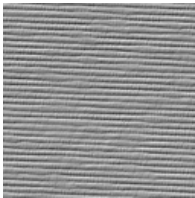
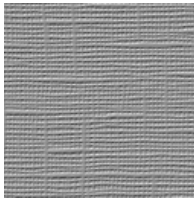
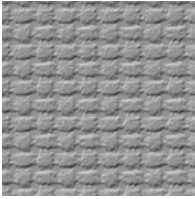
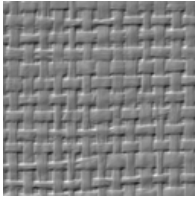
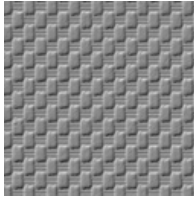
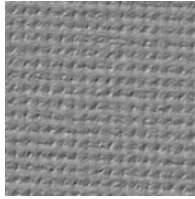
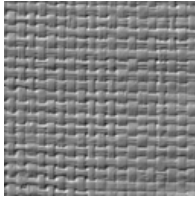
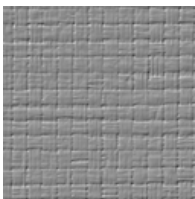
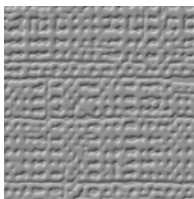
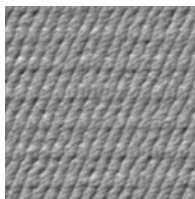
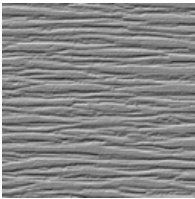
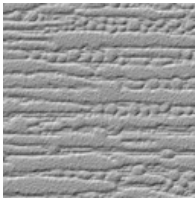
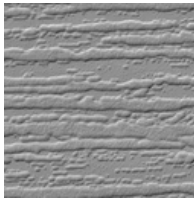
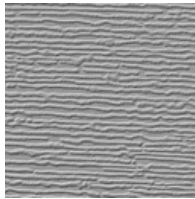
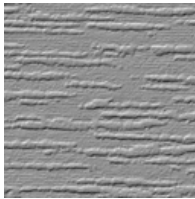
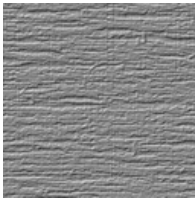
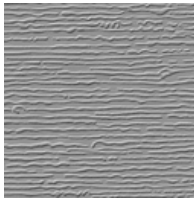
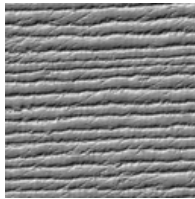
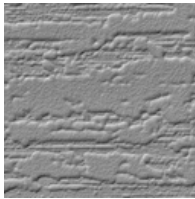
Observer	Mean Task Time		
	SOM _G	SOM _A	SOM _F
1	47.000	143.375	209.375
2	33.500	81.125	115.250
3	71.500	32.125	160.625
4	48.000	53.125	72.125
5	83.000	235.625	226.250
6	102.875	165.500	148.375
7	94.500	39.500	162.000
8	97.125	46.500	146.500
9	48.375	123.375	187.625
10	97.500	50.500	290.875
11	109.375	48.750	213.250
12	47.750	113.500	213.375
Means	73.375	94.417	178.802

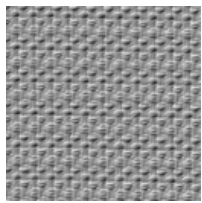
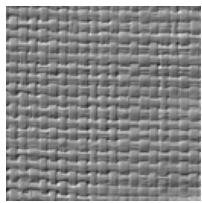
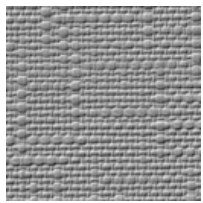
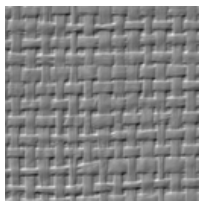
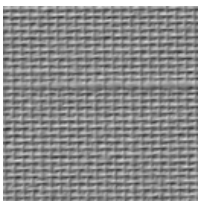
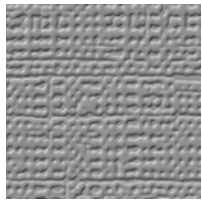
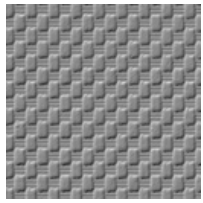
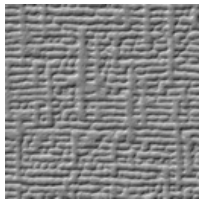
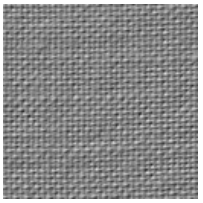
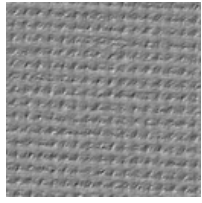
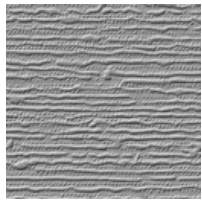
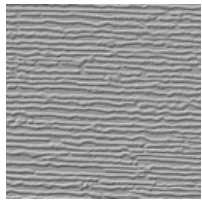
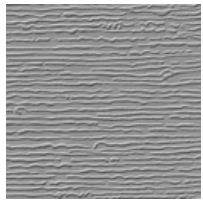
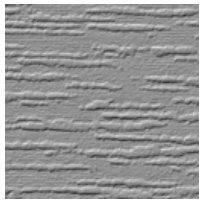
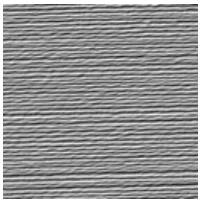
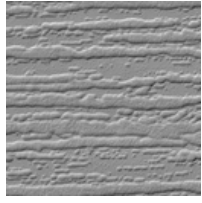
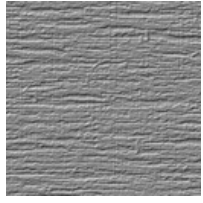
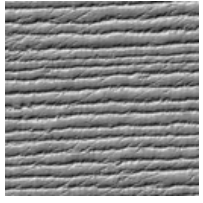
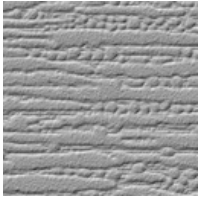
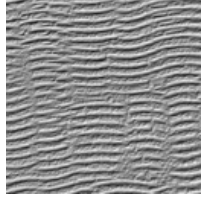
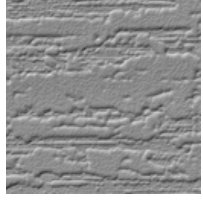
B.5 Crowdsourcing Dataset Augmentation

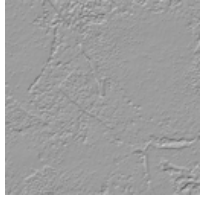
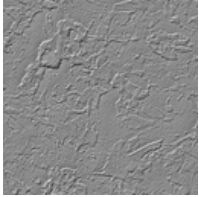
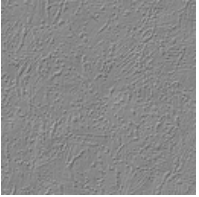
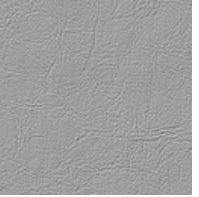
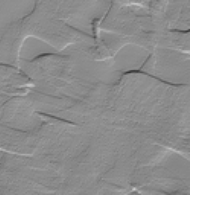
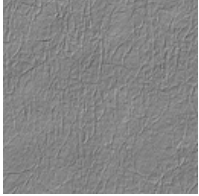
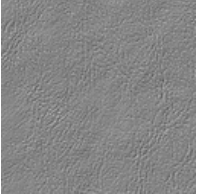
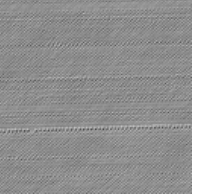
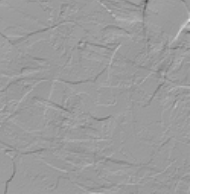
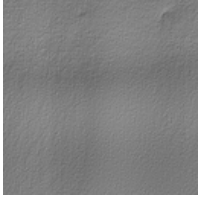
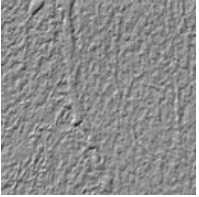
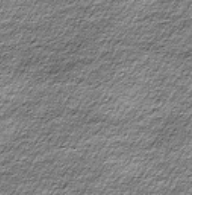
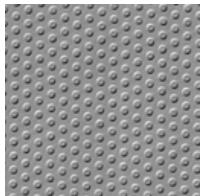
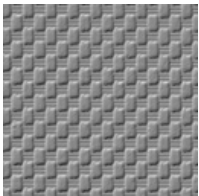
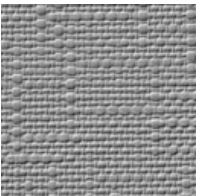
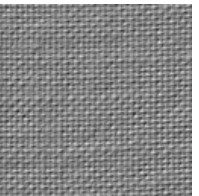
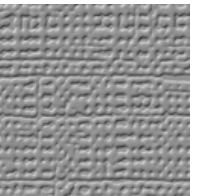
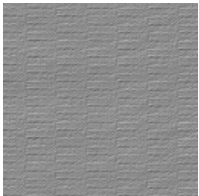
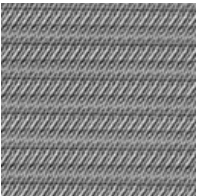
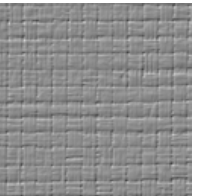
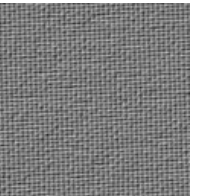
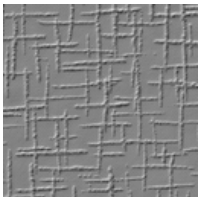
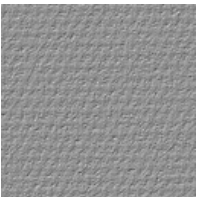
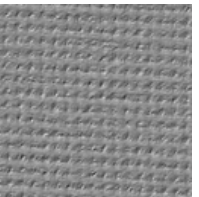
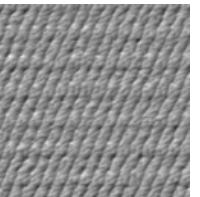
B.5.1 Acceptable Response Examples

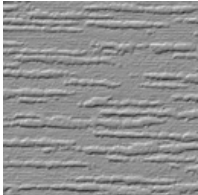
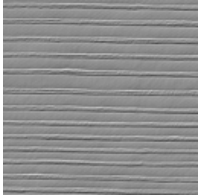
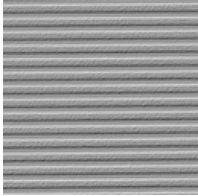
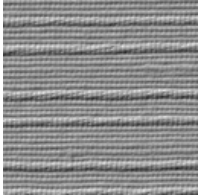
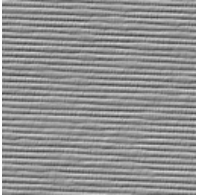
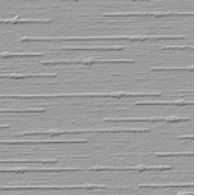
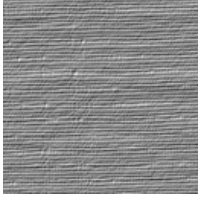
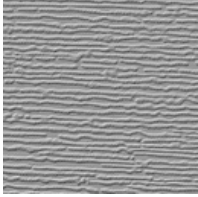
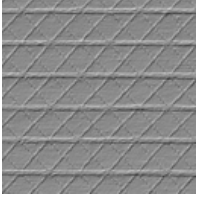
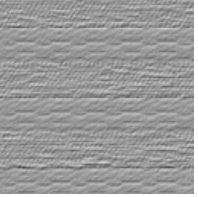
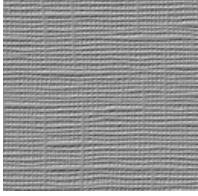
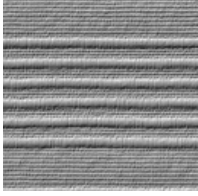
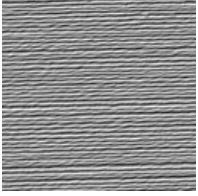
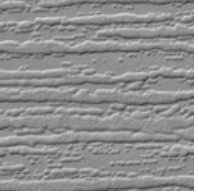
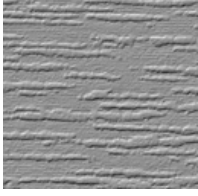
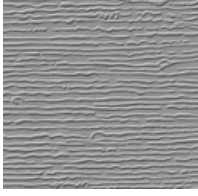
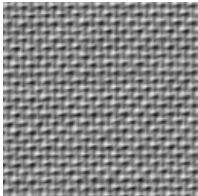
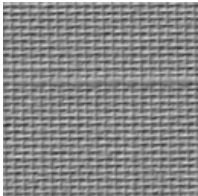
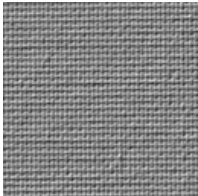
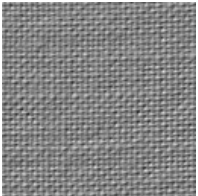
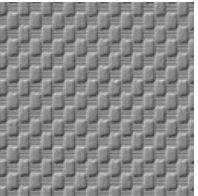
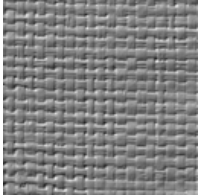
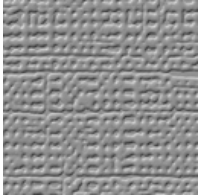
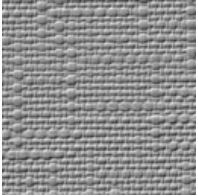
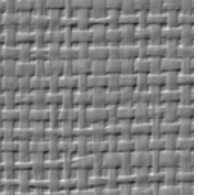
As each of the 400 stimuli in this experiment elicited between 6 and 28 unique responses, it would be much too space consuming to illustrate the complete result here. Instead, one randomly selected example response is shown for each response set size to illustrate the full range of observer agreement found in the experiment. The possible unique response sizes are 6–21 (inclusive), 23, and 28.

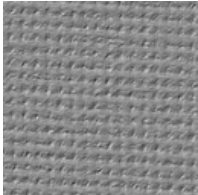
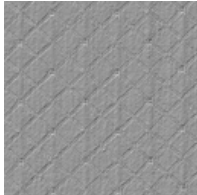
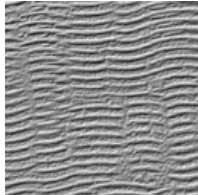
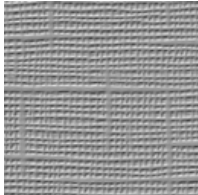
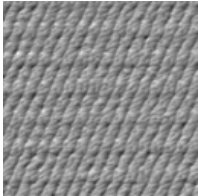
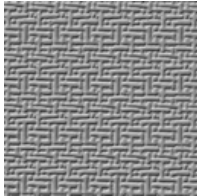
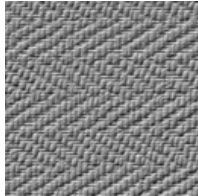
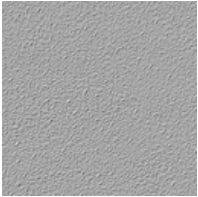
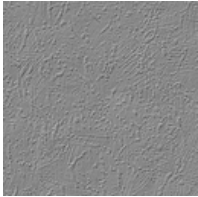
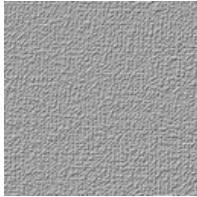
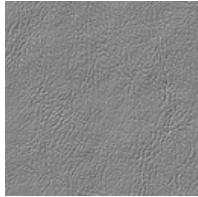
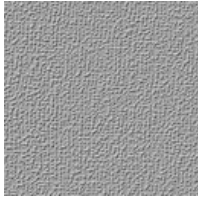
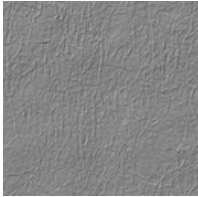
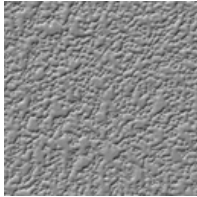
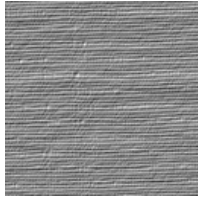
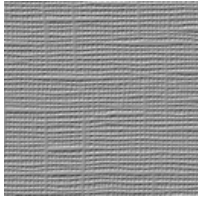
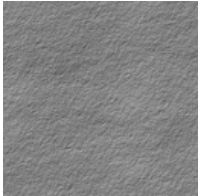
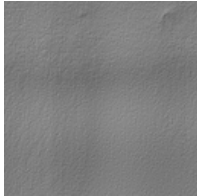
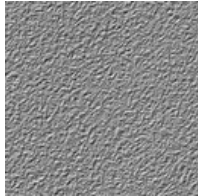
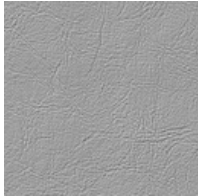
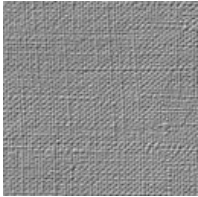
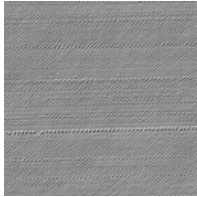
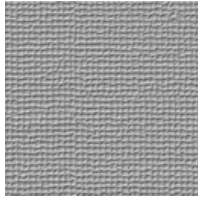
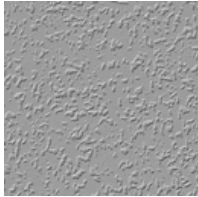
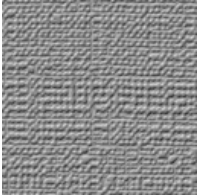
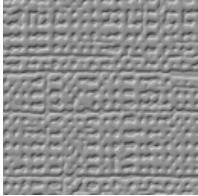
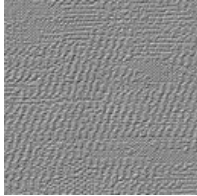
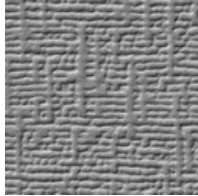
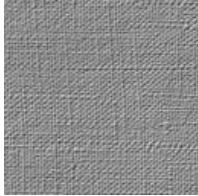
Note: Figure in brackets denotes number of votes received.

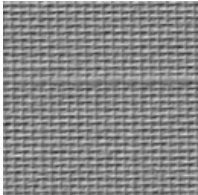
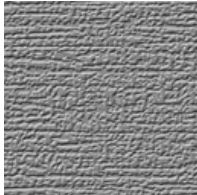
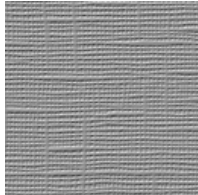
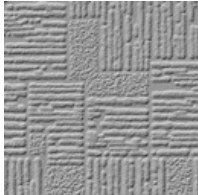
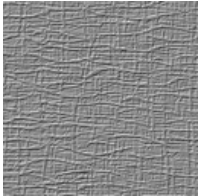
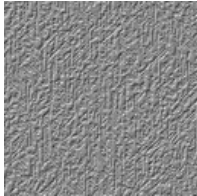
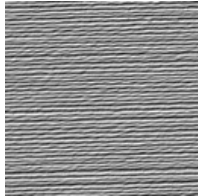
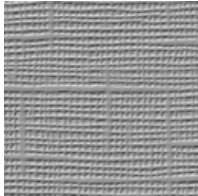
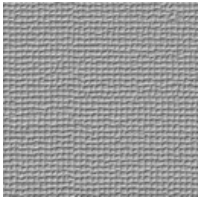
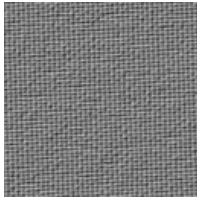
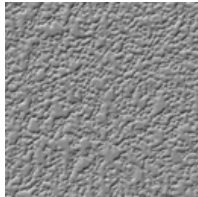
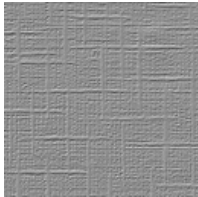
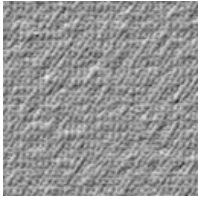
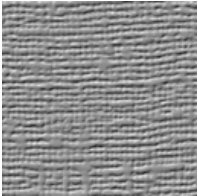
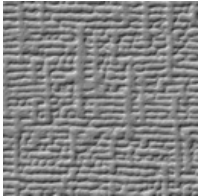
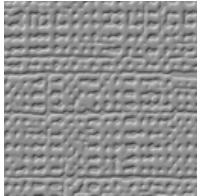
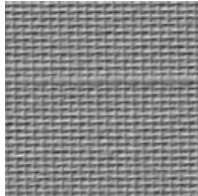
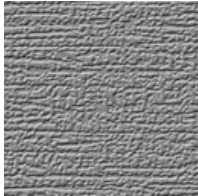
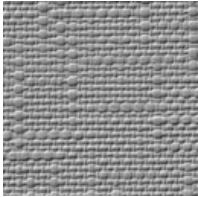
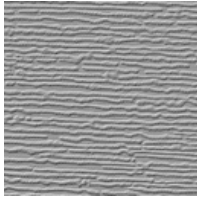
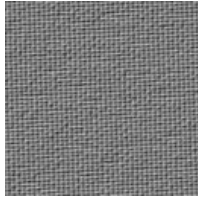
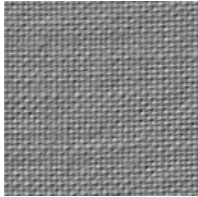
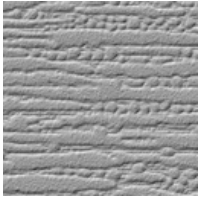
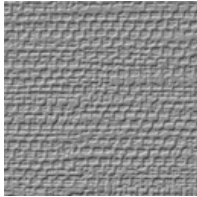
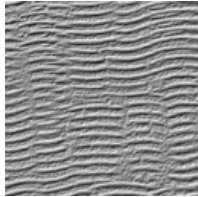
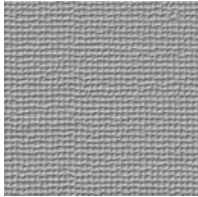
Stimuli	Responses				
 287	 345 (10)	 236 (9)	 056 (8)	 362 (2)	
	 221 (2)	 378 (1)			
 474	 176 (9)	 369 (7)	 101 (3)	 172 (2)	
	 169 (2)	 447 (1)	 088 (1)		
 002	 482 (7)	 099 (6)	 345 (5)	 079 (5)	
	 395 (3)	 056 (2)	 228 (1)	 095 (1)	

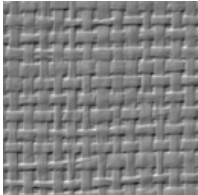
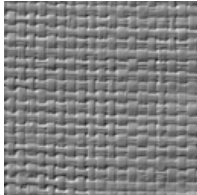
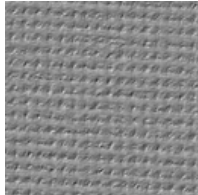
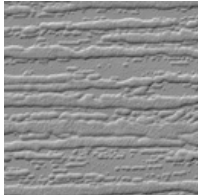
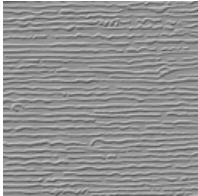
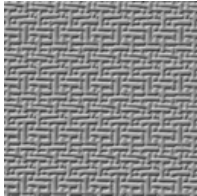
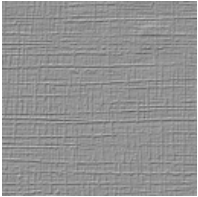
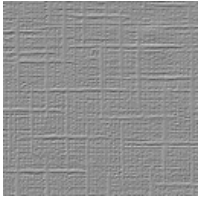
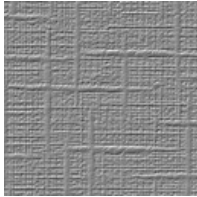
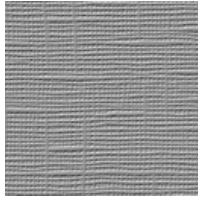
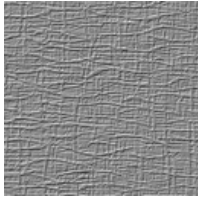
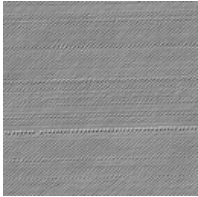
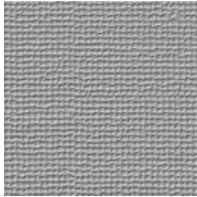
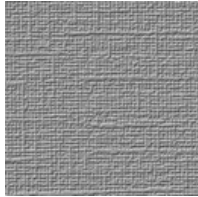
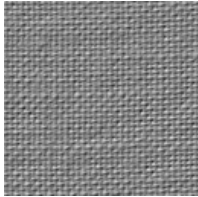
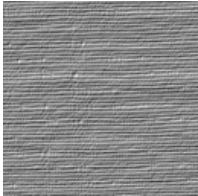
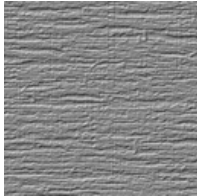
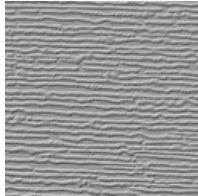
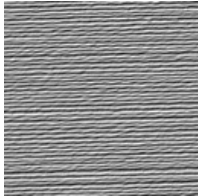
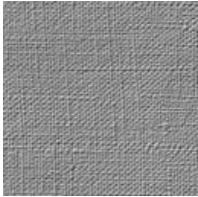
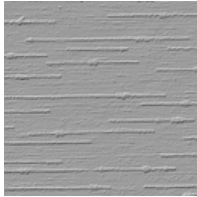
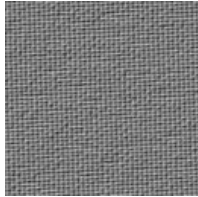
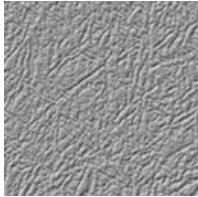
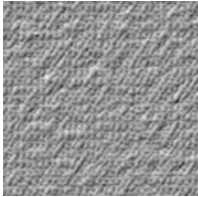
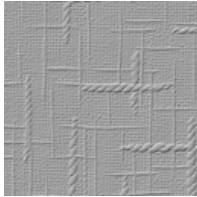
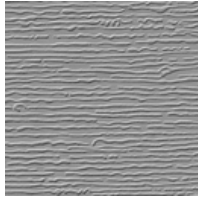
				
151	172 (5)	411 (4)	176 (4)	141 (4)
				
	447 (3)	369 (2)	007 (2)	499 (1)
				
	101 (1)			
				
367	345 (7)	056 (6)	079 (5)	236 (3)
				
	099 (3)	395 (2)	228 (2)	482 (1)
				
	311 (1)	095 (1)		

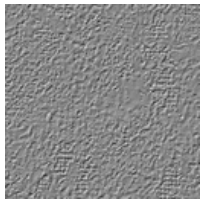
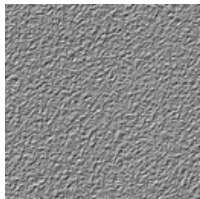
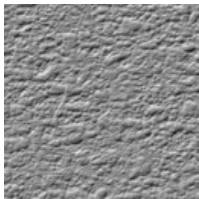
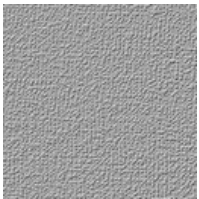
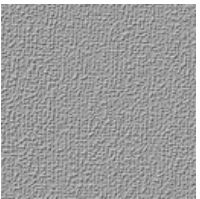
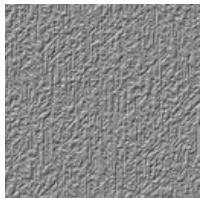
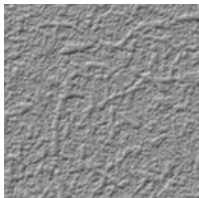
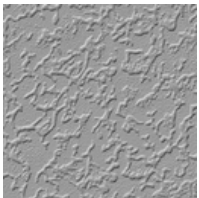
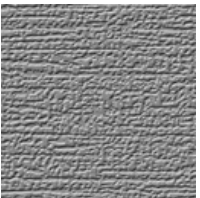
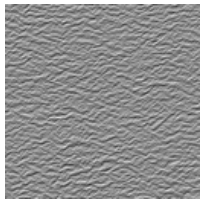
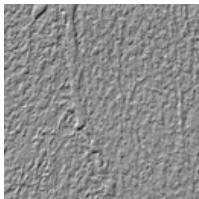
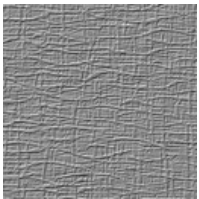
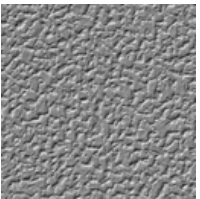
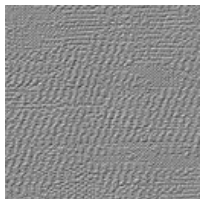
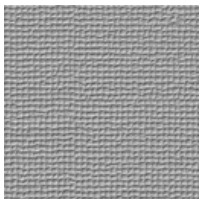
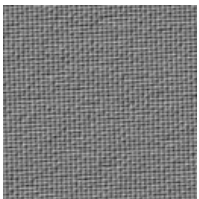
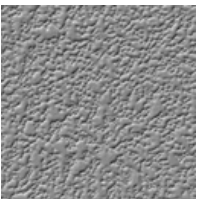
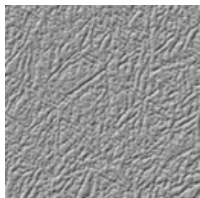
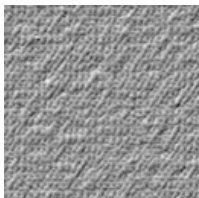
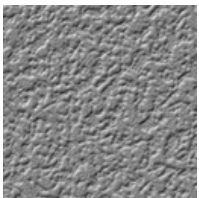
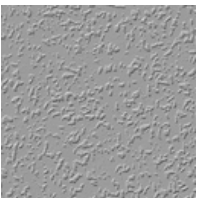
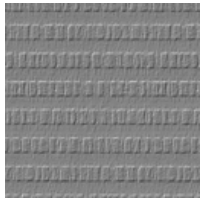
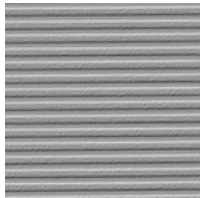
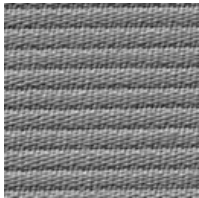
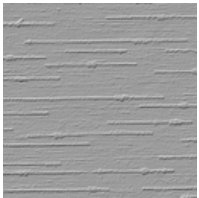
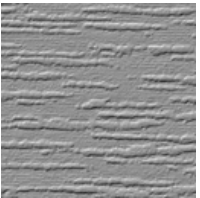
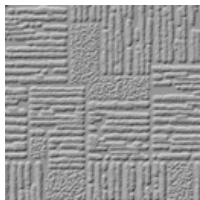
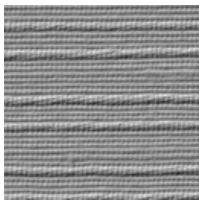
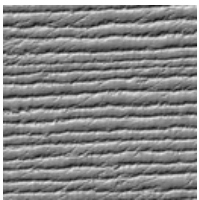
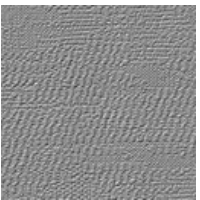
				
408	271 (7)	290 (4)	201 (4)	384 (3)
				
	324 (3)	204 (3)	197 (3)	390 (2)
				
	250 (2)	320 (1)	304 (1)	
				
231	369 (9)	411 (6)	499 (1)	447 (1)
				
	252 (1)	249 (1)	169 (1)	146 (1)
				
	109 (1)	102 (1)	101 (1)	088 (1)

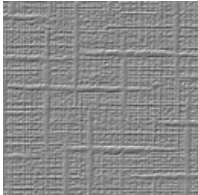
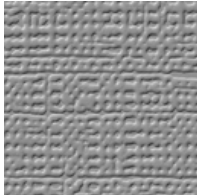
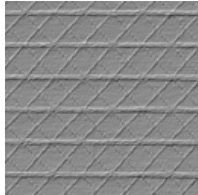
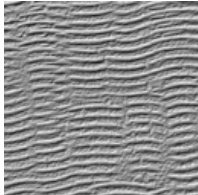
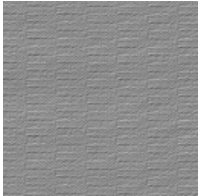
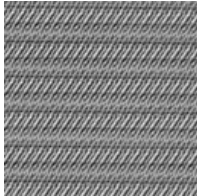
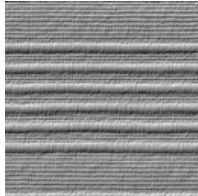
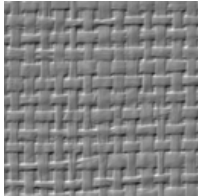
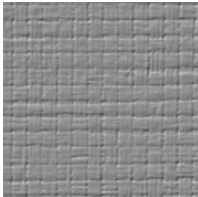
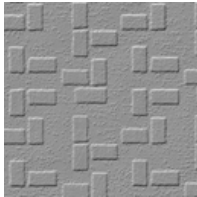
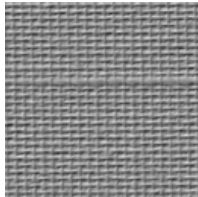
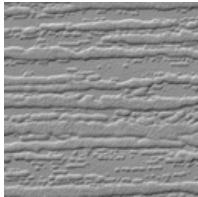
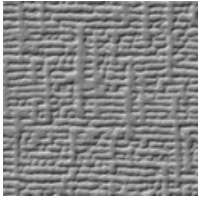
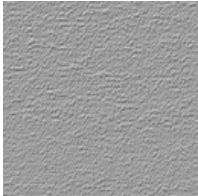
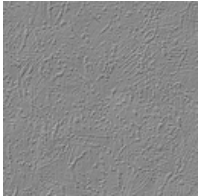
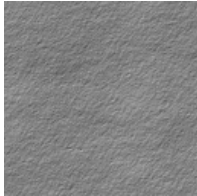
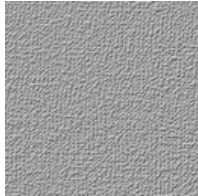
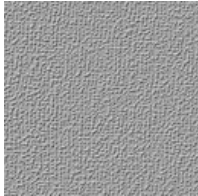
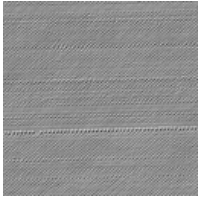
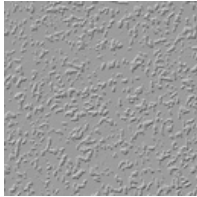
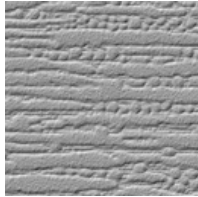
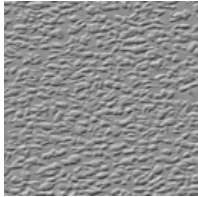
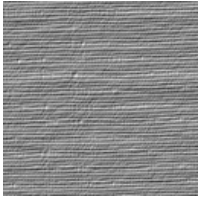
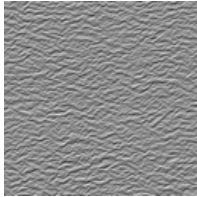
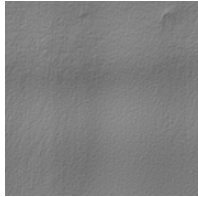
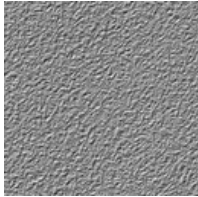
				
	079 (1)			
				
308	362 (5)	334 (4)	221 (4)	187 (4)
				
	430 (2)	345 (2)	437 (1)	434 (1)
				
	378 (1)	243 (1)	236 (1)	099 (1)
				
	079 (1)	056 (1)		
				
110	141 (6)	333 (5)	499 (3)	369 (3)
				
	172 (3)	447 (2)	411 (2)	176 (2)

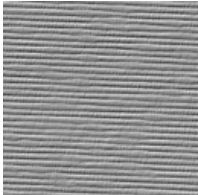
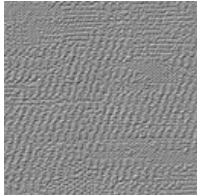
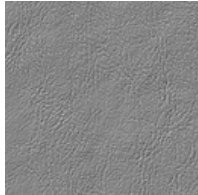
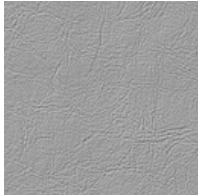
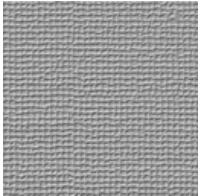
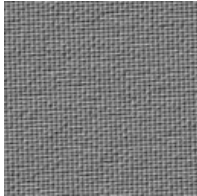
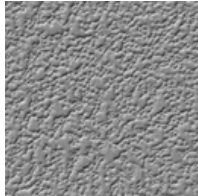
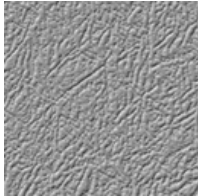
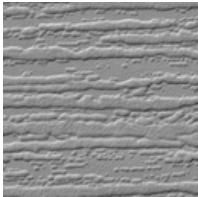
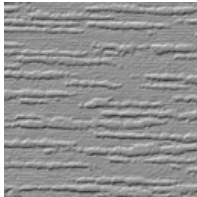
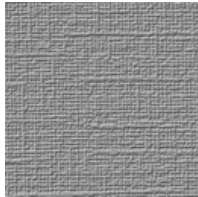
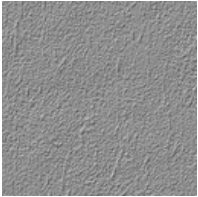
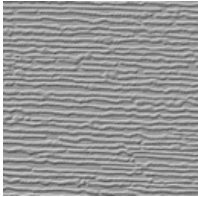
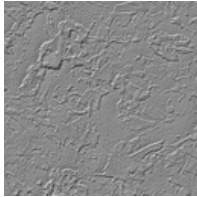
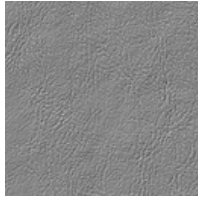
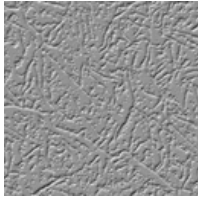
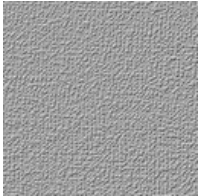
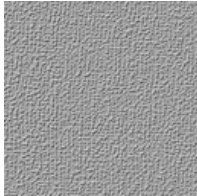
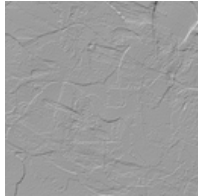
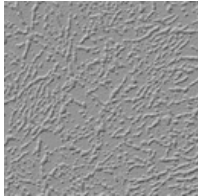
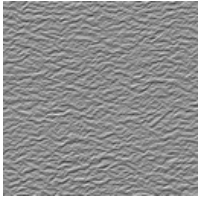
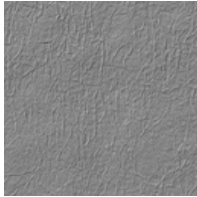
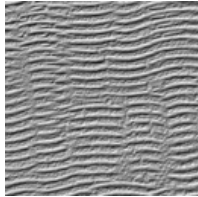
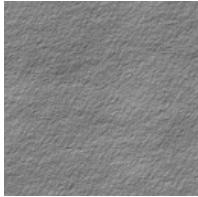
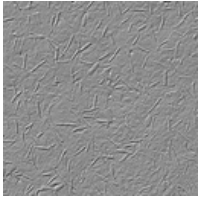
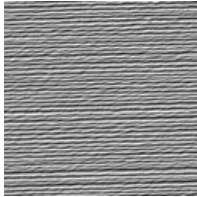
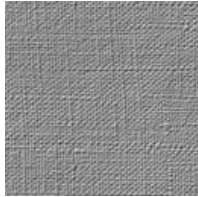
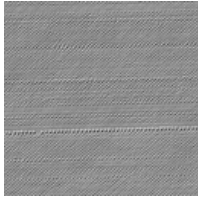
				
	101 (2)	436 (1)	311 (1)	191 (1)
				
	088 (1)	034 (1)	014 (1)	
				
454	290 (6)	455 (4)	204 (3)	446 (2)
				
	324 (2)	136 (2)	430 (1)	378 (1)
				
	304 (1)	250 (1)	222 (1)	201 (1)
				
	200 (1)	197 (1)	184 (1)	059 (1)
				
126	447 (6)	213 (3)	007 (3)	200 (2)

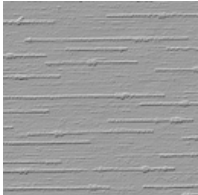
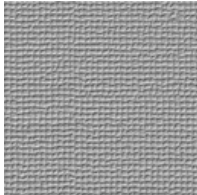
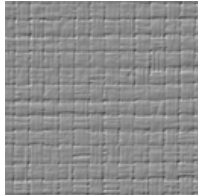
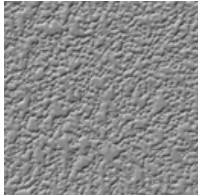
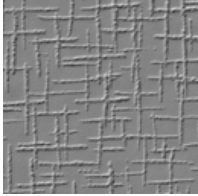
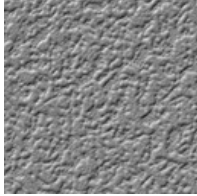
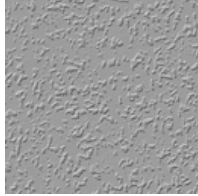
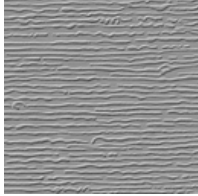
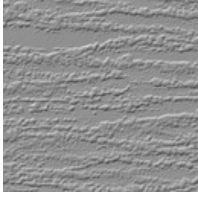
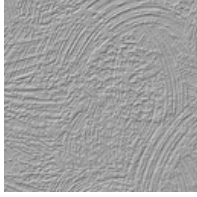
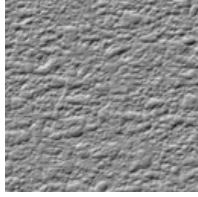
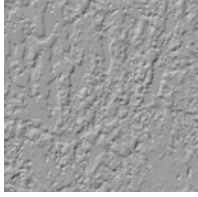
				
	141 (2)	451 (1)	378 (1)	365 (1)
				
	282 (1)	273 (1)	236 (1)	191 (1)
				
	184 (1)	146 (1)	136 (1)	117 (1)
				
	091 (1)			
				
114	007 (4)	447 (3)	141 (3)	451 (2)
				
	411 (2)	345 (2)	146 (2)	499 (1)
				
	482 (1)	462 (1)	311 (1)	184 (1)

				
	176 (1)	172 (1)	101 (1)	099 (1)
				
	056 (1)	034 (1)		
				
118	117 (6)	466 (5)	378 (2)	282 (2)
				
	197 (2)	184 (2)	045 (2)	499 (1)
				
	430 (1)	395 (1)	345 (1)	236 (1)
				
	200 (1)	187 (1)	146 (1)	127 (1)
				
	091 (1)	058 (1)	056 (1)	

				
274	222 (4)	017 (3)	455 (2)	446 (2)
				
	273 (2)	023 (2)	479 (1)	451 (1)
				
	326 (1)	320 (1)	282 (1)	225 (1)
				
	213 (1)	184 (1)	146 (1)	136 (1)
				
	127 (1)	091 (1)	072 (1)	059 (1)
				
406	362 (4)	105 (4)	187 (3)	079 (3)
				
	365 (2)	334 (2)	228 (2)	213 (2)

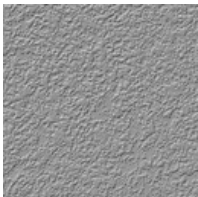
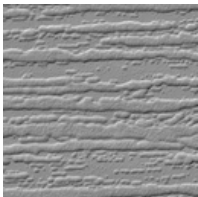
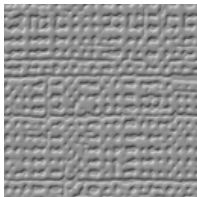
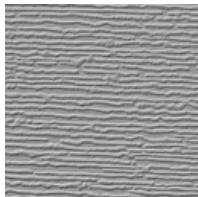
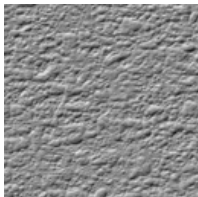
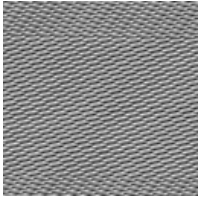
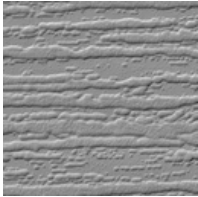
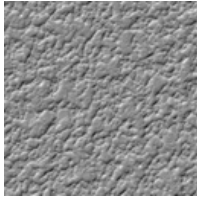
				
	466 (1)	447 (1)	437 (1)	311 (1)
				
	252 (1)	249 (1)	243 (1)	176 (1)
				
	169 (1)	159 (1)	141 (1)	099 (1)
				
	007 (1)			
				
153	290 (5)	304 (4)	455 (3)	446 (2)
				
	197 (2)	059 (2)	482 (1)	468 (1)
				
	430 (1)	326 (1)	250 (1)	222 (1)

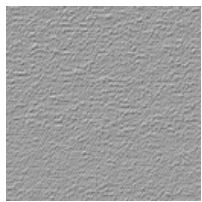
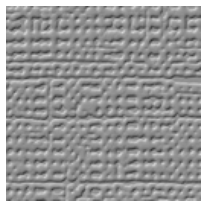
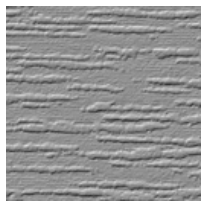
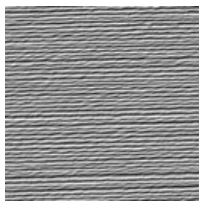
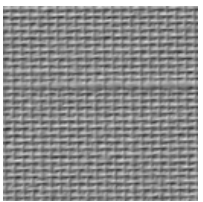
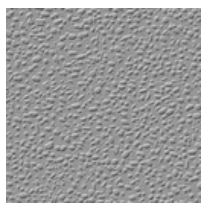
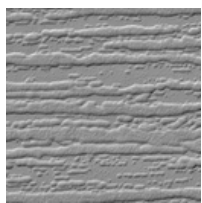
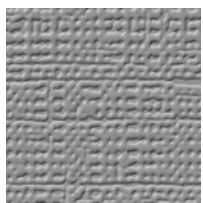
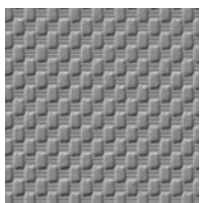
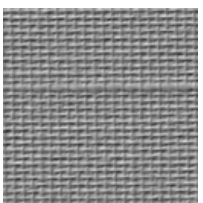

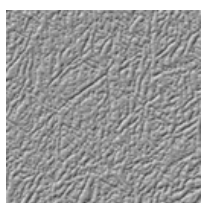
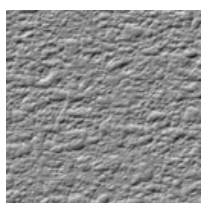
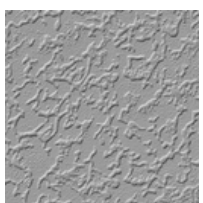
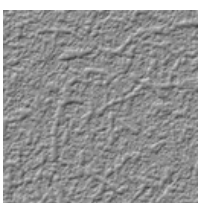
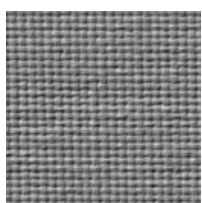
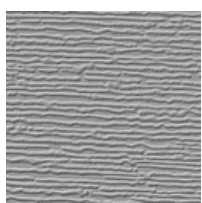
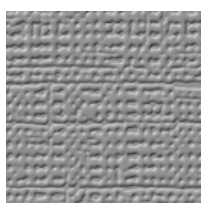
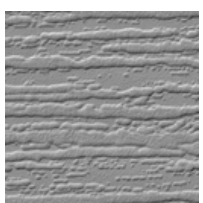
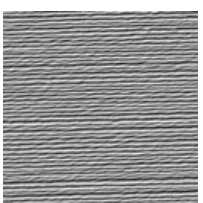
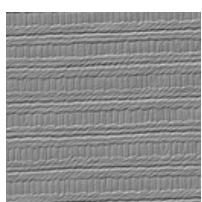
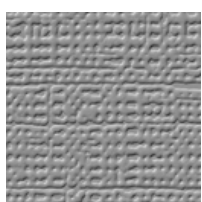
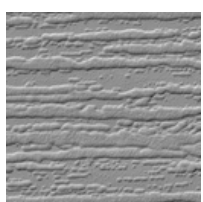
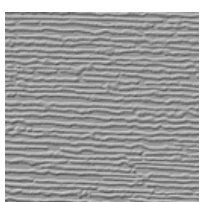
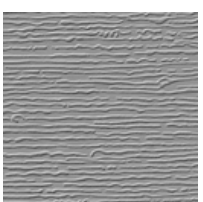
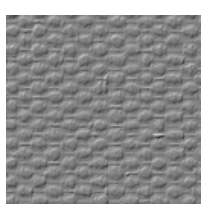
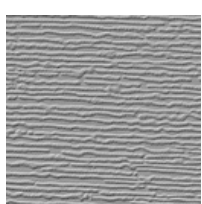
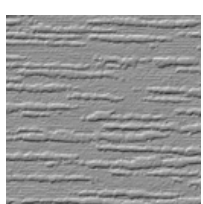
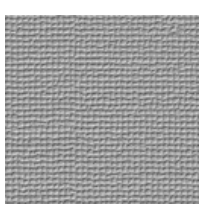
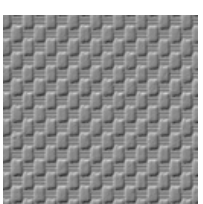
				
	221 (1)	213 (1)	204 (1)	201 (1)
				
	184 (1)	146 (1)	136 (1)	127 (1)
				
	099 (1)	079 (1)	045 (1)	
				
317	345 (2)	271 (2)	204 (2)	116 (2)
				
	455 (1)	446 (1)	390 (1)	346 (1)
				
	326 (1)	324 (1)	311 (1)	304 (1)
				
	284 (1)	236 (1)	200 (1)	197 (1)

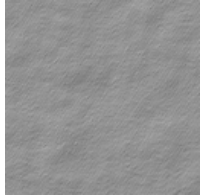
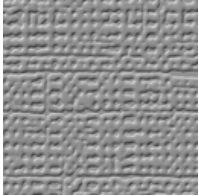
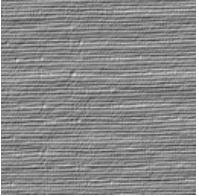
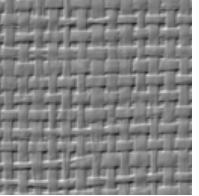
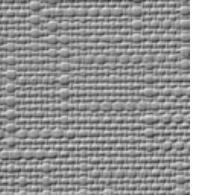
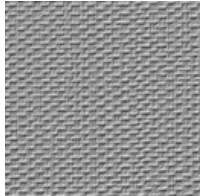
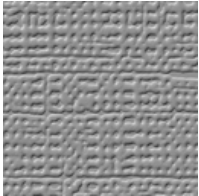
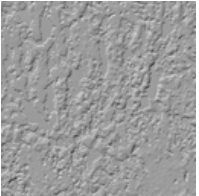
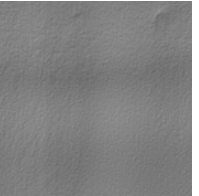
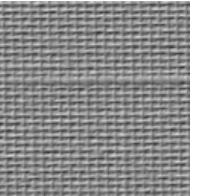
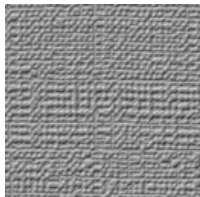
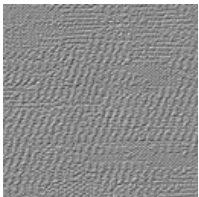
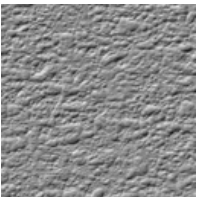
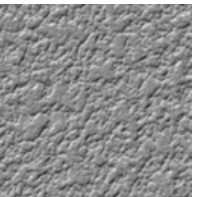
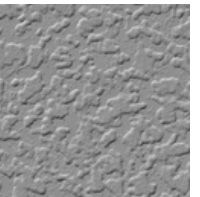
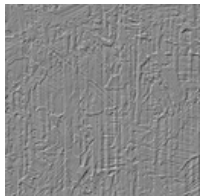
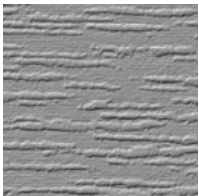
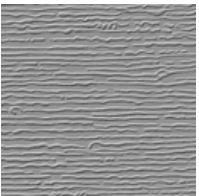
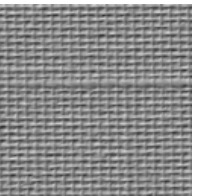
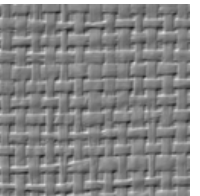
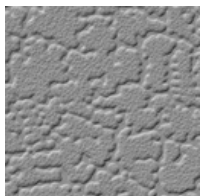
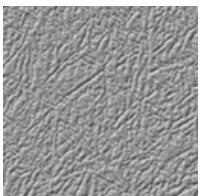
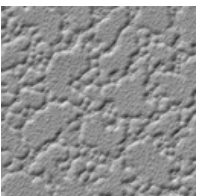
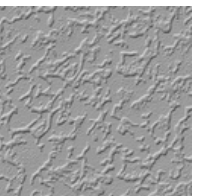
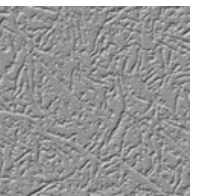
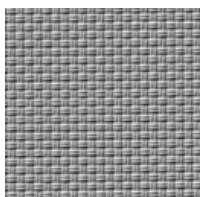
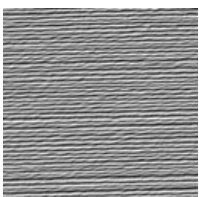
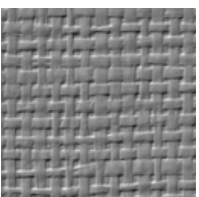
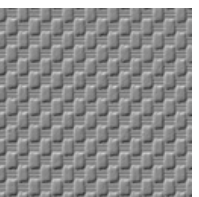
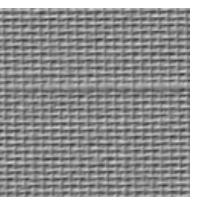
			
187 (1)	184 (1)	169 (1)	136 (1)
			
109 (1)	072 (1)	059 (1)	056 (1)
			
054 (1)	049 (1)	017 (1)	010 (1)

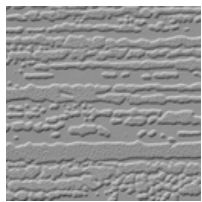
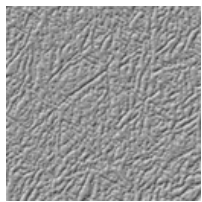
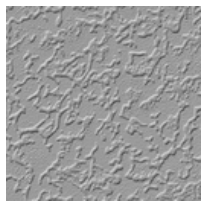
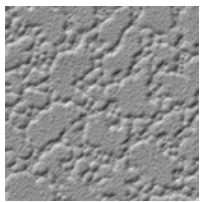
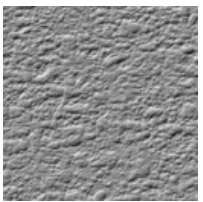
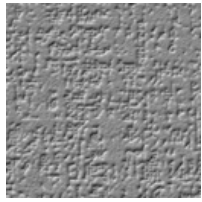
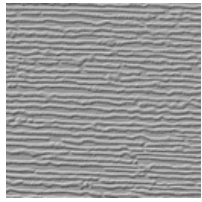
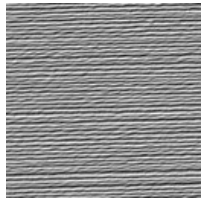
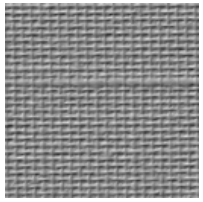
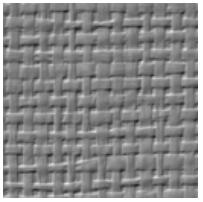
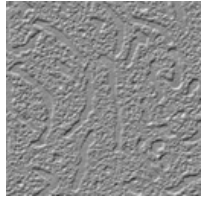
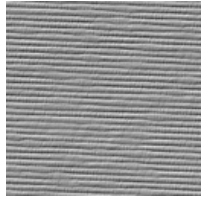
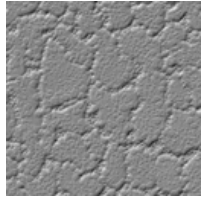
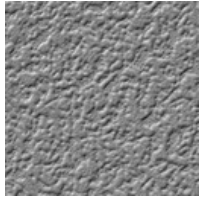
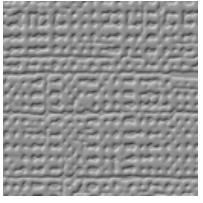
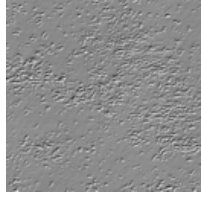
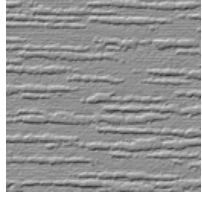
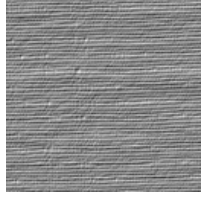
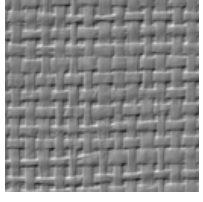
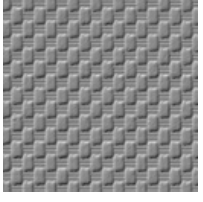
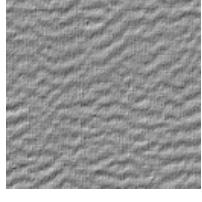
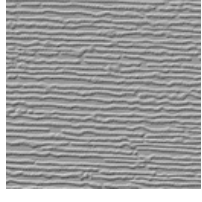
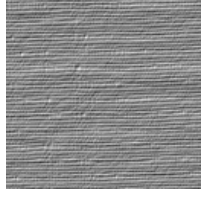
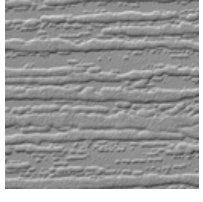
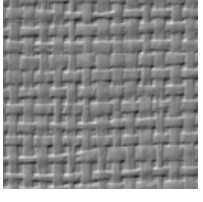
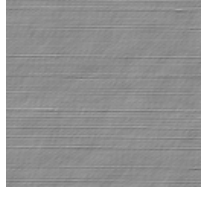
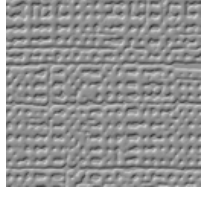
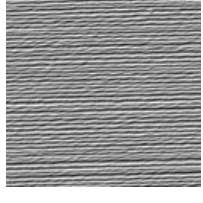
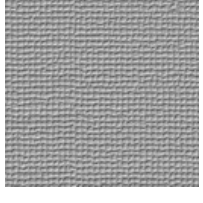
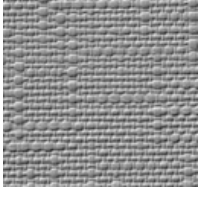
B.5.2 Rejected Response Example

The following observer's session was rejected as their result did not meet the minimum criteria for payment.

Stimuli	Response				
					
048	099	447	345	017	
					
289	099	063			

				
153	447	079	236	141
				
057	099	447	369	141
				
247	127	017	479	023
				
143	345	447	099	236
				
230	447	099	345	056
				
398	345	079	184	369

				
315	447	430	176	411
				
461	447	010	250	141
				
126	213	017	063	166
				
272	079	056	141	176
				
376	127	129	479	116
				
103	236	176	369	141

				
356	127	479	129	017
				
472	345	236	141	176
				
018	221	066	072	447
				
245	079	430	176	369
				
312	345	430	099	176
				
321	447	236	184	411

B.6 SOM_G & SOM_{A2} Browsing Environments Comparison

Observer	Mean Task Time	
	SOM _G	SOM _{A2}
1	108.500	86.583
2	51.583	26.417
3	57.000	33.667
4	122.750	57.833
5	55.250	156.917
6	48.000	83.500
7	37.583	64.250
8	34.417	129.500
9	41.583	88.417
10	70.833	139.833
11	51.667	67.500
12	99.833	44.417
Means	64.917	81.569

Bibliography

- [1] N. Abbadeni. Perceptual image retrieval. In *VISUAL*, pages 259–268, 2005.
- [2] S. Aksoy and R. M. Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recogn. Lett.*, 22:563–582, April 2001.
- [3] M. Amadasun and R. King. Textural features corresponding to textural properties. *T-SMC*, 19:1264–1274, 1989.
- [4] Amazon Mechanical Turk. <https://www.mturk.com/>.
- [5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. *J. Mach. Learn. Res.*, 3:1107–1135, 2003.
- [6] D. Benyon. The new hci? navigation of information space. *Knowl.-Based Syst.*, 14(8):425–430, 2001.
- [7] P. Brodatz. *Textures; a photographic album for artists and designers*. Dover Publications New York, 1966.
- [8] M. J. Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Proc. on Visual Image and Signal Processing*, 142(4):199–206, 1994.
- [9] C. Chen, G. Gagaudakis, and P. Rosin. Similarity-based image browsing. In *Proceedings of the 16th IFIP World Computer Congress (International Conference on Intelligent Information Processing)*, 2000.
- [10] J. Chen, C. A. Bouman, and J. C. Dalton. Active browsing using similarity pyramids. In *Proceedings of 1998 Asilomar Conference on Signals, Systems, And Computers*, pages 144–154, 1999.
- [11] J. Chen, C. A. Bouman, and J. C. Dalton. Hierarchical browsing and search of large image databases. *IEEE Transactions on Image Processng*, 9:442–455, 2000.

- [12] P. Chiu, A. Girgensohn, S. Lertsithichai, W. Polak, and F. Shipman. Mediametro: browsing multimedia document collections with a 3d city metaphor. In *Proceedings of the 13th annual ACM international conference on Multimedia, MULTIMEDIA '05*, pages 213–214, New York, NY, USA, 2005. ACM.
- [13] A. D. F. Clarke, F. Halley, P. R. Green, and M. J. Chantler. Similar symmetries: The role of wallpaper groups in perceptual texture similarity. *Symmetry: Special issue on Symmetry Processing in Perception and Art*, 3(2):246–264, 2011.
- [14] A. D. F. Clarke, F. Halley, A. J. Newell, L. D. Griffin, and M. J. Chantler. Perceptual similarity: A texture challenge. In *Proceedings of the British Machine Vision Conference*, page 52, 2011.
- [15] D. A. Clausi and M. E. Jernigan. Designing gabor filters for optimal texture separability. *Pattern Recognition*, 33(11):1835–1849, 2000.
- [16] T. T. A. Combs and B. B. Bederson. Does zooming improve image browsing? In *Proceedings of the Fourth ACM International Conference on Digital Libraries*, 1999.
- [17] K. J. Dana, B. Van-Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and Texture of Real World Surfaces. *ACM Transactions on Graphics (TOG)*, 18(1):1–34, Jan 1999.
- [18] M. De Marsicoi, L. Cinque, and S. Levialdi. Indexing pictorial documents by their content: a survey of current techniques. *Image and Vision Computing*, 15(2):119 – 141, 1997.
- [19] O. Dekel and O. Shamir. Vox populi: Collecting high-quality labels from a crowd. In *In Proceedings of the 22nd Annual Conference on Learning Theory*, 2009.
- [20] G. Demontis, M. Mosconi, and M. Porta. Experimental interfaces for visual browsing of large collections of images. In *Proceedings of the Tenth International Conference on Human-Computer Interaction*, volume 2 of *Human factors and ergonomics*, pages 1081–1085, 2003.
- [21] D. Dewangan, V. J. Samar, R. M. Rao, and P. Paul. Factors influencing psychophysically valid taxonomies of image texture. In *ICIP (3)*, pages 1196–1199, 2005.
- [22] R. M. Downs and D. Stea. Cognitive representations. In *Image and Environment*, pages 79–86, 1973.

- [23] J. M. H. du Buf and P. Heitkämper. Texture features based on gabor phase. *Signal Process.*, 23:227–244, June 1991.
- [24] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience Publication, 2000.
- [25] K. Emrith. *Perceptual Dimensions for Surface Texture Retrieval*. PhD thesis, Heriot Watt University, Edinburgh, 2008.
- [26] X. Fan, X. Xie, W. Ma, H. Zhang, and H. Zhou. Visual attention based image browsing on mobile devices. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, volume 1, pages I – 53–6 vol.1, July 2003.
- [27] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning object categories from google’s image search. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1816–1823, Washington, DC, USA, 2005. IEEE Computer Society.
- [28] A. Field. *Discovering Statistics Using SPSS (Introducing Statistical Methods series)*. Sage Publications Ltd, second edition edition, 2005.
- [29] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The qbic system. *Computer*, 28(9):23–32, 1995.
- [30] Flickr Photo Sharing Service. <http://www.flickr.com>.
- [31] C. Fraley and A. E. Raftery. How many clusters? which clustering method? answers via model-based cluster analysis. *Computer Journal*, 41(8):578–588, 1998.
- [32] Google Image Labeler. <http://images.google.com/imagelabeler>.
- [33] A. D. Gordon. A review of hierarchical classification. *Journal of the Royal Statistical Society, Series A*, 150:119–137, 1987.
- [34] A. S. Gordon. Browsing image collections with representations of common-sense activities. *Journal of the American Society for Information Science and Technology*, 52(11):925–929, 2001.
- [35] J. Gordon, B. Van Durme, and L. K. Schubert. Evaluation of commonsense knowledge with mechanical turk. In *In NAACL Workshop on Creating Speech and Language Data With Amazons Mechanical Turk*, 2010.
- [36] R. M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, 1984.

- [37] V. N. Gudivada and V. V. Raghavan. Content-based image retrieval systems. *Computer*, 28(9):18–22, 1995.
- [38] R. Gurnsey and D. J. Fleet. Texture space. *Vision Research*, 41:745–757, 2001.
- [39] D. Heesch. A survey of browsing models for content based image retrieval. *Multimedia Tools and Applications*, 40:261–284, 2008.
- [40] Y.-X. Ho, M. S. Landy, and L. T. Maloney. Conjoint measurement of gloss and surface texture. *Psychological Science*, 19(2):196–204, 2008.
- [41] L. E. Holmquist. Focus+context visualization with flip zooming and the zoom browser. In *CHI '97 extended abstracts on Human factors in computing systems: looking to the future*, CHI EA '97, pages 263–264, New York, NY, USA, 1997. ACM.
- [42] L. E. Holmquist and S. Bjork. A hierarchical focus & context method for image browsing. In *ACM SIGGRAPH 98 Conference abstracts and applications*, SIGGRAPH '98, pages 282–, New York, NY, USA, 1998. ACM.
- [43] T. Honkela, S. Kaski, K. Lagus, and T. Kohonen. Websom - self-organizing maps of document collections. In *Neurocomputing*, pages 101–117, 1998.
- [44] A. Jain, M. Murty, and P. Flynn. Data clustering: A review. *ACM Computing Survey*, 31(3):264–323, 1999.
- [45] Z. Jiang, X. Luo, Y. Mao, B. Zang, H. Lin, and H. Bao. Interactive browsing of large images on multi-projector display wall system. In *Human-Computer Interaction. Interaction Platforms and Techniques*, volume 4551 of *Lecture Notes in Computer Science*, pages 827–836. 2007.
- [46] S. Jul and G. W. Furnas. Navigation in electronic worlds. *SIGCHI Bulletin*, 29:44–49, 1997.
- [47] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290:91–97, 1981.
- [48] R. Kadobayashi and K. Tanaka. 3d viewpoint-based photo search and information browsing. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 621–622, New York, NY, USA, 2005. ACM.
- [49] A. Kadyrov and M. Petrou. The trace transform and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:811–828, 2001.

- [50] S. Kawabata, R. Postle, and M. Niwa. Objective specification of fabric quality, mechanical properties and performance. In *Japan-Australian Joint Symposium*. The Textile Machinery Society of Japan, 1982.
- [51] I. Keller, T. Meiers, T. Ellerbrock, and T. Sikora. Image browsing with pca-assisted user-interaction. In *Content-Based Access of Image and Video Libraries, 2001. (CBAIVL 2001). IEEE Workshop on*, pages 102–108, 2001.
- [52] A. Kittur, H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proc. CHI 2008, ACM Press*, pages 453–456, 2008.
- [53] T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, December 2000.
- [54] A. K. König. A survey of methods for multivariate data projection, visualisation and interactive analysis. In *Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems*, pages 55–59, 1998.
- [55] P. Korpipää and J. Remes. Website navigation and profiling based on self-organising maps. In *Proceedings of the IJCAI'03 Workshop on Artificial Intelligence, Information Access, and Mobile Computing*, pages 59–64, August 2003.
- [56] S. Krishnamachari and M. Abdel-Mottaleb. Image browsing using hierarchical clustering. In *Computers and Communications, 1999. Proceedings. IEEE International Symposium on*, pages 301–307, 1999.
- [57] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129, June 1964.
- [58] K. Lagus, T. Honkela, S. Kaski, and T. Kohonen. Self-organizing maps of document collections: A new approach to interactive exploration. In *Second International Conference on Knowledge Discovery and Data Mining*, pages 238–243. AAAI Press, 1996.
- [59] K. Lam and R. Spence. Image browsing: a space-time trade-off. In *Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction, INTERACT '97*, pages 611–612, London, UK, 1997. Chapman & Hall, Ltd.
- [60] S. Lim, L. Chen, G. Lu, and R. Smith. Browsing texture image databases. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 328 – 333, January 2005.
- [61] 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*, 20:64–76, 2005.

- [62] H. Long, W. K. Leow, and F. K. Chua. Perceptual texture space for content-based image retrieval. In *In Proc. Int. Conf. on Multimedia Modeling (MMM 2000*, pages 167–180, 2000.
- [63] H. Long, C. W. Tan, and W. K. Leow. Invariant and perceptually consistent texture mapping for content-based image retrieval. In *In Proc. ICIP*, pages 117–120, 2001.
- [64] D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Norwell, MA, USA, 1985.
- [65] J. MacCuish, A. McPherson, J. Barros, and P. Kelly. Interactive layout mechanisms for image database retrieval. In *Proceedings of SPIE Conference on Visual Data Exploration and Analysis III*, pages 104–115, 1996.
- [66] J. Martinez and E. Loissant. Browsing image databases with galois’ lattices. In *Proceedings of the 2002 ACM symposium on Applied computing, SAC ’02*, pages 791–795, New York, NY, USA, 2002. ACM.
- [67] W. Mason and D. J. Watts. Financial incentives and the ”performance of crowds”. In *HCOMP ’09: Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 77–85, New York, NY, USA, 2009. ACM.
- [68] A. Mead. Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(1):pp. 27–39, 1992.
- [69] MeasTex Image Texture Database.
<http://www.texturesynthesis.com/meastex/meastex.html>.
- [70] M. Mirmehdi, X. Xie, and J. Suri, editors. *Handbook of texture analysis*. Imperial College Press, 2008.
- [71] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson. Image statistics and the perception of surface qualities. *Nature*, 447:206–209, April 2007.
- [72] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.
- [73] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002.

- [74] M. Oren and S. K. Nayar. Generalization of lambert’s reflectance model. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, SIGGRAPH ’94, pages 239–246, New York, NY, USA, 1994. ACM.
- [75] OuTex Texture Database. <http://www.outex.oulu.fi/>.
- [76] S. Padilla, F. Halley, and M. J. Chantler. Improving product browsing whilst engaging users. In *Proceeding of Digital Engagement Conference*, 2011.
- [77] W.-M. Pang. An intuitive texture picker. In *Proceeding of the 14th international conference on Intelligent user interfaces*, IUI ’10, pages 365–368, New York, NY, USA, 2010. ACM.
- [78] R. Passini. *Wayfinding in Architecture*. Van Nostrand Reinhold, New York, 1992.
- [79] J. S. Payne, L. Hepplewhite, and T. J. Stonham. Perceptually based metrics for the evaluation of textural image retrieval methods. In *ICMCS, Vol. 2*, pages 793–797, 1999.
- [80] Janet S. Payne and T. J. Stonham. Mapping perceptual texture similarity for image retrieval. In *SCIA*, pages 960–969, 2005.
- [81] Z. Pecenovio, M. N. Do, M. Vetterli, and P. Pu. Integrated browsing and searching of large image collections. In R. Laurini, editor, *VISUAL*, volume 1929 of *Lecture Notes in Computer Science*, pages 279–289. Springer, 2000.
- [82] M. Petrou and A. Kadyrov. Affine invariant features from the trace transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:30–44, 2004.
- [83] M. Petrou, A. Talebpour, and A. Kadyrov. Reverse engineering the way humans rank textures. *Pattern Anal. Appl.*, 10:101–114, April 2007.
- [84] PhoTex Photometric Texture Database at Texture Lab. <http://www.macs.hw.ac.uk/texturelab/resources/databases/photex/>.
- [85] W. Plant and G. Schaefer. Navigation and browsing of image databases. In *Soft Computing and Pattern Recognition, 2009. SOCPAR ’09. International Conference of*, pages 750 –755, December 2009.
- [86] J. C. Platt, M. Czerwinski, and B. A. Field. Phototoc: automatic clustering for browsing personal photographs. In *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, volume 1, pages 6 – 10 Vol.1, December 2003.

- [87] M. Porta. Browsing large collections of images through unconventional visualization techniques. In *AVI '06: Proceedings of the working conference on Advanced visual interfaces*, pages 440–444, New York, NY, USA, 2006. ACM.
- [88] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vision*, 40:49–70, October 2000.
- [89] A. R. Rao and G. L. Lohse. Identifying high level features of texture perception. *CVGIP: Graph. Models Image Process.*, 55:218–233, May 1993.
- [90] A. R. Rao and G. L. Lohse. Towards a texture naming system: identifying relevant dimensions of texture. In *Proceedings of the 4th conference on Visualization '93*, VIS '93, pages 220–227, Washington, DC, USA, 1993. IEEE Computer Society.
- [91] K. Rodden. How do people organise their photographs? In *Proceedings of the BCS IRSG Colloquium*, 1999.
- [92] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Evaluating a visualisation of image similarity as a tool for image browsing. In *Information Visualization, 1999. (Info Vis '99) Proceedings. 1999 IEEE Symposium on*, pages 36–43, 143, 1999.
- [93] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 190–197, New York, NY, USA, 2001. ACM.
- [94] K. Rodden and K. R. Wood. How do people manage their digital photographs? In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 409–416, New York, NY, USA, 2003. ACM.
- [95] B. E. Rogowitz, T. Frese, J. R. Smith, C. A. Bouman, and E. Kalin. Perceptual image similarity experiments. In *SPIE Conference on Human Vision and Electronic Imaging*, pages 576–590, 1998.
- [96] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173, 2008.
- [97] S. Santini and R. Jain. Integrated browsing and querying for image databases. *IEEE Multimedia*, 7:26–39, 2000.

- [98] G. Schaefer. A next generation browsing environment for large image repositories. *Multimedia Tools Appl.*, 47:105–120, March 2010.
- [99] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or ”how do i organize my holiday snaps?”. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 414–431, London, UK, 2002. Springer-Verlag.
- [100] M. Schroeder and P. Noy. Multi-agent visualisation based on multivariate data. In *In Working Notes of the Fourth UK Workshop on Multi-Agent Systems UKMAS-01*, pages 85–91. ACM Press, 2001.
- [101] R. N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 27(3):219–246, 1962.
- [102] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3 - Volume 3*, ICIP '95, pages 3444–, Washington, DC, USA, 1995. IEEE Computer Society.
- [103] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, pages 835–846, New York, NY, USA, 2006. ACM Press.
- [104] P. H. A. Sneath and R. R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. Freeman, San Francisco, 1973.
- [105] R. Snow, B. O’Connor, D. Jurafsky, and A. Y. Ng. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *EMNLP '08: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 254–263, Morristown, NJ, USA, 2008. Association for Computational Linguistics.
- [106] R. S. Torres, C. G. Silva, C. B. Medeiros, and H. V. Rocha. Visual structures for image browsing. In *Proceedings of the twelfth international conference on Information and knowledge management, CIKM '03*, pages 49–55, New York, NY, USA, 2003. ACM.
- [107] G. Strong and M. Gong. Browsing a large collection of community photos based on similarity on gpu. In *Advances in Visual Computing*, volume 5359 of *Lecture Notes in Computer Science*, pages 390–399. 2008.
- [108] H. Tamura, S. Mori, and T. Yamawaki. Texture features corresponding to visual perception. *IEEE Transactions on System, Man and Cybernatic*, 6, 1978.

- [109] C. D. Taylor. *The Analysis of Woven Fabric Structure*. PhD thesis, Heriot Watt University, Edinburgh, 2003.
- [110] J. T. Todd., J. F. Norman, J. J. Koenderink, and A. M. L. Kappers. Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, 26:807–822, 1997.
- [111] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky. Video manga: Generating semantically meaningful video summaries. pages 383–392. ACM Press, 1999.
- [112] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas. Self-organizing map in matlab: the som toolbox. In *In Proceedings of the Matlab DSP Conference*, pages 35–40, 2000.
- [113] VisTex Texture Lab Database at MIT.
<http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>.
- [114] J. A. Walter and H. Ritter. On interactive visualization of high-dimensional data using the hyperbolic plane. In *KDD*, pages 123–132, 2002.
- [115] J. A. Walter, D. Weßling, K. Essig, and H. Ritter. Abstract interactive hyperbolic image browsing towards an integrated multimedia navigator. In *ACM Conference on Knowledge Discovery and Data Mining*, page 111, 2006.
- [116] A. A. Ward, S. J. McKenna, A. Buruma, P. Taylor, and J. Han. Merging technology and users: Applying image browsing to the fashion industry for design inspiration. In *Content-Based Multimedia Indexing, 2008. CBMI 2008. International Workshop on*, pages 288 –295, June 2008.
- [117] M. Wish and J. D. Carroll. Concepts and applications of multidimensional scaling. *ASTM STP*, 545:91–108, 1973.
- [118] K. Wittenburg, W. Ali-Ahmad, D. Laliberte, and T. Lanning. Rapid-fire image previews for information navigation. In *Proceedings of Advanced Visual Interfaces (May, L'Aquila, ITALY)*, ACM, pages 76–82. ACM Press, 1998.
- [119] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 1980.
- [120] D. Wu, A. Singh, D. Agrawal, A. El Abbadi, and T. R. Smith. Efficient retrieval for browsing large image databases. In *Proceedings of the fifth international conference on Information and knowledge management, CIKM '96*, pages 11–18, New York, NY, USA, 1996. ACM.