To the Graduate Council:

I am submitting herewith a dissertation written by Anthony C. Copeland entitled "Image Metrics for Human Search and Discrimination of Textured Targets and Backgrounds." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Electrical Engineering.

Prof. Paul B. Crilly, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of The Graduate School

IMAGE METRICS FOR HUMAN SEARCH AND DISCRIMINATION OF TEXTURED TARGETS AND BACKGROUNDS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Anthony C. Copeland

December 1996

ACKNOWLEDGEMENTS

This work would not have been possible without the advice and assistance of many people. I would particularly like to thank:

- my major professor, Prof. Paul B. Crilly, and the other committee members at the University of Tennessee,
- Prof. Mohan M. Trivedi of the University of California, San Diego, who served as research advisor and mentor during the preparation of this work,
- Dr. Gopalan Ravichandran of SRI International, who provided much technical advice and assistance, particularly relating to texture synthesis,
- Mr. Nils Lassiter of U.C. San Diego, who assisted greatly in the acquisition and assembly of hardware necessary for the eye tracking testbed,
- all volunteers who donated their time to serve as test subjects in the perceptual experiments,
- Mr. Jim McManamey of the U.S. Army CECOM Research, Development, and Engineering Center at Fort Belvoir, VA, who provided technical advice as well as some editing comments for portions of Chapter 5,
- the U.S. Army and U.C. San Diego for providing financial support,
- and Dr. Arun Dalmia and Datacube, Inc., in Danvers, MA, for providing the MaxTD image processing system containing a MaxVideo200 pipeline processor used in the eye tracking testbed.

ABSTRACT

This dissertation is a study of the use of quantitative image metrics to measure the signature strength of a camouflaged target in a sensed image. These measurements of signature strength are compared to the results of experiments in which human observers perform the tasks of search and discrimination with digital imagery. The real-world analogy of these experiments is that of a military observer looking upon the sensed image of a static scene to find camouflaged enemy targets that are reported to be in the area. Several pure discrimination experiments are conducted with image stimuli possessing target patterns in a fixed location. The psychological scale values from human judgments of the stimuli are compared to both existing and newly developed target signature metrics. Using this technique, it is demonstrated that some newly developed texture-based metrics most effectively quantify the perceptual distinctness of the target patterns. The task of search is integrated with the task of discrimination in an experiment in which observers are asked to view image stimuli containing targets randomly distributed throughout real-world outdoor scenes. The observer's fixation point in the image is measured by tracking his eye movements during the experiment. From the data, various statistics are computed for each target to describe how easily the observers located it, including the likelihood the target was fixated or identified and the time required to do so. The target signature metrics are compared to these statistics, and the texture-based metrics are again found to be the most highly correlated with the data.

TABLE OF CONTENTS

CHAPTER PAGE

1. Introduction and Motivation **1**

1.1 Camouflage on the Battlefield 1

1.2 Sensed Images and Image Texture 4

1.3 Tasks Constituting Human Target Acquisition 5

1.4 Research Methodology 8

1.5 Organization of the Dissertation 10

2. The Past: A Review of the Literature **11**

2.1 Methods of Image Texture Analysis and Modeling 11

2.1.1 General Methods 11

2.1.2 Gray Level Cooccurrence (GLC) Models 15

2.2 Studies of Human Visual Search and Discrimination 16

2.2.1 Discrimination 17

2.2.2 Search 19

3. Experiments in the Perceptual Qualities of a Texture Contained Within its Cooccurrence Probabilities **22**

3.1 Texture Synthesis using GLC Models 22

3.2 Visualizing the GLC Model of a Texture 26

3.2.1 Images for GLC Visualization 26

3.2.2 Surface Plots for GLC Visualization 28

CHAPTER PAGE

3.3 Importance of GLC Ridge Sharpness 36

3.4 Importance of GLC Displacement Vector 38

3.5 Gaussian Shaped GLC Models 40

3.6 Nature of GLC Error in Synthesized Texture 41

3.6.1 Visualizing GLC Synthesis Error 41

3.6.2 Analysis of GLC Synthesis Error 47

3.7 Conclusions 50

4. Computer Testbeds and Experimental Procedures for Studying Human Search and Discrimination 53

4.1 XPET – The X-Based Perceptual Experiment Testbed 53

4.1.1 Overview of XPET 53

4.1.2 Conducting an Experiment with XPET 55

4.1.3 Comparison Experiments to Investigate Human Texture Discrimination 60

4.2 ITEMS – The Integrated Testbed for Eye Movement Studies 66

4.2.1 Overview and Utility of ITEMS 66

4.2.2 Conducting a Visual Search Experiment Using ITEMS 68

4.2.3 ITEMS Hardware Configuration 70

4.2.4 ITEMS Software Configuration 72

5. First-order Image Statistics and Human Perception 97

5.1 Purpose of the Experiment 97

5.2 Creation of the Image Stimuli 97

CHAPTER PAGE

5.3 Data Collection 99

5.4 LCJ Analysis of the Data 101

5.5 Evaluation of Selected Target Signature Metrics 103

5.6 Doyle Metric Derivatives 105

5.7 A Discrimination Threshold Metric 108

5.8 Conclusions 109

6. Second-order Image Statistics and Human Perception 111

6.1 Creation of the Image Stimuli 111

6.2 Design and Conduct of the Experiments 116

6.3 Results of the Experiments 117

6.4 Conclusions 121

7. Hybrid Textures and Human Perception 123

7.1 Paired Comparisons of Hybrid Textures 123

7.1.1 Creation of the Image Stimuli 123

7.1.2 Design and Conduct of the Experiment 127

7.1.3 Results of the Experiment 129

7.1.4 Conclusions 134

7.2 Triadic Comparisons of Hybrid Textures 135

7.2.1 Creation of the Image Stimuli 136

7.2.2 Design and Conduct of the Experiment 143

7.2.3 Results and Analysis of the Data 144

7.2.4 Conclusions 163

CHAPTER PAGE

8. Natural Texture Patterns and Human Perception 164

8.1 Texture Selection and Stimulus Creation 166

8.2 Design and Conduct of the Experiment 169

8.3 Results and Analysis of the Data 171

8.3.1 Estimation of Psychological Scale Values for the Stimuli . . . 171

8.3.2 Measures of Image Boundary Strength 172

8.3.3 Evaluation of Individual Target Signature Metrics 176

8.3.4 Multivariable Linear Regression and Multiple Correlation . . . 184

8.4 Conclusions 188

9. Integrating Visual Search with Discrimination 190

9.1 Creation of the Image Scene Stimuli 190

9.2 Design and Conduct of the Experiment 192

9.3 Results and Analysis of the Data 194

9.3.1 Target Fixation and Identification Statistics 194

9.3.2 Evaluation of Target Signature Metrics 197

9.4 Conclusions 201

10. Concluding Remarks 209

10.1 Review of Accomplishments 209

10.2 Summary of Findings 211

10.3 Relevance to Other Research 212

10.4 Future Research Directions 214

BIBLIOGRAPHY 216

CHAPTER PAGE

APPENDICES 224

A. The 42 Triadic Comparisons Experiment Stimuli 225

B. The Raw Decision Data from the Triadic Comparisons Experiment . 229

C. The 15 Natural Texture Experiment Stimuli 233

D. Target Locations for the Search Experiment Stimuli 235

E. Sample Fixation Data for the Search Experiment 237

F. Computed Statistics for the Search Experiment Targets 268

G. Computed Target Signature Metrics for the Search Experiment Targets270

VITA 271

LIST OF TABLES

TABLE PAGE

4.1 An example of the data resulting from a short I_SPY experiment session. 75

4.2 An example of the data resulting from an I_TRACK session. 80

4.3 The measurement data used by I_WHERE. 88

4.4 An example of the fixation point data generated by I_WHERE. . . . 91

5.1 The raw data for the 12 stimuli used in the LCJ analysis. The element in row j and column k is the number of times (out of 18 observations) that stimulus k was judged greater than stimulus j. 102

5.2 Computed values for selected metrics and their correlation coefficients with respect to the LCJ scale values. 103

5.3 Computed values for some more recently developed metrics and their correlation coefficients with respect to the LCJ scale values. 105

6.1 The LCJ scale values for the 10 stimuli in each of the three texture synthesis experiments. 118

6.2 The correlation coefficients for the six GLC computational measures considered, computed relative to the LCJ scale values for the texture synthesis experiments. 120

TABLE PAGE

7.1 The raw data from the texture hybridization experiment. The element in row j and column k is the number of times that stimulus k was judged greater than stimulus j. The last line contains the sums over each column. 129

7.2 The traditional and adjusted scale values for the 11 stimuli in the texture hybridization experiment. 130

7.3 The correlation coefficients for the six GLC computational measures considered, computed relative to the traditional and the adjusted LCJ scale values for the texture hybridizations experiment. 134

7.4 The five texture models used in creating the training session. 142

7.5 The scheme used to create the four stimuli for the training session. . . 143

7.6 The six GLC metrics calculated for every possible pair of the 7 prototype texture models. 155

7.7 The correlation coefficients between the 17 finite psychological comparative distances and the corresponding metric calculations. 156

7.8 The computed proportions of correct predictions for each of the six GLC-based metrics. 158

8.1 Descriptions of each of the six real texture samples. 166

8.2 The scheme for creating the 15 stimuli from the six real texture samples. 168

8.3 The raw data for the 15 stimuli used in the LCJ analysis. The element in row j and column k is the number of times (out of 20 observations) that stimulus k was judged greater than stimulus j. 172

8.4 The scale values for the 15 stimuli. 173

8.5 The computed values of average boundary strength (ABS) and relative average boundary strength (RABS) for the 15 real texture experiment stimuli. 176

8.6 The sample correlation coefficient (r) computed between the vector of stimulus scale values and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric. 179

8.7 The correlation coefficients computed between pairs of selected metrics. 183

8.8 The multiple correlation coefficients for selected pairs of metrics. . . . 186

8.9 The multiple correlation coefficients and corresponding regression parameters for selected linear combinations of metrics. 187

9.1 The number of targets and the raw images used to contribute the target and background pixels for the ten stimuli in the search experiment. 191

9.2 The number of targets identified (out of 52) and the average distances in pixels of the fixation point estimates at the instant of identification for the 12 observers in the search experiment. 196

9.3 The correlation coefficients between pairs of computed statistics from the search experiment. 197

9.4 The sample correlation coefficient (r) computed between the vector of computed identification statistics and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric. . . 198

9.5 The sample correlation coefficient (r) computed between the vector of computed fixation statistics and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric. . . . 199

9.6 The correlation coefficients computed between pairs of selected metrics. 200

9.7 The multiple correlation coefficients computed between each possible pair of metrics and P_{ID} computed from the search experiment data. . 206

9.8 The multiple correlation coefficients computed between each possible pair of metrics and $\overline{T_{<ID}}$ computed from the search experiment data. 206

9.9 The multiple correlation coefficients computed between each possible pair of metrics and P_{fix} computed from the search experiment data. . 207

9.10 The multiple correlation coefficients computed between each possible pair of metrics and $\overline{T_{<fix}}$ computed from the search experiment data. 207

9.11 The multiple correlation coefficients computed between each possible pair of metrics and $\overline{T_{fix}}$ computed from the search experiment data. . 208

LIST OF FIGURES

FIGURE		PAGE
1.1	Static scene with three tank targets.	3
1.2	The general tasks involved in target acquisition.	6
2.1	A typical spatial neighborhood for Markov/Gibbs models of order 1-4.	13
2.2	The coordinate definitions for determining the displacement vector from the pixel with gray level i to the pixel with gray level j.	16
3.1	A target pattern extracted directly from a portion of the background is visible due to phase differences along its boundaries.	25
3.2	Progressive stages in the synthesis of a camouflaged target into a background, with blending along the boundaries.	26
3.3	The three 256x256 texture samples selected for GLC matrix visualization.	27
3.4	The GLC matrices of the pigskin texture for four selected displacements.	29
3.5	The GLC matrices of the raffia texture for four selected displacements.	30
3.6	The GLC matrices of the evergreen texture for four selected displacements.	31
3.7	The original wood image.	32
3.8	The GLC matrices extracted from the wood texture image for four selected displacement vectors: (a) and (b) small and large horizontal displacements, (c) and (d) small and large vertical displacements.	33

3.9 Spatial indexing for the GLC matrices extracted from the wood texture image: (a) for gray level transition $i = 4, j = 4$, (b) for gray level transition $i = 6, j = 10$. 35

3.10 Demonstration of how GLC ridge sharpness affects the resulting synthesized texture. In (a) the diagonal elements of each of the 12 GLC's is equal to the non-diagonal elements. In (e) the diagonal elements are 16 times greater in magnitude than the non-diagonal elements. . . 37

3.11 Demonstration of how flattening the shape of one GLC matrix in the model can affect the resulting synthesized texture. 39

3.12 (a) and (b) Two normal GLC matrices with different covariance matrices. (c) Best synthesis result using (a). (d) Best synthesis result using (b). 42

3.13 (a) and (b) Two normal GLC matrices with different covariance matrices. (c) Best synthesis result using (a). (d) Best synthesis result using (b). 43

3.14 (a) and (b) Two normal GLC matrices with different covariance matrices. (c) Best synthesis result using (a). (d) Best synthesis result using (b). 44

3.15 The synthesis process, from the initial random distribution (top row) to the best result (bottom row). The first column shows the image at each stage, the second column shows the GLC matrix for displacement [0 1] at each stage, and the third column shows the corresponding GLC error matrix (current GLC minus goal GLC). 45

3.16 Contribution to GLC error of pixels in the final synthesized image: (a) considering only displacements [x y] where $1 \leq (x \quad \text{and} \quad y) \leq 6$, (b) considering only displacements [x y] where $7 \leq (x \quad \text{or} \quad y) \leq 12$. . 49

3.17 Spatial indexing for the GLC error in the final synthesized image. Shown, for each displacement, is the sum of absolute error between the model for the synthesized image and the goal model. 50

4.1 An example of the XPET display in set-up mode. 54

4.2 A test subject using XPET to register his responses. 55

4.3 The distributions of discriminal processes for four possible stimuli. . . 62

4.4 A possible distribution of discriminal differences $(d_k - d_j)$, where the shaded area corresponds to the proportion p_{jk}. 63

4.5 The author collecting eye tracking data using ITEMS. 66

4.6 The chronological scheme of image presentation for ITEMS data collection. 70

4.7 The interconnectedness of the various hardware components of ITEMS. 71

4.8 The observer control sequence for I_SPY in experiment mode. 73

4.9 The general flow of the "I_TRACK" program. 77

4.10 The geometrical model of the eye tracking problem. 81

4.11 The vectors of interest in the eye tracking problem. 83

4.12 An illustration of the 28 calibration regions defined by the 40 calibration targets. 94

4.13 An illustration of four possible fixation point estimates (circles) along with the true locations of the targets (squares). The arrows represent the error vectors (estimate minus truth) for the estimates. 95

5.1 The 4-bar pattern that was used when superimposing the targets on the background. 98

5.2 The mean (μ) and standard deviation (σ) parameters of the 12 target patterns as well as the background pattern. 99

5.3 The twelve image stimuli used in the pilot experiment. 100

5.4 The stimuli plotted with their LCJ scale values and computed values for various metrics. 106

6.1 The binary mask showing the relative location of the 96x96 target field within the 256x256 background textures. 112

6.2 The stimuli used in the pigskin experiment. 113

6.3 The stimuli used in the raffia experiment. 114

6.4 The stimuli used in the evergreen experiment. 115

6.5 The 10 stimuli plotted using their LCJ scale values and the ACE measure computations for each of the three texture synthesis experiments. 121

7.1 The original texture images for the texture hybridization experiment. 124

7.2 The histogram to which the texture images where specified. This histogram is normal with a mean of 3.5 and a standard deviation of 1.5 gray levels. 125

7.3 The stimuli used in the texture hybridization experiment. 128

7.4 The 10 finite stimuli in the texture hybridization experiment plotted using the ACE metric computations and both the traditional and adjusted LCJ scale values. 135

7.5 The three images selected from the VisTex database. 137

7.6 The three 256x256 texture samples after a histogram normalization. . 138

7.7 The scheme that was used in creating the synthetic models from the real models. The texture models on the corners of the triangle shape were computed from the real textures. Texture models #4, #5, and #6 were created by combining the two real models at the neighboring corners and model #7 was created by combining all three. 139

7.8 The binary mask showing the relative location of the 224×224 target field within the 512×512 background textures. 140

7.9 The four image stimuli used in the training session. 144

7.10 The GLC matrices of the water texture for the four smallest displacements. 160

7.11 Plots of the energy in the GLC matrices for every displacement for the three original natural textures. 161

8.1 The six 256×256 real texture samples. 167

8.2 The binary mask showing the relative location of the 48×48 target field within each 256×256 background texture. 169

8.3 The three 256×256 hybrid texture experiment stimuli that are used in the training session. 170

FIGURE PAGE

8.4 The relative locations of the scale values along the perceptual continuum representing target distinctness. 174

8.5 The stimuli plotted with their LCJ scale values and computed values for various metrics. 178

9.1 The 52 targets in the search experiment plotted with their probability of detection and the computed values for various metrics. 202

9.2 The 52 targets in the search experiment plotted with their average time to detection and the computed values for various metrics. 203

9.3 The 52 targets in the search experiment plotted with their probability of fixation and the computed values for various metrics. 204

9.4 The 52 targets in the search experiment plotted with their average time to fixation and the computed values for various metrics. 205

E.1 Ground truth for image stimulus A. 238

E.2 Observer a's fixations for image stimulus A. 239

E.3 Image stimulus A. 240

E.4 Ground truth for image stimulus B. 241

E.5 Observer a's fixations for image stimulus B. 242

E.6 Image stimulus B. 243

E.7 Ground truth for image stimulus C. 244

E.8 Observer a's fixations for image stimulus C. 245

E.9 Image stimulus C. 246

E.10 Ground truth for image stimulus D. 247

E.11 Observer a's fixations for image stimulus D. 248

E.12 Image stimulus D. 249

E.13 Ground truth for image stimulus E. 250

E.14 Observer a's fixations for image stimulus E. 251

E.15 Image stimulus E. 252

E.16 Ground truth for image stimulus F. 253

E.17 Observer a's fixations for image stimulus F. 254

E.18 Image stimulus F. 255

E.19 Ground truth for image stimulus G. 256

E.20 Observer a's fixations for image stimulus G. 257

E.21 Image stimulus G. 258

E.22 Ground truth for image stimulus H. 259

E.23 Observer a's fixations for image stimulus H. 260

E.24 Image stimulus H. 261

E.25 Ground truth for image stimulus I. 262

E.26 Observer a's fixations for image stimulus I. 263

E.27 Image stimulus I. 264

E.28 Ground truth for image stimulus J. 265

E.29 Observer a's fixations for image stimulus J. 266

E.30 Image stimulus J. 267

CHAPTER 1

Introduction and Motivation

1.1 Camouflage on the Battlefield

On the modern battlefield, life and death are often decided in mere seconds. Anything that can give our soldiers even the slightest technological edge against the enemy could mean the difference between utter defeat and complete victory. Camouflage is an attempt to obscure the signature of a target and also to match its background. The goal is to make the detection and recognition of the target more difficult, whether the observer is a man, a machine, or both. Improved camouflage techniques will make our military forces more effective at achieving strategic and tactical objectives while minimizing sustained casualties.

Camouflage takes many forms in the U.S. Army. Our soldiers wear uniforms with coloration patterns intended to blend into their specific fighting environment. Occasionally soldiers paint their faces and other exposed skin areas, and attach various pieces of vegetation to themselves and their rifles to obscure their outline. Vehicles and equipment are painted in colors to match the battlefield environment, with special paint that is designed to absorb certain wavelengths of electromagnetic radiation and to scatter radar signals. But in the defensive, when units are in stationary positions, camouflage primarily takes the form of specialized nets that are spread over vehicles and equipment, staked to the ground at the edges, and

elevated and spread in the center by special supports.

At present, the emphasis in developing the coloration patterns for camouflage is to define four possible battlefield environments and to use patterns that are generally appropriate for each. The four environments are desert, woodland, urban, and arctic. Of the four, desert and woodland receive the most emphasis. The approach is to generate three or four patterns that seem to be good candidates, manufacture a prototype for each of the patterns, and then to qualitatively judge their relative effectiveness in the field. Modernizing the camouflage development and assessment processes with computer techniques could reduce time and cost requirements as well as improving the final result.

The ultimate aim of the research reported in this dissertation is the development of a robust and quantitative means for characterizing the signature strength of a target in a sensed image. The signature strength measurements should be closely correlated to the ease or difficulty of a human observer attempting to detect it. The real-world analogy is that of a military observer, perhaps an artillery spotter, looking upon the sensed image of a static scene to find camouflaged enemy targets that are reported to be in the area. In this situation, the observer may have some information of the size of suspected targets as well as other cues for which to look, such as straight lines. As an example of what an observer might see in this situation, see Figure 1.1. In the top portion of this image, there are three simulated tank targets possessing various patterns against various background patterns. The perceptual experiments reported in this work were conducted to learn about measuring the fundamental characteristics of target and background patterns that serve as impor-



Figure 1.1: Static scene with three tank targets.

tant discrimination cues for humans and to increasingly resemble the conditions of the real-world analogy.

Image metrics that can quantify target signature strength could ultimately be used to synthesize camouflage patterns and to assess camouflage effectiveness. The metrics could be particularly useful for "active" camouflage systems of the future, where a camouflage pattern is created in the field. This "site specific" camouflage would be designed for the particular location of the vehicle or structure. Computer control of the process would allow the soldiers to concentrate on their duties rather than pausing to erect the camouflage net, and could even allow a vehicle on the

move to constantly update its camouflage pattern as it moves across various terrain.

The metrics could also be useful for development and assessment of automatic target recognition (ATR) systems. Most current ATR research disregards the case of targets that are camouflaged, primarily because as of yet, there are no adequately robust methods for even uncamouflaged targets. But perhaps an even more promising application for the metrics is in computer assisted targeting, where computer and human cooperate. Although ATR is receiving a tremendous amount of attention by the military research community, commanders in the field may be reluctant to allow the targeting of weapons to be completely controlled by computers. In light of our most recent large scale military engagement in the Persian Gulf region, where more casualties on our side were caused by friendly fire than by enemy fire, it is likely that any targeting system actually deployed in the future will include a human in the loop at some point. Computer assisted targeting could give us a tremendous advantage on the battlefield, saving our soldiers crucial seconds in combat. Probably the most helpful role that the computer could play would be to assist in the identification of candidates for attention, i.e., the search process. This is especially so since the visual displays used by tank gunners have a rather narrow field of view. The human will still most likely perform the discrimination process, as it will be his ultimate decision to pull the trigger.

1.2 Sensed Images and Image Texture

An important phenomenon to remember in target signature modeling is that the appearance of an object in a sensed image varies greatly, dependent upon both

environmental conditions and sensor parameters such as resolution and frequency sensitivity. A warm vehicle may be practically indistinguishable from its background in visible wavelengths but be clearly distinct in a thermal image. The same thing can be true of the vehicle in the darkness when viewed through a night vision sight. Very often, the soldier can select one or another sensor, but the output from any single sensor is a one-channel image, meaning that the image can be represented by pixels that each have only one value. In this dissertation all images used are gray scale, or single-channel.

In this context, the signature strength of a target in a sensed image is equivalent to the distinctness of the image pattern representing the target from the pattern of its specific background. Certainly, this is strongly influenced by the first-order (histogram) statistics of the target and background. Such statistics are capable of characterizing the relative brightness and variance of patterns. But often the perceptual differences between target and background are functions of image texture [ST92], which cannot be adequately described by only first-order statistics that ignore spatial relationships. This may be especially true if the target is camouflaged. To generalize the wide variety of potential target and background patterns in a sensed image, we will utilize second-order probabilities as contained in gray level cooccurrence (GLC) matrices as a model for image texture.

1.3 Tasks Constituting Human Target Acquisition

It is known [Cha86, CH95] that a human observer performing target acquisition fixates only a small portion of the visual field at any given time. This fixation lasts

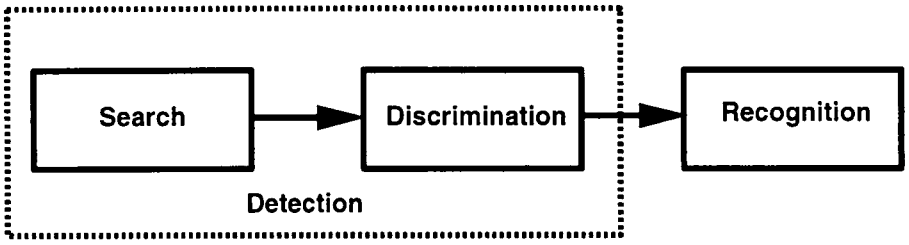


Figure 1.2: The general tasks involved in target acquisition.

for a short duration, and is followed by a quick movement, or saccade, to a different location in the visual field. For the purpose of defining the scope of the perceptual studies reported in this dissertation, we will consider target acquisition to involve target detection followed by target recognition, as illustrated in Figure 1.2. The detection task is that which establishes the existence and location of the object. Recognition is the task of determining the characteristics of the object which indicate its identity, such as man-made or natural object, vehicle or fixed structure, military or civilian, wheeled or tracked, and friend or foe. Further, we will consider target detection to consist of the combination of the individual tasks of search and discrimination. Search is the process of locating areas of a scene in which to direct our attention. Discrimination is the process of segregating a potential object from its immediate background. This approach is very similar to the conclusion of O'Kane *et al.* [OWD93]. In this dissertation we are studying the target detection task, comprising search and discrimination, without considering recognition. The task of recognition is generally performed by comparing the shape and structure of the object to some knowledge base, whether it be the experience and memory of a

human observer or the digital memory of computer database files. In the perceptual experiments reported in this dissertation, the targets in any one experiment are a constant size and shape.

There is a physiological basis for considering search and discrimination as two distinct processes. This is that there are two types of light receptors distributed about the retina. Rods are distributed over a majority of the retina, while cones are distributed mainly in a small portion of the retina called the fovea. Several rods are connected to a single nerve end, while each cone is connected to its own nerve end [GW92]. The result of this is that most of the resolving power of the human eye is concentrated in the fovea. A human fixates a particular point in his visual field by rotating the eyeball until the image of that point falls on the fovea. The peripheral portions of the visual field contain considerably less discernible detail, meaning that it is often necessary for the observer to fixate the object to segregate it from a complex background pattern.

The result is that in order to detect targets in a wide area, the observer must rely largely upon cues from his peripheral vision to direct him to fixate particular locations. It is only after he has obtained greater resolved details of the location by fixating it that he can perform the task of discrimination. Indeed, this result has been observed by other researchers studying eye movements during search and discrimination [Cha86,CH95], in that the saccades during search are longer and more wandering, while saccades during discrimination are shorter and more concentrated.

1.4 Research Methodology

Since we wish to generalize the wide variety of potential target and background patterns in a sensed image, we utilize a model for image texture. Experiments are performed to study the perceptual qualities of a texture in relationship to the shape of the second-order probability distributions within its gray level cooccurrence (GLC) matrices. A texture synthesis technique is used to generate realizations of image texture possessing specified second-order probabilities. The GLC matrices of these synthesized textures are displayed so as to visualize the information content of the matrices and to compare this to the perceptual qualities of the texture. The results of these experiments lead to the design of GLC-based texture metrics to measure the distinctness of target and background patterns.

To study human target detection, data from psychophysical experiments using human observers is collected and analyzed. The psychophysical data are used to evaluate various target signature metrics as well as to develop new metrics to explain the data. Discrimination is studied separately from search by using stimulus images that contain a target pattern in a constant location known to the observer. In the discrimination experiments, human observers view image stimuli consisting of textured targets and backgrounds and provide responses as to the perceived distinctness of the targets. Each experiment begins with the creation of the visual stimuli. The precise nature of these stimuli will depend on what we hope to learn from the particular experiment. The various variables are carefully controlled, so that those that we wish to study are allowed to vary while those not of interest in the experiment are equal in all of the stimuli. Experiments are performed using

both synthetic patterns possessing specified texture statistics and natural patterns extracted from real-world imagery. As we iteratively test and refine the metrics over a series of experiments, we allow variation in more variables, so that we more closely approach the conditions of the real-world analogy.

The actual conduct of the discrimination experiments utilizes an unique testbed, developed to handle the presentation of stimuli, registering of observer responses, and control of environmental conditions. This testbed, called the X-Based Perceptual Experiment Testbed (XPET), implements the experimental methods of paired comparison and triadic comparison. It features a graphical user interface, which allows the observer to make his responses to stimulus pairs by simply using the mouse to point and click on his selections. The review and analysis of the resulting data makes use of psychological scaling techniques as well as multivariable statistical methods. It will be important to refer back to the original stimuli in order to interpret the statistical data. Intuition will play an important role in the identification of the important perceptual cues that were utilized by the observers in making their responses. It should be expected that every observer may use different criteria in their judgments and that every stimulus may be interpreted differently by the observers. Most importantly, the data is reviewed with an eye out for any deviation from expectation, as it is these deviations that yield new insights into the process of human target detection.

The scope of the investigation is extended with an experiment in which observers must perform both search and discrimination together to identify targets in a scene. By integrating search and discrimination, we are able to study not just how dis-

tinct a target is in its local background, but how this affects the time required for observers to detect it and even the likelihood it will be detected at all. During the experiment, the observer's fixation points in the image are measured by tracking his eye movements. The computer testbed developed for this experiment is called the Integrated Testbed for Eye Movement Studies (ITEMS). From the data, various statistics are computed for each target to describe how easily the observers located it, including the likelihood the target was fixated or identified and the time required to do so. This final experiment is the one most closely resembling the real-world analogy.

1.5 Organization of the Dissertation

This dissertation begins with a literature review in Chapter 2. Relevant literature in the areas of image texture analysis and modeling as well as studies of human visual search and discrimination are presented. Chapter 3 presents some experiments in which the perceptual qualities of a texture are studied in relationship to the shape of its cooccurrence probability distributions within its GLC matrices. Chapter 4 describes the development of various computer testbeds and selection of experimental procedures appropriate for experiments in human performance of search and discrimination. Chapter 5 through Chapter 9 describe the perceptual experiments in chronological order, progressing from first-order statistics to second-order statistics to hybrid synthesized textures to natural textures to search integrated with discrimination. Chapter 10 concludes the dissertation.

CHAPTER 2

The Past: A Review of the Literature

2.1 Methods of Image Texture Analysis and Modeling

First, the wide range of texture modeling techniques that have been used by computer vision researchers will be briefly surveyed. Then, since a large portion of the research conducted for this dissertation to this point is based on gray level cooccurrence (GLC) matrices for modeling image texture, this one method is reviewed in detail in Section 2.1.2.

2.1.1 General Methods

Tuceryan and Jain [TJ93] present a taxonomy of a wide variety of methods for modeling image texture. They classify these methods into the four categories of statistical, geometrical, model based, and signal processing methods. The statistical methods are given as cooccurrence matrices and autocorrelation features. The geometrical methods are given as Voronoi tessellation features and structural methods. The model based methods are given as random field models and fractals. The signal processing methods are given as spatial domain filters, Fourier domain filtering, and Gabor and wavelet models. There are several questionable points to this taxonomy. For example, the model based methods are described as those methods based on a model that can not only describe texture, but also synthesize it. However,

there exists a technique for synthesizing texture patterns possessing a particular set of cooccurrence matrices [GDM85, Gag81], but cooccurrence matrices are listed as a statistical method rather than model based. Additionally, cooccurrence matrices have been shown to consist of the same information as random field models but only tabulated in a more general way [PM95]. Nevertheless, this taxonomy is adequate for defining the organization of a review of the literature in this area.

Statistical Methods

Statistical methods of texture analysis consider the spatial distribution of gray values. Cooccurrence matrices will be discussed in Section 2.1.2. Mao and Jain [MJ92] propose using simultaneous autoregressive models for texture. The autoregression model, the Fourier power spectrum, and the autocorrelation functions are all the same subset of second-order statistics, and each of them can be mathematically derived from another [TT90]. The approach is typically to divide the frequency domain into rings and wedges, and then to use the computed energy in each of the regions as texture features.

Geometrical or Structural Methods

The geometrical, or structural, methods characterize image texture as being composed of texture primitives. The various methods aim to describe both the texture primitives and their spatial placement [Har79]. Tuceryan and Jain [TJ90] propose extracting texture primitives using the Voronoi tessellation of the image. This technique essentially defines the local spatial neighborhoods in an image and describes the local spatial distributions of texture primitives in the shapes of Voronoi polygons.

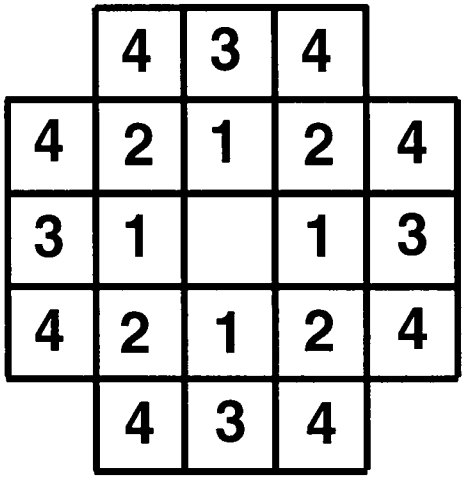


Figure 2.1: A typical spatial neighborhood for Markov/Gibbs models of order 1-4.

Blostein and Ahuja [BA89] extract texture primitives by examining the response of Laplacian of Gaussian (LoG) filters at multiple scales. Tomita and Tsuji [TT90] extract texture primitives by performing a medial axis transform on the connected components of a segmented image. Zucker [Zuc76] regards real textures to be distorted versions of ideal textures. The texture primitive placement rule is defined for the ideal texture by an isomorphic graph and then transformed to generate the real texture.

Model Based Methods

A Markov random field models a texture by assuming that the gray level of a pixel depends only on the gray levels of the neighboring pixels. The neighborhood of a pixel is defined in various ways according to the order of the model. Figure 2.1 shows a typical spatial neighborhood for models of order 1-4. The neighborhood of order p contains all points with labels $\leq p$ [EP94]. A Gibbs random field assigns an exponential probability mass function to the entire neighborhood [PEP91]. Random

field models are well suited for texture synthesis, either by Markov chain [YH79, MSM81] or the Metropolis spin-flip algorithm [MRR+53, CCB85, Car92].

Fractals are best suited to model the self-similarity of textures at different scales, and are also suited for texture synthesis [Dod87]. The fractal dimension is a measure of the roughness of a texture. Other features computed from fractals are lacunarities and fractal signatures [DD94].

Signal Processing Methods

The earliest attempts to apply spatial domain filtering to image texture concentrated on measuring edge density per unit area. This measurement is usually computed using simple edge masks [RW90]. Fourier domain filtering attempts to decompose an image into its frequency and orientation components. This method appears to parallel human preattentive vision mechanisms, as it is known that the visual cortex contains neurons that are tuned to specific frequencies and orientations [HW90]. To effectively analyze texture, the Fourier analysis must be localized in the spatial domain. This is usually accomplished by the window Fourier transform, where the image is multiplied by a localized window function before the Fourier transform is computed. If the window function is Gaussian, then it is called a Gabor transform [FS89]. Two-dimensional Gabor filters have been shown to optimally and uniquely achieve simultaneous localization in the spatial and frequency domains [BCG90]. The wavelet transform is accomplished in a similar way, but using a window with a width that increases as the central frequency of the analyzing filter is increased [SS93].

2.1.2 Gray Level Cooccurrence (GLC) Models

A GLC model is essentially a 4-dimensional representation of the second-order (cooccurrence) probabilities of a texture pattern. As such, the GLC model is usually classified as a statistical method of texture analysis [TJ93]. For a given spatial displacement vector of $d = [x\ y]$, the joint probability of a pixel m at location (m_x, m_y) having a gray level i and the pixel at location $(m_x + x, m_y + y)$ having a gray level j in the texture is $P(i, j|d)$. If we consider every possible pair of gray levels i and j, then we obtain a 2-dimensional joint probability distribution for every displacement vector d. This is represented with a $G \times G$ GLC matrix of cooccurrence probabilities, where G is the number of possible gray levels. Varying x and y to cover a spatial neighborhood, we obtain a 4-dimensional set of cooccurrence probabilities indexed by i, j, x, and y, i.e, two gray levels and two spatial coordinates. In that GLC matrices consist of tabulated second-order probabilities, they contain essentially the same information as some other texture analysis tools, such as Gibbs/Markov random fields [EP94].

Figure 2.2 shows the spatial coordinate definitions for determining the displacement from the pixel with gray level i to the pixel with gray level j. Notice that using symmetry $P(i, j|[x\ y]) = P(j, i|[-x\ -y])$, so we need to consider only displacements in directions varying over 180^o. Thus, while the x-coordinate for the displacement vector is varied from $-nx$ to $+nx$, the y-coordinate is varied only from 1 to ny, plus there are nx additional displacements from [1 0] to [nx 0]. The total number of displacements considered is therefore $(2 \times nx \times ny) + nx + ny$, and for each displacement we obtain a $G \times G$ GLC matrix of joint probabilities.

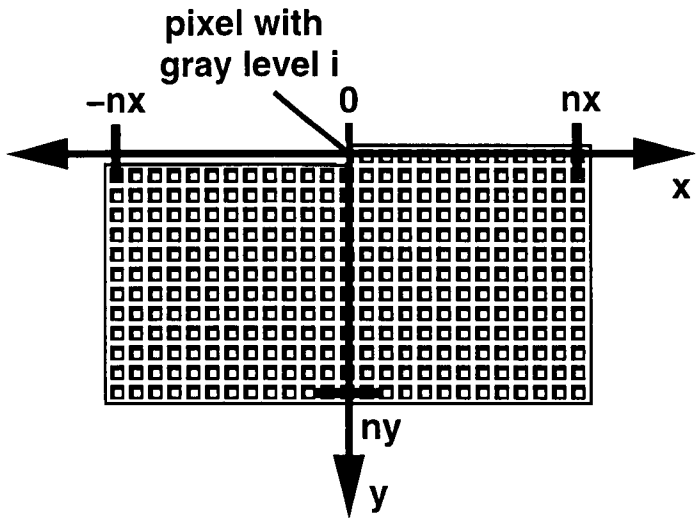


Figure 2.2: The coordinate definitions for determining the displacement vector from the pixel with gray level i to the pixel with gray level j.

The use of GLC matrices for image texture analysis became popular after Haralick, *et al.*, used features computed from GLC's for image classification [HSD73]. In several studies to compare the relative power of various texture analysis techniques to perform texture discrimination, GLC matrices generally outperformed other methods [WDR76, CH80, OD92]. GLC's have also been used for object detection [THCG84], scene analysis [HTCP86], as well as texture synthesis [Gag81, GDM85, Loh94]. Other studies have demonstrated the wealth of texture information contained within GLC's [FVPBGG87, PSO90, GK90].

2.2 Studies of Human Visual Search and Discrimination

Researchers often make the distinction between preattentive and attentive vision processes in humans. Preattentive, or early, vision is considered to occur effortlessly, within the first moments of the presentation of a stimulus [Mar76]. Attentive vision

is that which requires some cognition, and is presumed to occur only if the observer is allowed to study the stimulus. Although this is a popular dichotomy for those who wish to understand the physiological processes that occur in human vision up to and including the visual cortex, our interest is in predicting the overall result, *i.e.*, modeling the difficulty that a human observer will encounter in detecting the target. As mentioned previously, the probability that an observer detects a particular target in any particular instant is the probability that the target is within his foveal area times the probability that he will detect the target, given that it is located in his foveal area. We see that the observer must first **search** for a place to fixate, and then must perform **discrimination** to decide if he is fixating the target. For our purposes, it is more convenient to study the two separate tasks of search and discrimination, and realize that both preattentive and attentive processes may be occurring simultaneously.

2.2.1 Discrimination

The first significant studies of human pattern discrimination were conducted by perceptual psychologist Bela Julesz, *et al.* [Jul62, CJ78, CJG78]. Julesz's original conjecture was that a human cannot preattentively discriminate between patterns with identical first- and second-order statistics. After finding several counterexamples to his own conjecture [JGV78], he theorized that preattentive discrimination was based on local conspicuous features called "textons," which are elongated blobs with specific properties [Jul81, JB83]. Textons are essentially the same features as what other researchers have called texture "primitives," "elements," or "tiles."

Other perceptual psychologists have studied similar issues. Callaghan [Cal89] used textures composed of arrays of various geometric shapes to investigate the relationships between hue, shape, and orientation cues in texture segregation. Analyses revealed no dominance of one property over another. Sutter *et al.* [SBG89] used channeled Gabor filters to predict observers' segregation of array textures. Bergen and Landy [BL91] also use banks of oriented filters and at multiple resolutions to mimic human texture segregation.

Computer vision researchers have also studied human pattern discrimination. There has been a significant effort among them to develop computational measures of local clutter. Schmieder and Weathersby [SW83] recognized the role of background clutter on target discrimination. Their clutter metric was based on the variance of pixel gray levels in cells surrounding the target. Waldman, *et al.* [WWHL88], define a clutter measure based on the differences between the target and the background in the length of their "typical" texture tiles. This information was inferred from the relative strength of the main diagonals of the GLC matrices computed from the background and from the target. Shirvaikar and Trivedi [ST92] utilize GLC matrices and the size of the target to define clutter. Their metric is based on the inertia measure which also reflects the relative strength of the main diagonals of the GLC matrices. Rotman et al. [RTK94] measure local clutter using the histogram of pixel intensities after the image has been processed with a difference of offset Gaussians (DOOG) filter to enhance edges.

Other computer vision researchers have attempted to identify the features of image texture which are important for perception. Tamura *et al.* [TMY78] asked

human observers to judge texture patterns in terms of coarseness, contrast, directionality, line-likeness, regularity, and roughness, and then estimated psychological scale values for each texture for all six features. They also defined computational measures of each of the six features, and then computed rank correlations between the psychological scale values and the computational measures. The rank correlations were generally high, but the authors were unsuccessful at using combinations of the computational measures to predict human judgments of similarity among the textures. Amadasun and King [AK89] performed similar experiments with similar results using features they call coarseness, contrast, busyness, complexity, and texture strength. Rao and Lohse [RL93] asked observers to group texture patterns into as many classes as desired, allowing the observers to decide their own similarity criteria for grouping. They performed both cluster analysis and multidimensional scaling on the resulting data. Their conclusion is that observers tended to utilize periodicity, directionality, and structural complexity to judge the textures.

2.2.2 Search

Some of the earliest pioneering studies of human visual search were performed by the Russian biophysicist Alfred Yarbus [Yar67]. Yarbus recorded the eye movements of observers presented with various photographs of scenes and objects. Among many things, Yarbus learned that an observer's search process is determined both by the nature of the objects in the scene and the particular information about the scene the observer hopes to obtain. Juhola [Juh86] describes a syntactic method for recognizing saccadic eye movements from digitalized eye position signals. The

signals are transformed into a symbol string, valid saccade beginnings are searched for, and thereafter the whole saccade is identified by using a parser. Hacisalihzade *et al.* [HSA92] labeled regions of interest in images as states in a Markov process, then modeled sequences of fixations with transition probability matrices. They also present a method of measuring the similarity of two different sequences. Tidhar *et al.* [TRA+94] conducted experiments using synthetic rural scenes containing various vehicular targets. They compared detection probability to several measures of global clutter. Rotman *et al.* [RKG94] conducted similar experiments, but compared the popular fixation points in the scene to the value of clutter metrics computed within square blocks of the image. Stark *et al.* [SYTN93] recorded the eye movements of observers identifying targets in both 2-D and 3-D scenes. The 2-D scenes had targets randomly distributed throughout the entire image, in the sky, mountains, etc. The targets in the 3-D scenes were vehicles located in natural locations relative to the other objects in the scenes: on roads, in parking lots, etc. The search paths of the 2-D scenes tended to be in a regular scanning pattern, while the searches of the 3-D scenes were based on the spatial content of the scene. Nicoll [Nic94] describes a "neoclassical" field-of-view search model. This model is similar to the Markov process of Hacisalihzade *et al.* [HSA92], with the inclusion of one additional state for "wandering," when the observer is performing a general search for another interesting point to fixate. Probabilities are assigned to all possible state transitions, and the detection process for each target is modeled as an exponential function of time. The model is flexible in that it forms the structure of a workable model of visual search, but presumes no particular method for obtaining the state transition

probabilities or the exponential detection rates. Generally, the search process is considered completely independent of the various detection processes. The detection process for any one point of interest is independent of the detection processes of the other points of interest. However, the higher order detection correlations (such as the probability that neither target A nor target B has been detected) are linked by the search process, by the simple fact that the observer cannot attend to two locations at the same time.

CHAPTER 3

Experiments in the Perceptual Qualities of a Texture Contained Within its Cooccurrence Probabilities

In this chapter, we present a technique for visualization and analysis of the structural information that is contained in a gray level cooccurrence (GLC) model of a texture. This is accomplished by observing 2-dimensional components of the 4-dimensional model, indexed by both the gray level coordinates and the spatial coordinates. The spatial indexing in particular allows several observations to be made about the texture tile characteristics. We then synthesize a new pattern with similar second-order probabilities, using the Metropolis spin-exchange algorithm. We examine the GLC models of the image at various points during the process to make some observations about what is happening during synthesis. We analyze the GLC error that still remains in the synthesis result, and develop a technique for identifying the pixels which contribute the most to the total error.

3.1 Texture Synthesis using GLC Models

As explained earlier (Section 2.1.2), a GLC model is essentially a 4-dimensional representation of the second-order (cooccurrence) probabilities of a texture pattern. For example, for a given spatial displacement vector of $d = [x \quad y]$, the joint probability of a pixel m at location (m_x, m_y) having a gray level i and the pixel at location

$(m_x + x, m_y + y)$ having a gray level j in the texture is $P(i, j|d)$. Varying x and y to cover a spatial neighborhood, we obtain a 4-dimensional set of cooccurrence probabilities indexed by i, j, x, and y, i.e, two gray levels and two spatial coordinates. Using a GLC model of image texture, texture synthesis becomes a process of generating a random pattern with a particular set of GLC probabilities. We begin with a random image possessing the desired histogram, and then sequentially modify this image to bring its second-order probabilities closer to the desired probabilities. There are two basic algorithms that have been used to accomplish this iterative modification of the image. In the Metropolis spin-flip algorithm [MRR+53, Car92], individual pixels are visited randomly and their gray levels are changed if such a change brings the second-order probabilities closer to the desired probabilities. This is essentially the method of Gagalowicz and Ma [GDM85]. In the Metropolis spin-exchange algorithm, introduced in image analysis by Cross and Jain [CJ83] and applied to GLC models by Ravichandran et al. [RKT94] and by Lohmann [Loh94], pairs of pixels are chosen randomly to be considered for a gray level exchange. If exchanging the gray levels of the two pixels is deemed beneficial then the exchange is executed. A temperature annealing schedule can also be used [PEP91], which essentially means that the exchange occurs with probability $\min(1, \exp(-\Delta E/T))$, where ΔE is a measurement of the change in probabilities and T is a temperature which is gradually reduced. The synthesis method used to generate the patterns to be used as stimuli in our experiments is the spin-exchange algorithm, but without annealing. This is the method most often used, and is preferred over spin-flip because it maintains the image histogram (first-order probabilities) during the process.

For the synthesis process, we need a synthesis error metric in order to make the decision as to whether or not a gray level exchange moves the cooccurrence probabilities closer to the desired probabilities. This metric essentially gauges how different two GLC texture models are. We define the cooccurrence error (CE) for any displacement d as the sum over all pairs of gray levels of the absolute difference of the two GLC matrices:

$$CE(d) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |P_c(i,j|d) - P_g(i,j|d)|, \tag{3.1}$$

where G is the number of gray levels used, d is an index for the displacement vector, P_c is the GLC matrix of the current image, and P_g is the GLC matrix of the synthesis goal texture model. Thus, the maximum CE for any GLC matrix is 2.0 (if the two probability distributions have zero overlap) and the minimum is 0.0 (if the two are identical). The average cooccurrence error (ACE) is simply the average value of CE over all displacements considered, and can be used to gauge how different two GLC texture models are:

$$ACE = \frac{1}{D} \sum_{d=1}^{D} CE(d), \tag{3.2}$$

where $D = (2 \times nx \times ny) + nx + ny$ is the total number of $[x\ y]$ displacement vectors considered.

If we synthesize a texture in a location within a background, boundary effects can cause the pattern to be distinct even if the texture is matched well. An example of target/background boundary effects is illustrated in Figure 3.1. A target pattern in the shape of a tank profile is placed in a background wood grain pattern. Even though the target pattern was extracted directly from another area in the same background texture, it is visible due to phase differences along its boundaries.

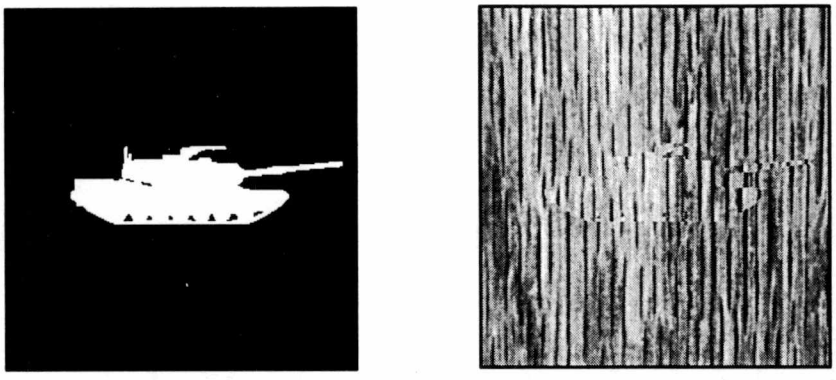


Figure 3.1: A target pattern extracted directly from a portion of the background is visible due to phase differences along its boundaries.

These boundary effects can be minimized if we "blend" the synthesized target into the background, such that only target pixels can be considered for gray level exchanges, but the intensity values of the background pixels are taken into account when exchanges are considered. As the synthesis process attempts to minimize the GLC error, pixels near the border will tend to fit in with the tiling structure of the background texture in that area. This blending process is an attempt to minimize the edge effects at the border between target and background. Figure 3.2 is an example of this, showing progressive stages in the synthesis of a camouflaged target into the background, where the pixels along the target/background boundary are blended to reduce phase differences. The target pattern initially consists of randomly placed pixels with first-order (histogram) probabilities the same as the background, and then the locations of these pixels are adjusted in order to bring the second-order probabilities closer to those of the background.

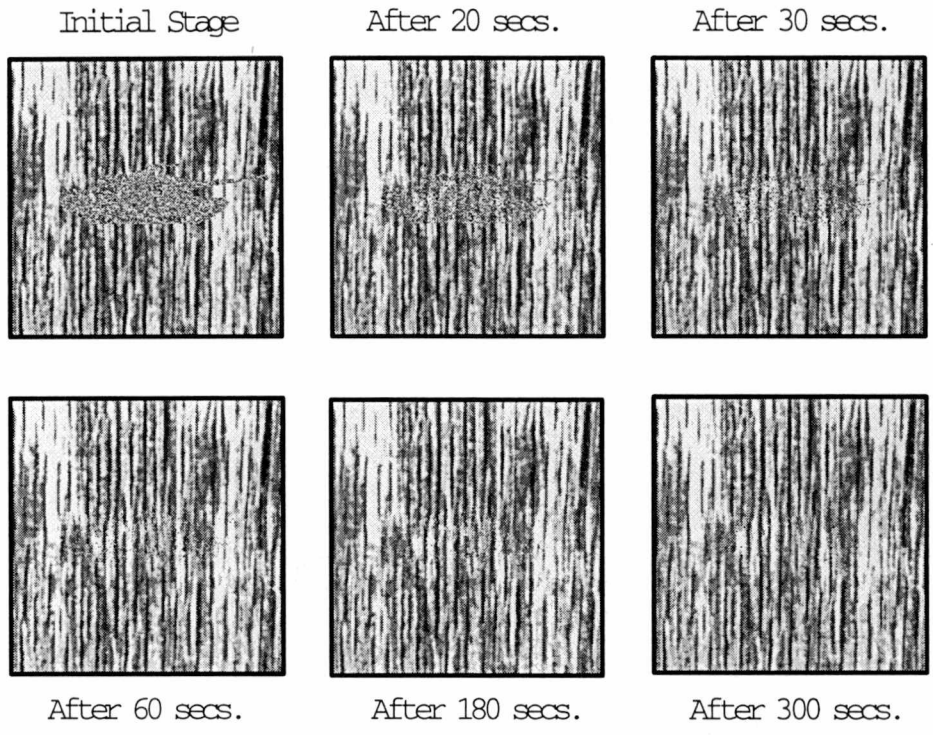
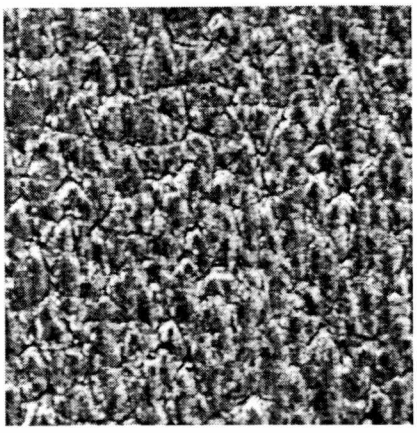


Figure 3.2: Progressive stages in the synthesis of a camouflaged target into a background, with blending along the boundaries.

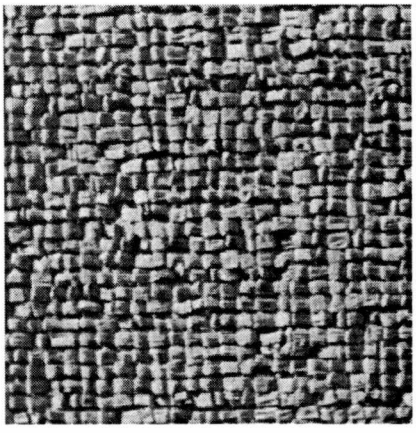
3.2 Visualizing the GLC Model of a Texture

3.2.1 Images for GLC Visualization

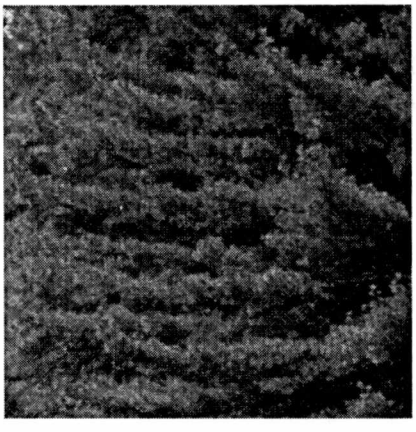
We can visualize GLC matrices if we plot them as gray scale images. In this section, this is done for three different texture patterns selected as case studies. Two of the patterns, a pigskin texture and a raffia weave texture, were extracted from the Brodatz album of texture images [Bro66]. In Rao's taxonomy of these textures [Rao90], the pigskin texture is described as a "disordered" texture, and the raffia weave is described as a "strongly ordered" texture. The third texture pattern was extracted from a database of digitized photographs of scenes at Fort Belvoir, Virginia. This



(a) Pigskin



(b) Raffia



(c) Evergreen

Figure 3.3: The three 256x256 texture samples selected for GLC matrix visualization.

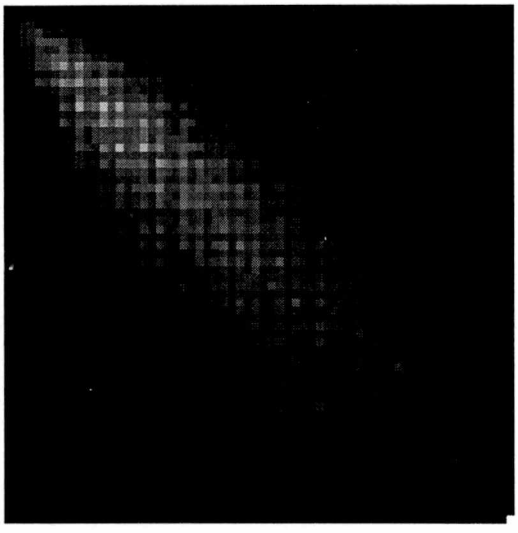
texture is a portion of an evergreen tree, and was included as an example of a "real-world" texture. The three raw textures are shown in Figure 3.3.

Figure 3.4 through Figure 3.6 show four selected GLC matrices displayed as gray scale images. We have quantized each of the three images to 64 gray levels, calculated the 144 GLC matrices for all displacements of up to 8 pixels, and selected four of these GLC matrices to display as gray scale images. In each figure, (a) and (b) are the GLC matrices for horizontal displacements of 1 pixel and 8 pixels, while

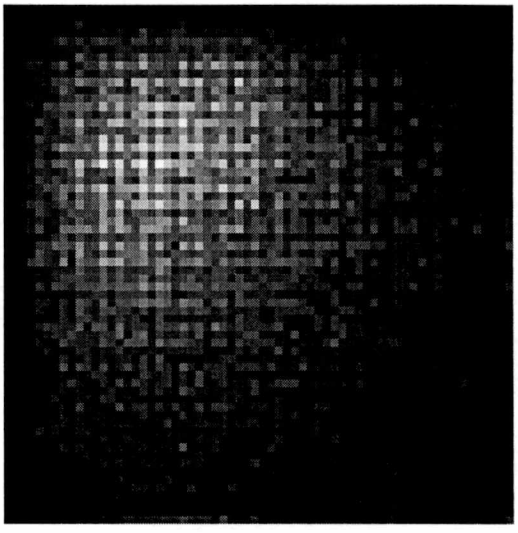
(c) and (d) are for vertical displacements. From these GLC visualizations, we can see that the GLC matrices for the three textures have various forms. Since a GLC matrix is essentially a joint probability distribution, all elements in a particular GLC matrix will sum to unity. Thus, we can say the only differences between the GLC matrices is in their **topographic shape**. Another observation from these GLC matrices calculated from existing patterns is that most have higher probability levels concentrated along the diagonal from upper-left to lower-right. The diagonal elements represent pairs of pixels that have the same gray level. Thus, it makes sense that textures with some form of tile structure will have GLC matrices with strong diagonals, especially for the shortest displacements [PE92].

3.2.2 Surface Plots for GLC Visualization

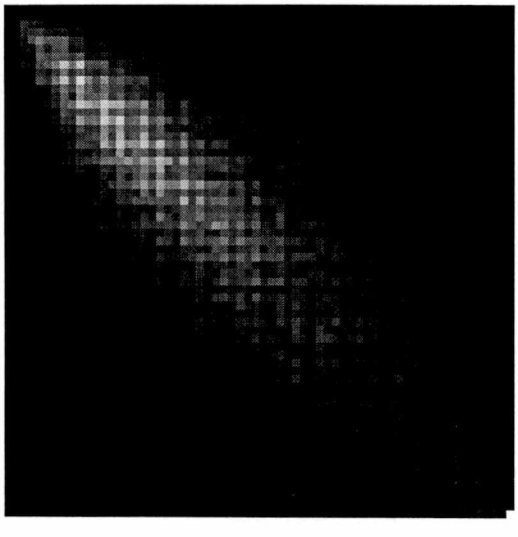
The texture that we have chosen to illustrate another form of GLC visualization is shown in Figure 3.7. This 256 × 256 gray scale image is a portion of texture D68 (wood grain pattern) from the Brodatz album of textures [Bro66]. The most distinctive features of this texture are the dark, narrow, elongated vertical stripes. Also evident are some larger dark or light colored areas, of various shapes. Before the GLC texture model was extracted, the image was fitted to a 16 gray level Gaussian, or normal, histogram. This histogram has a mean of 7.5 and a standard deviation of about 3 gray levels (where gray level 0 is black and 15 is white). The GLC model was calculated from the histogram-normalized wood image. The second-order probabilities were calculated for all displacements up to $nx = 12$ and $ny = 12$. The total number of displacements considered is therefore $(2 \times nx \times ny) + nx + ny =$



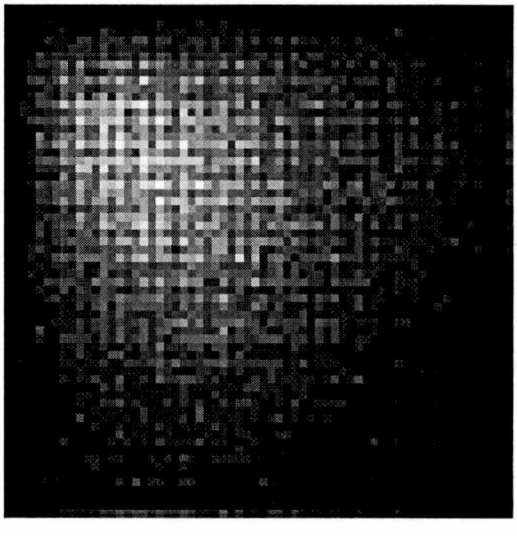
(a) Pigskin GLC (1,0)



(b) Pigskin GLC (8,0)

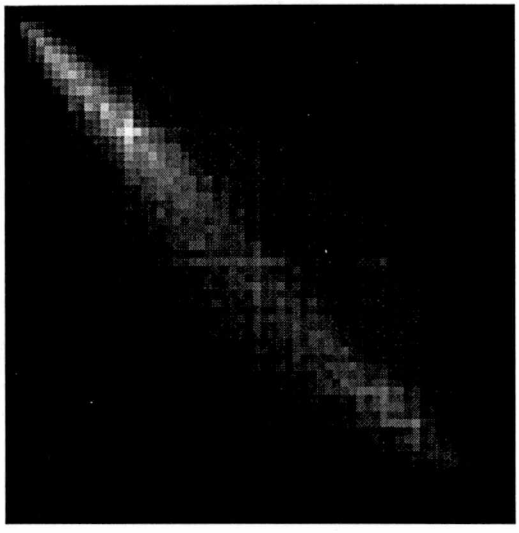


(c) Pigskin GLC (0,1)

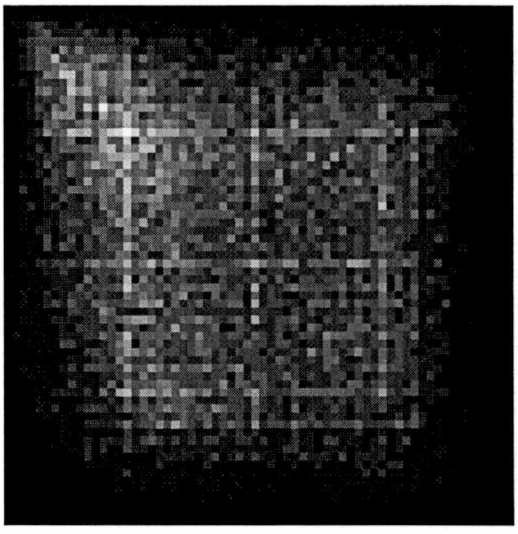


(d) Pigskin GLC (0,8)

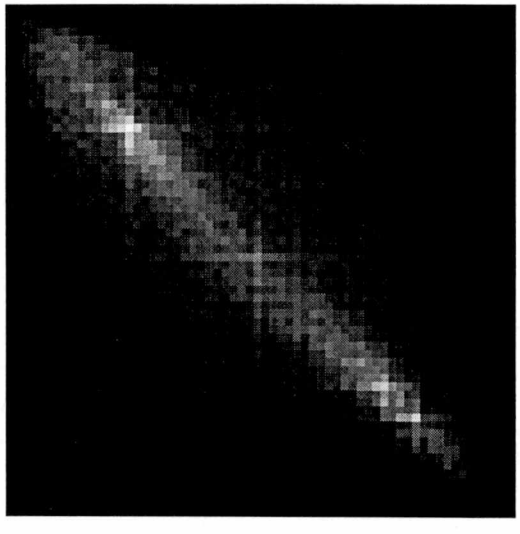
Figure 3.4: The GLC matrices of the pigskin texture for four selected displacements.



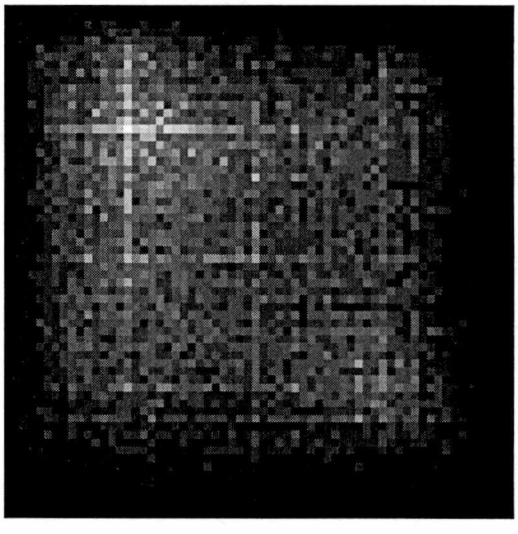
Raffia GLC (1,0)



Raffia GLC (8,0)

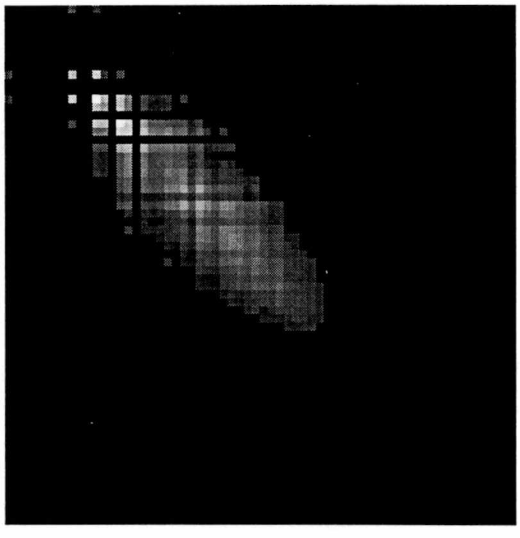


Raffia GLC (0,1)

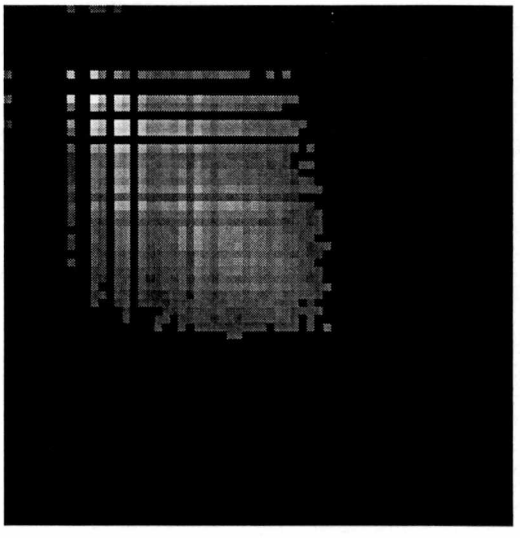


Raffia GLC (0,8)

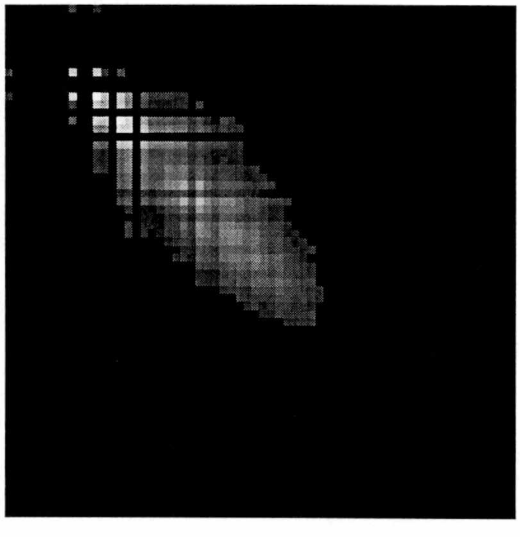
Figure 3.5: The GLC matrices of the raffia texture for four selected displacements.



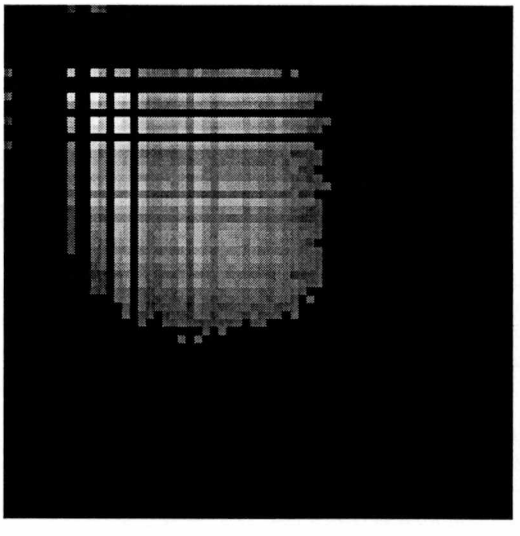
Evergreen GLC (1,0)



Evergreen GLC (8,0)



Evergreen GLC (0,1)



Evergreen GLC (0,8)

Figure 3.6: The GLC matrices of the evergreen texture for four selected displacements.

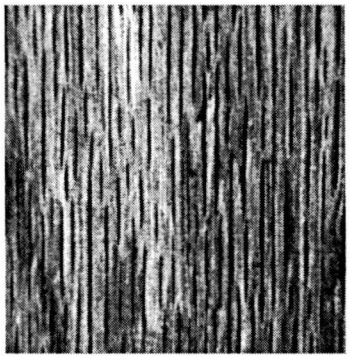
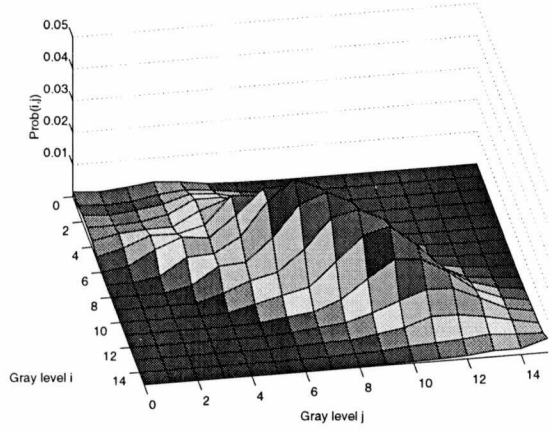


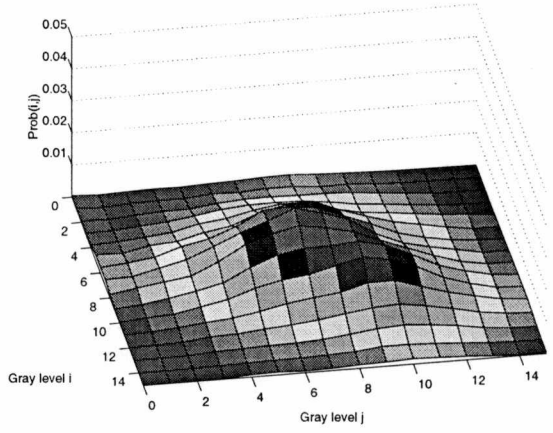
Figure 3.7: The original wood image.

$(2 \times 12 \times 12) + 12 + 12 = 312$, and for each displacement we create 16×16 GLC matrix of second-order probabilities, since there are 16 gray levels. Since this GLC model is 4-dimensional, we can examine the information contained in it by plotting the 2-dimensional surfaces of the GLC matrices for particular $[x \quad y]$ displacement vectors. This is what is shown in Figure 3.8 for the GLC model computed from the wood texture. We have selected just four representative displacements (out of the 312 in our model) and plotted the 16×16 GLC matrix for each. In (a) is shown the GLC matrix for displacement [1 0], which means that the pixel with gray level j is the horizontal neighbor of the pixel with gray level i. In (b) is shown the GLC matrix for the longest (12 pixels) horizontal displacement considered. In (c) and (d), the GLC matrices for the corresponding small and large vertical displacements are shown.

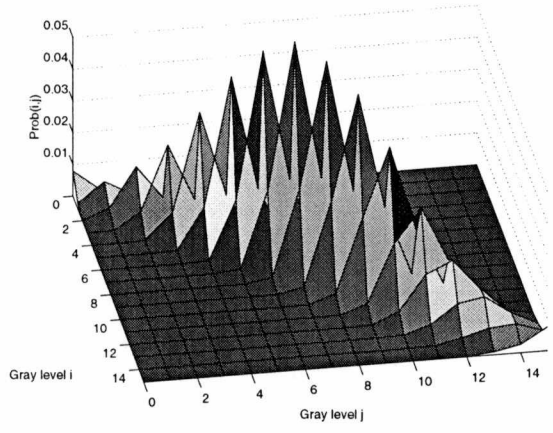
As mentioned previously, since the elements along the main diagonal of the matrix represent the likelihood of the two pixels having the same gray level, the relative strength of the elements along or near the main diagonal represent texture element, or tile, structure at that particular distance. Thus the main diagonal of the wood



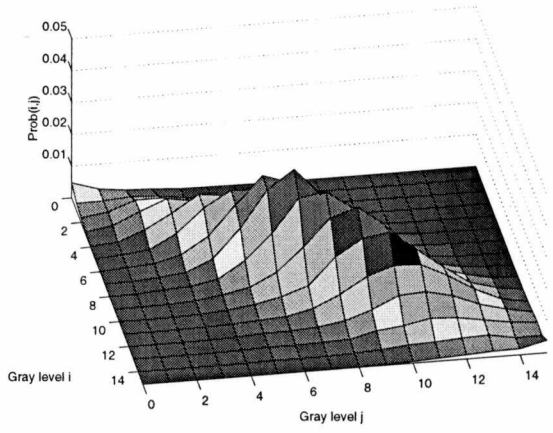
(a) Displacement [1 0]



(b) Displacement [12 0]



(c) Displacement [0 1]

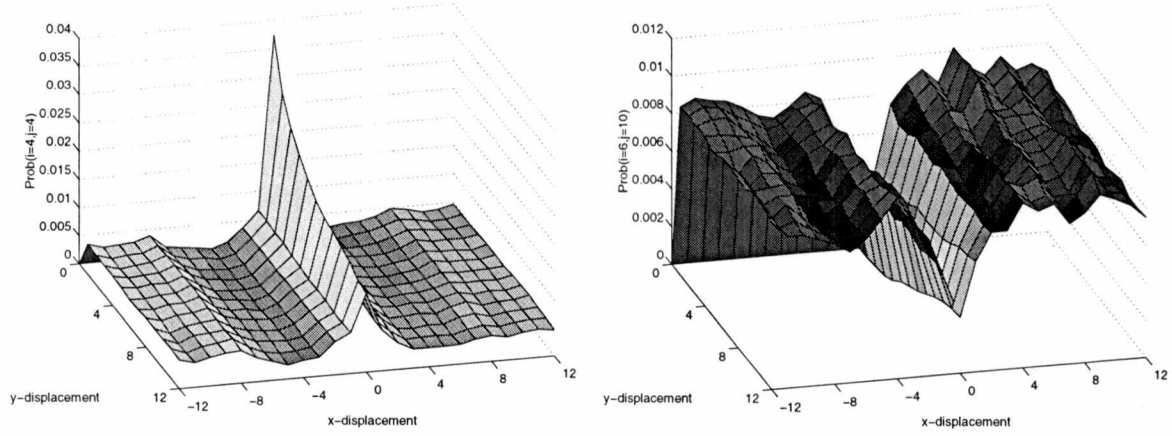


(d) Displacement [0 12]

Figure 3.8: The GLC matrices extracted from the wood texture image for four selected displacement vectors: (a) and (b) small and large horizontal displacements, (c) and (d) small and large vertical displacements.

texture GLC matrices for the vertical displacements are much sharper than for the corresponding vertical displacements because the wood texture has mostly vertical structure. Notice also that for both the horizontal and the vertical directions, the GLC matrices for the small displacements have sharper main diagonals than for the large displacements. This will be true for most textures, as a change in gray level is more likely if the pixel with gray level i lies on one texture tile while the pixel with gray level j lies on another. Notice also from these plots that the Gaussian shaped histogram is evident if the function is projected onto either gray level axis. An estimate of the histogram can be obtained from any GLC matrix by either summing all the rows or all the columns.

We can also get an idea of the structure of the texture pattern by fixing the gray level indices i and j and observing the relative probabilities indexed spatially. For instance, in Figure 3.9(a) the cooccurrence probabilities for only the gray level transition $i = 4, j = 4$ are plotted with the $[x \quad y]$ spatial displacement coordinates. In this plot, the coordinate definitions are as shown earlier in Figure 2.2 (Section 2.1.2). Note that indexing the GLC matrices in this manner does not constitute a proper 2-dimensional probability distribution, as we have only plotted relative values sampled from the same point in a number of probability distributions. Since the transition $i = 4, j = 4$ is one of the elements along the main diagonal for each GLC matrix, we should see high relative probabilities for displacements where the pixel with gray level i and the pixel with gray level j are part of the same texture tile or lie on two similar tiles. The main spine of the surface extends down the vertical (y) axis, slowly tapering off. This is an indication of the typical tile shape for darker-shaded



(a) Gray level transition $i = 4, j = 4$ (b) Gray level transition $i = 6, j = 10$

Figure 3.9: Spatial indexing for the GLC matrices extracted from the wood texture image: (a) for gray level transition $i = 4, j = 4$, (b) for gray level transition $i = 6, j = 10$.

tiles. Also notice that the surface dips low moving both in the positive and negative x-direction before rising somewhat at x-displacements of about 8 pixels. This should be an indication of about how far apart the darker-shaded tiles are typically located. Figure 3.9(b) shows the same type of plot but for gray level transition $i = 6, j = 10$. Since this transition is located off the main diagonal for each GLC matrix, we should see high relative probabilities for displacements where the pixel with gray level i and the pixel with gray level j are not part of the same texture tile. We see that the lowest points of the surface are along the vertical axis, slightly increasing with larger vertical displacements. This again is indicative of the vertical structure of the texture. Moving horizontally, the surface rises significantly and exhibits some rippling. These waves and ripples also indicate the vertical structure

of the texture.

3.3 Importance of GLC Ridge Sharpness

We have seen that for the GLC matrices of the textures considered, the diagonal components dominate the non-diagonal components. This is especially true for the GLC-matrices associated with short displacement vectors, corresponding to the formation of single-intensity "blobs" or "tiles" of various shapes, sizes, and orientations. Additionally, the GLC-matrices associated with displacement vectors that are long enough to extend between blobs have a shape that is flatter, without a distinct "ridge" along its diagonal. Experiments in which textures are synthesized from artificially created GLC models instead of ones that have been computed from an existing pattern can be used to investigate this phenomenon. Figure 3.10 demonstrates such an experiment. The textures shown are the results of the spin-flip synthesis algorithm for five different GLC-based texture models. In each case, the GLC-model considered displacement vectors of length no greater than 2 pixels in either the x- or y-component (i.e., $nx = ny = 2$ pixels). Since only three gray levels were used, each of the 12 GLC matrices was of size 3×3. Another constraint for each synthesized texture was that they be histogram equalized before application of the iterative process to minimize GLC model error, i.e., each pixel has a $\frac{1}{3}$ chance of being any one of the three gray-levels.

In Figure 3.10(a), each GLC-matrix in the model had its diagonal elements equal to its non-diagonal elements ($a = b$). This means the shape of each matrix is flat, with no ridge along its diagonal. Thus, the probabilities of intensity change at

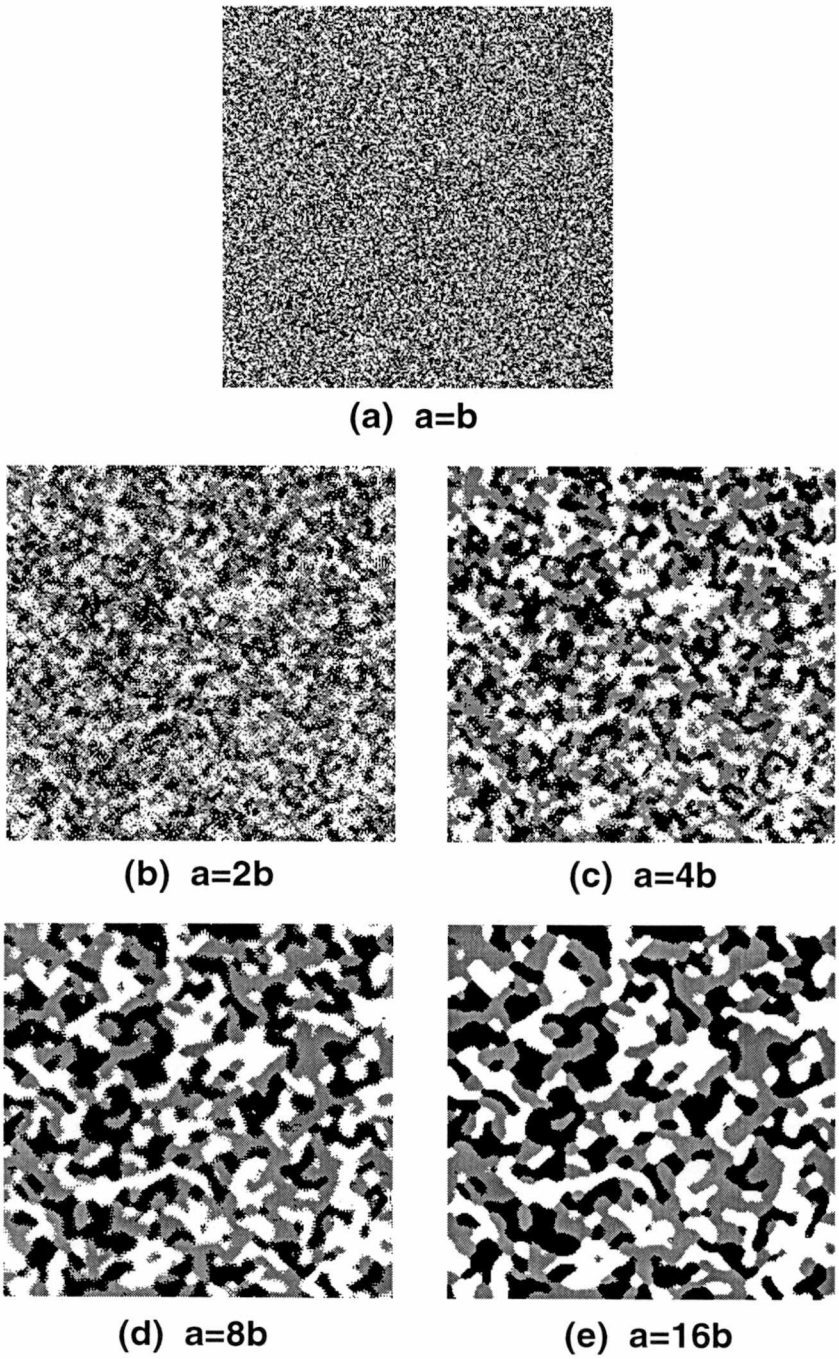


Figure 3.10: Demonstration of how GLC ridge sharpness affects the resulting synthesized texture. In (a) the diagonal elements of each of the 12 GLC's is equal to the non-diagonal elements. In (e) the diagonal elements are 16 times greater in magnitude than the non-diagonal elements.

every distance are always equal, so the result is a purely random 3-gray level image of uniform distribution. In parts (b) through (e), the magnitude of the diagonal elements was increased to $a = 2b$, $a = 4b$, $a = 8b$, and $a = 16b$, respectively. In each case, **all 12** GLC matrices in the model were the same. Notice that as the diagonal ridge is made sharper, the resulting texture blobs become more defined and to some extent larger. This result raises the question as to whether certain elements of the matrices, such as the diagonal ones, are more important in the description of a particular texture than the other matrix elements.

3.4 Importance of GLC Displacement Vector

Other experiments suggest that individual displacement vectors in the texture model have an important influence in the generated texture. Figure 3.11 shows some results of an experiment to demonstrate this. In (b) is shown the original texture with no modification. It was generated using 12 identical matrices as before, and is the same image as in Figure 3.10(e), where every diagonal element is 16 times larger than the every non-diagonal element ($a = 16b$). One at a time, the GLC matrix corresponding to each displacement vector considered was swapped with a different matrix, and the synthesis procedure was performed. The replacement matrix was the one used before in Figure 3.10(a), where the diagonal elements are equal to the non-diagonal elements ($a = b$). Thus we are changing the high-ridge matrix in the texture model to an all-flat matrix, and seeing what effect it has on the generated texture. Notice how the texture blobs in (a) tend to be elongated slanting up and to the left. This direction is exactly perpendicular to the displacement vector (d=[-2,2])

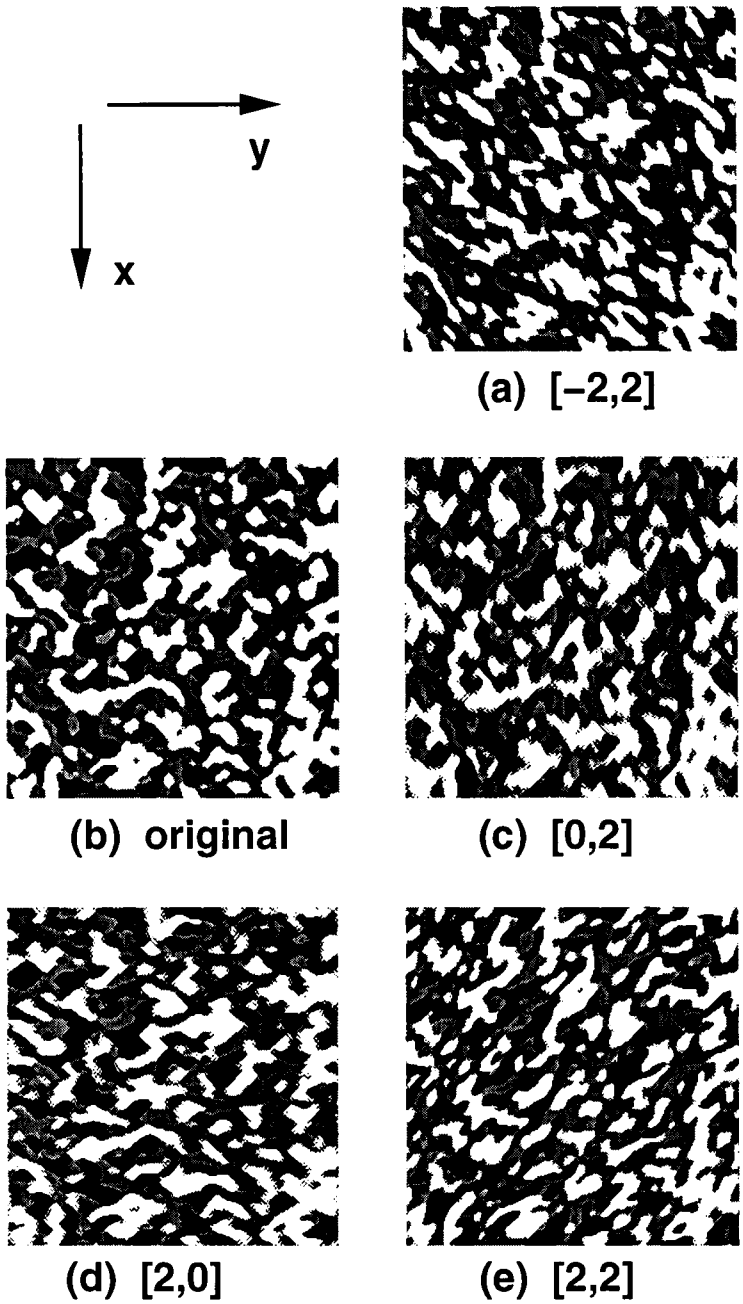


Figure 3.11: Demonstration of how flattening the shape of one GLC matrix in the model can affect the resulting synthesized texture.

corresponding to the GLC-matrix that was swapped. This result seems intuitive, if you consider that replacing a ridge matrix with a flat matrix is increasing the probability of intensity changes in that direction. This is equivalent to increasing the amount of AC energy oriented in that direction. Similarly, the texture blobs in (e) tend to be elongated up and to the right, perpendicular to displacement vector d=[2,2]. This same result is a little more difficult to see in (c) and (d), most likely because the displacement vectors d=[0,2] and d=[2,0] are not as long in the Euclidean sense as d=[-2,2] and d=[2,2]. However, the same texture blobs that appear in the original in (b) tend to stretch more vertically in (c) and horizontally in (d). This result raises the question as to whether certain GLC-matrices are more important in the description of a particular texture than the other GLC-matrices.

3.5 Gaussian Shaped GLC Models

In this section, we consider once again textures synthesized from artificially created GLC matrices, but with Gaussian shaped distributions. A bivariate normal distribution, given in Equation 3.3, can be completely determined by 5 parameters, which consist of the 2 mean parameters and the 3 independent elements in the 2x2 covariance matrix [DH73].

$$p(\mathbf{x}) = \frac{1}{2\pi|\mathbf{C}|^{1/2}} exp[-\frac{1}{2}(\mathbf{x}-\mathbf{m})'\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})] \tag{3.3}$$

In Equation 3.3, $\mathbf{x}$ is the 2x1 coordinate vector, $\mathbf{m}$ is the 2x1 mean vector, and $\mathbf{C}$ is the 2x2 covariance matrix. It was noted previously that most texture patterns with a certain amount of tile structure have GLC matrices with diagonal and near-

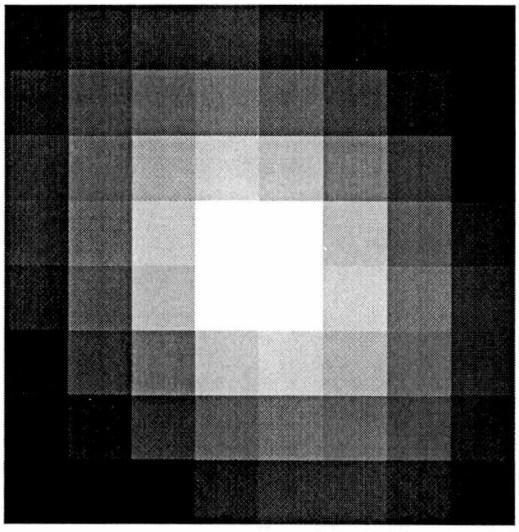
diagonal elements that are larger than the elements farther from the diagonal. By choosing proper covariance matrices, we can generate bivariate normal distributions with this characteristic.

In Figure 3.12 through Figure 3.14, several 8x8 bivariate normal distributions with various covariance matrices are plotted in (a) and in (b). The 2x1 mean vector has been specified so that the distribution is centered in the GLC matrix. Parts (c) and (d) are the best results of texture synthesis when all of the GLC matrices in the input model are of the bivariate normal distributions shown in (a) and (b). In each case, 8 gray levels are used and displacements of up to 8 pixels in the horizontal vertical directions are considered, resulting in a total of 144 GLC matrices in each model. These few examples of bivariate normal GLC models demonstrate the possibility of using synthetic GLC models with controllable parameters, instead of models computed from actual textures, to synthesize patterns to be used as perceptual experiment stimuli.

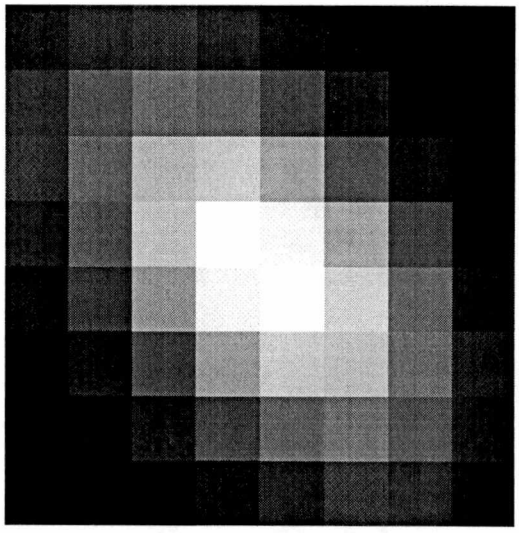
3.6 Nature of GLC Error in Synthesized Texture

3.6.1 Visualizing GLC Synthesis Error

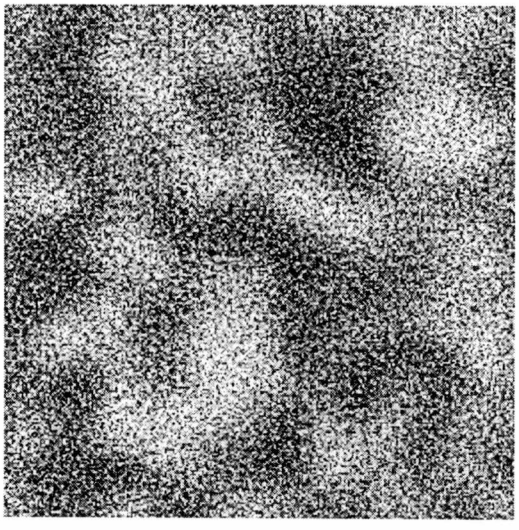
In Section 3.1, the spin-exchange method of texture synthesis was discussed. The average cooccurrence error (ACE), which was defined previously in Equation 3.2, is simply the cooccurrence error (CE) averaged over all displacements considered, and is used to gauge how different two GLC texture models are. The maximum CE for any GLC matrix is 2.0 (if the two probability distributions have zero overlap) and the minimum is 0.0 (if the two are identical). Figure 3.15 shows three 256×256



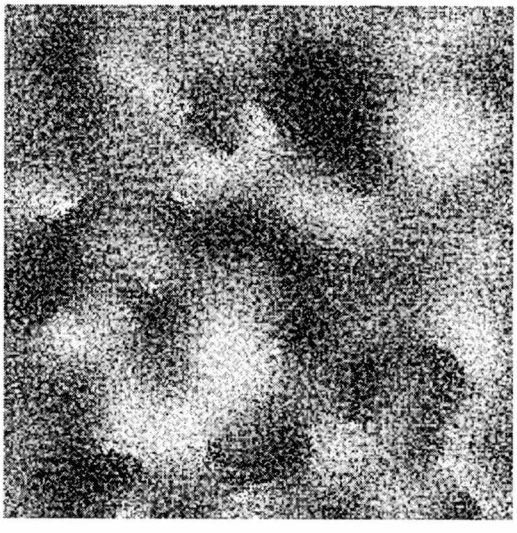
(a) Normal GLC [4,1 ; 1,4]



(b) Normal GLC [4,2 ; 2,4]

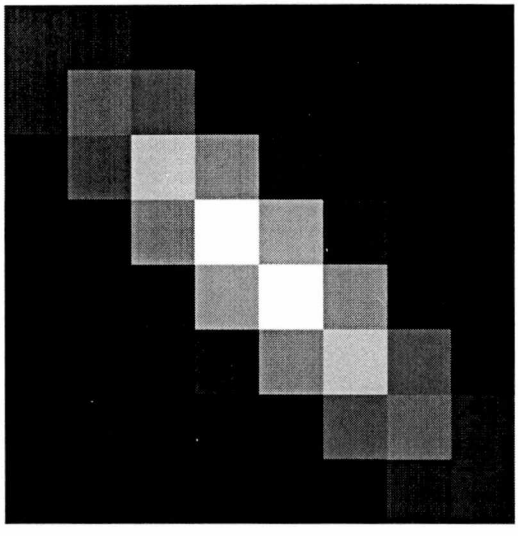


(c) Synthesis from (a)

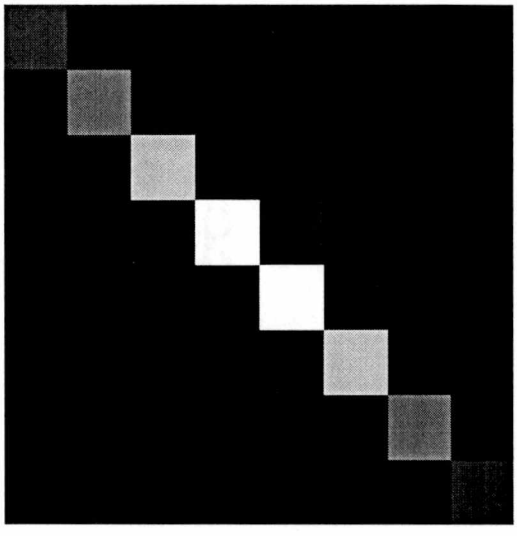


(d) Synthesis from (b)

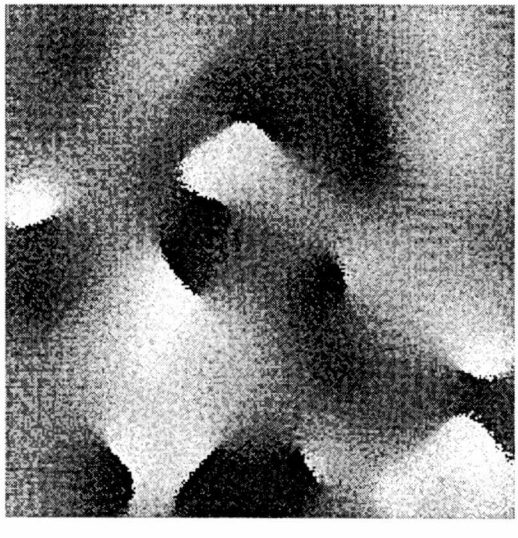
Figure 3.12: (a) and (b) Two normal GLC matrices with different covariance matrices. (c) Best synthesis result using (a). (d) Best synthesis result using (b).



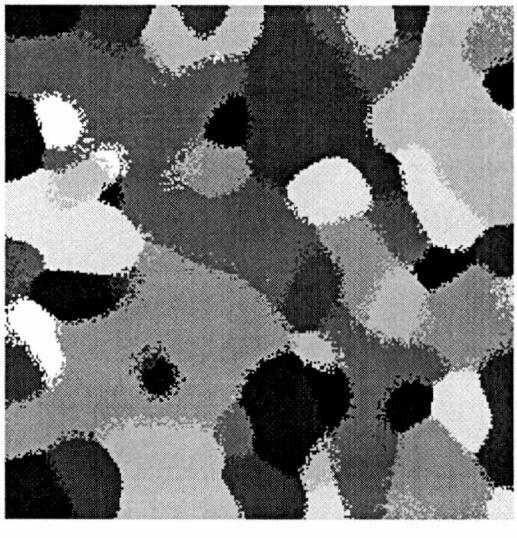
(a) Normal GLC [4,3.5 ; 3.5,4]



(b) Normal GLC [4,3.9 ; 3.9,4]

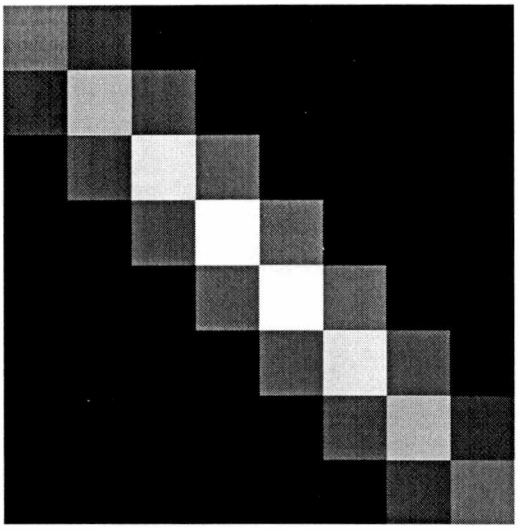


(c) Synthesis from (a)

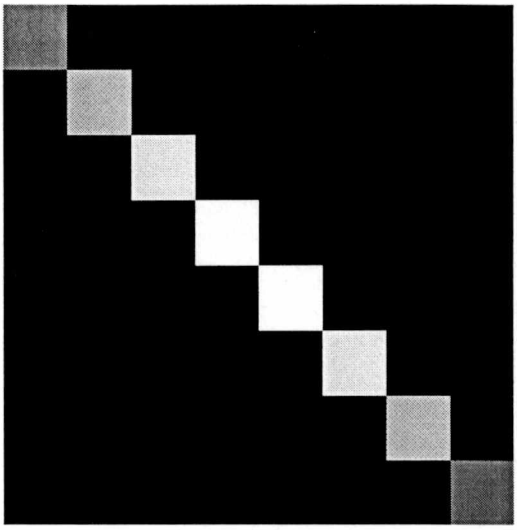


(d) Synthesis from (b)

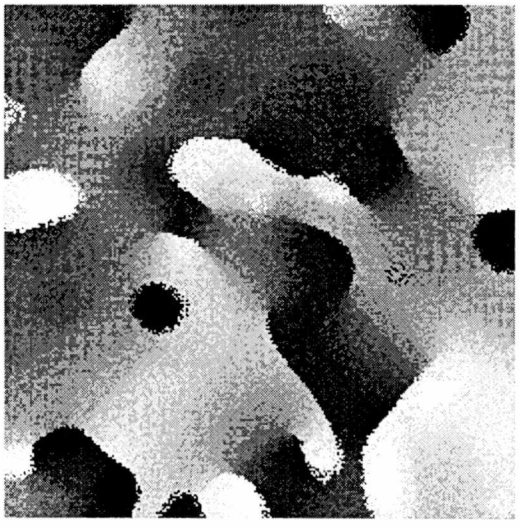
Figure 3.13: (a) and (b) Two normal GLC matrices with different covariance matrices. (c) Best synthesis result using (a). (d) Best synthesis result using (b).



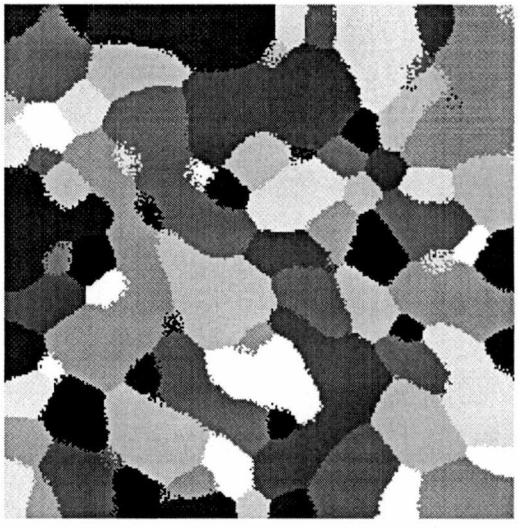
Normal GLC [8,7.7 ; 7.7,8]



Normal GLC [8,7.95 ; 7.95,8]

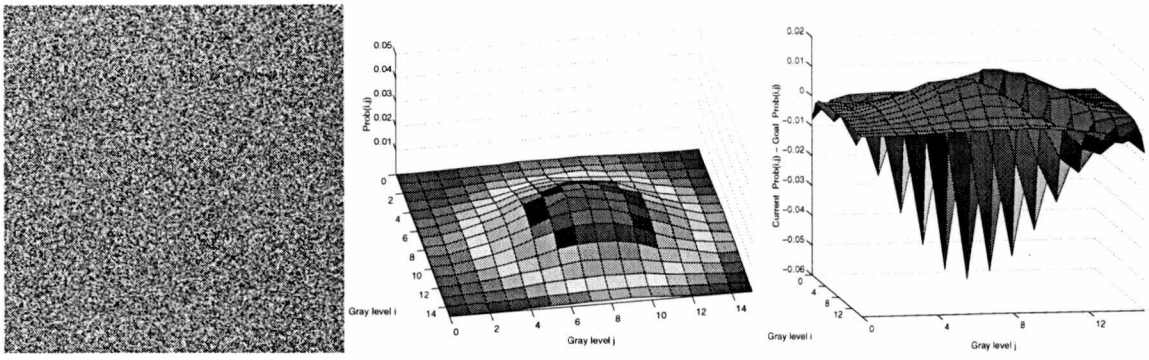


(c) Synthesis from (a)

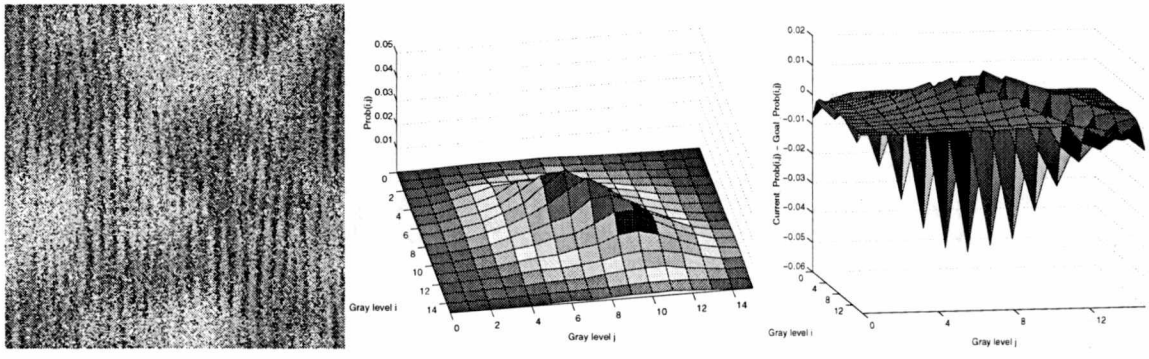


(d) Synthesis from (b)

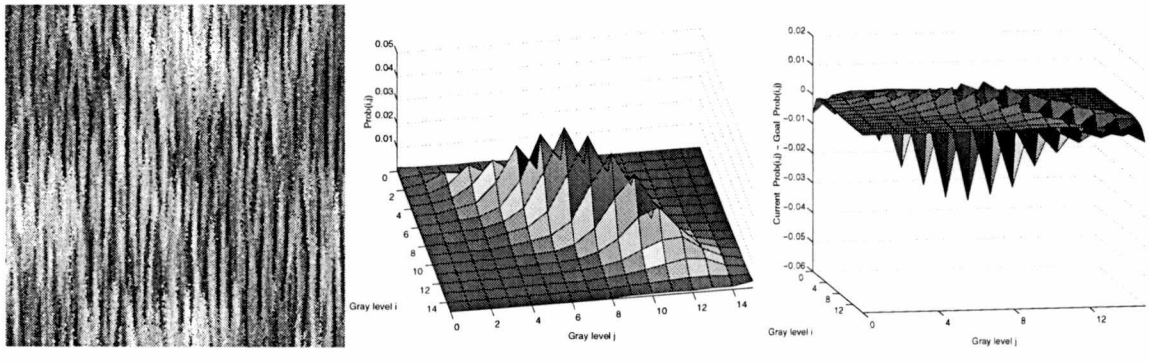
Figure 3.14: (a) and (b) Two normal GLC matrices with different covariance matrices. (c) Best synthesis result using (a). (d) Best synthesis result using (b).



Initial image GLC [0 1] GLC error [0 1]



Intermediate image GLC [0 1] GLC error [0 1]



Final image GLC [0 1] GLC error [0 1]

Figure 3.15: The synthesis process, from the initial random distribution (top row) to the best result (bottom row). The first column shows the image at each stage, the second column shows the GLC matrix for displacement [0 1] at each stage, and the third column shows the corresponding GLC error matrix (current GLC minus goal GLC).

gray scale images during the synthesis process for the GLC matrices computed from the wood grain texture, shown originally in Figure 3.7. Also shown for each image is its GLC matrix for the [0 1] (vertical neighbor) displacement and the corresponding GLC error matrix, which is defined as the current GLC matrix minus the GLC matrix from the goal model. As explained before, the initial image is specified to the desired histogram but the gray level values at any two pixels are independent. Because of this independence, we know that that $P(i,j|d) = P(i) \times P(j)$ for any displacement d, where $P(i)$ and $P(j)$ are the histogram probabilities of gray levels i and j. If we define H to be the histogram probability vector, then the corresponding GLC matrix will be $H^T H$. Since our specified histogram distribution was Gaussian, we see the 2-dimensional Gaussian shape in the [0 1] GLC matrix surface. Note that if we had initially chosen a uniform histogram, this GLC matrix surface would be flat. Since the goal [0 1] GLC matrix for the wood texture has a very steep main diagonal (Figure 3.8(c)), we see from the GLC error plot that our current [0 1] GLC matrix has far too little energy along its main diagonal and too much energy in the center but off the main diagonal. For this initial image, the GLC error metric is approximately $ACE = 0.28$.

The second row of Figure 3.15 shows an image at one particular point in the synthesis process. At this intermediate point, 46,281 gray level exchanges have occurred, and the GLC error metric is $ACE = 0.12$. Note that this intermediate image is beginning to demonstrate both the narrow, vertical striping as well as the formation of some larger, shapeless tiles that are characteristic of the wood texture. From the [0 1] GLC matrix plot for this image, we can see that the gray

level exchanges are causing a sharpening of the main diagonal as the vertical tiling structure is formed. From the GLC error plot, we see that the elements along the main diagonal still have much smaller values than we desire.

The third row of Figure 3.15 shows the synthesized image at the point where the GLC error metric is $ACE = 0.05$, and 93,814 gray level exchanges have occurred. At this point in the synthesis process, the image has reached a state where a texture pattern has formed and there are no further perceptible changes in its visual appearance. Since the GLC error metric has not reached zero, we presume that either the pattern that has formed has constrained the process so that the GLC error metric is caught in a local minimum, or that the random method of choosing pairs of pixels is beginning to have difficulty finding pairs that will lower the GLC error metric. From the [0 1] GLC matrix plot for this final image, we can see that the main diagonal has continued to sharpen. However, it is still not as sharp as the main diagonal of the [0 1] GLC matrix for the wood texture, as evidenced in the GLC error plot.

3.6.2 Analysis of GLC Synthesis Error

The pattern to which the synthesis process converged possessed a GLC error metric of about $ACE = 0.05$. If we could obtain a pattern with GLC error closer to zero, we would probably have a more satisfactory result. More knowledge about the nature and structure of the GLC error in the final result could help define a method for improved texture synthesis.

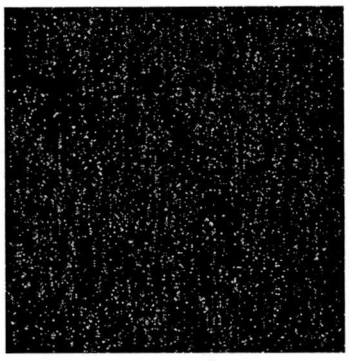
The error for only the [0 1] GLC matrix of the final synthesized image was

already plotted in Figure 3.15. But which pixels in the image are contributing most to the total error? To answer this question, we can define the GLC error at a given pixel. For any given displacement $d = [x \quad y]$, the gray level of a pixel m at location (m_x, m_y) contributes to two elements of the GLC matrix. For one element, it determines the row and the gray level of the pixel at $(m_x + x, m_y + y)$ determines the column. For the other element, it determines the column and the gray level of the pixel at $(m_x - x, m_y - y)$ determines the row. If i is the gray level of pixel m, j is the gray level of the pixel at $(m_x + x, m_y + y)$, and k is the gray level at $(m_x - x, m_y - y)$, then we can say that pixel m contributes to error:

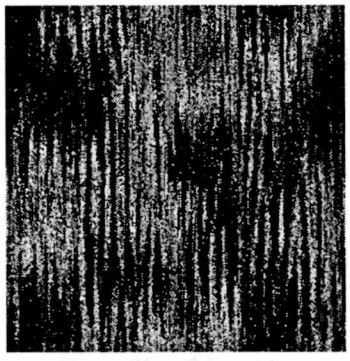
$$e_p(d) = (P_s(i,j|d) - P_g(i,j|d)) + (P_s(k,i|d) - P_g(k,i|d)), \tag{3.4}$$

where P_s is the GLC matrix for the synthesized texture and P_g is for the goal texture model. We can see from Equation 3.4 that if the GLC matrix probability for the goal texture is higher than the corresponding probability for the synthesized texture, then the pixel is rewarded for providing a gray level transition that is in shortage by having its error reduced by the amount of the difference.

If we sum the value of $e_p(d)$ for each pixel over several displacements, we can get some indication of which pixels are contributing the most to GLC error. In Figure 3.16(a), we show the result for small displacements $[x \quad y]$, where x and y are both 6 or less. In this image, if the error for a pixel is less than zero it is shown as black, and all positive values of pixel error are normalized and quantized to a gray level from 0 to 255. In (b) the displacements considered are large ones, where either x or y is 7 or larger. From these two images, we can see that the pixel error for small displacements seems to correspond to higher-frequency characteristics, where



(a) Small displacements



(b) Large displacements

Figure 3.16: Contribution to GLC error of pixels in the final synthesized image: (a) considering only displacements [x y] where $1 \leq (x \quad \text{and} \quad y) \leq 6$, (b) considering only displacements [x y] where $7 \leq (x \quad \text{or} \quad y) \leq 12$.

individual pixels don't belong well on a particular texture tile. The pixel error for large displacements corresponds more to lower-frequency characteristics, such as how the tiles are arranged with respect to one another.

In Figure 3.17, we have used spatial indexing to examine the structure of the GLC error in the synthesized image. For each displacement, we have plotted the value of $CE(d)$ (Equation 3.1). This shows us that the GLC error is concentrated in the vertical displacements, especially for short distances, and rather uniform elsewhere. It seems as if the synthesis process had difficulty in forming the vertical tile structure that was desired.

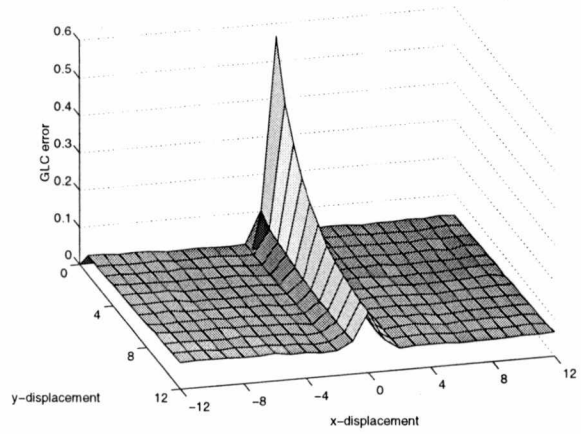


Figure 3.17: Spatial indexing for the GLC error in the final synthesized image. Shown, for each displacement, is the sum of absolute error between the model for the synthesized image and the goal model.

3.7 Conclusions

From our work with synthesizing a wood texture pattern, we can draw several conclusions. It is evident that GLC matrices, which are essentially a 4-dimensional feature space, offer a great deal of information about a texture pattern. We can see this visually by considering only one displacement and plotting the 2-dimensional GLC matrix for that displacement or by considering only one gray level transition and plotting that element over all 2-dimensional spatial displacements. With this spatial indexing, we can obtain much structural information about the texture pattern, such as tile size, shape, or orientation. Although GLC matrices are normally considered a statistical method of texture analysis, we see that they can also be

used as a method of structural analysis [CN89, HTCP86].

With the importance of the information in the spatial dimensions of GLC matrices, it would be detrimental to ignore or obscure this information. For instance, it may be unwise to average or otherwise combine information from different displacements. Shen and Bie [SB92] and Wu and Chen [WC92] compute scalar features from GLC matrices and then average their values over displacements of several directions. They use these averaged values to compare the performance of GLC matrices to other textural analysis methods. This will certainly obscure the directional information content of the GLC matrices, and we know that directionality is a very important textural characteristic. Using the same argument, we can also see that GLC matrices generalize some other texture analysis techniques. For instance, it has been observed that aura matrices (and, similarly, Gibbs/Markov random field models) are simply sums of GLC matrices, taken over the displacements contained in the neighborhood structuring element [EP94]. Comparing Figure 2.2 to Figure 2.1, which showed a typical spatial neighborhood for Gibbs/Markov models of order 1-4, we see that GLC matrices use generalized spatial information while a Gibbs model unnecessarily combines the information over several displacements.

As far as the Metropolis spin-exchange method of synthesizing a texture with specific cooccurrence probabilities, we have seen that while the method can change a random pixel distribution into one resembling the desired texture, the process converges to a pattern for which the GLC error does not reach zero. This may be because once the tiles form we have a pattern from which it is not possible to reach zero error, i.e., the process has become stuck in a local minimum. The use of

temperature annealing [PEP91] could allow the process to occasionally escape from such a local minimum. However, it is clear that the random method of choosing pairs of pixels in the spin-exchange algorithm becomes less effective as the pattern develops structure. Perhaps some knowledge of how much a pixel contributes to the total error could be used to develop a more efficient method of selecting pixels for possible gray level exchange, which could result in more satisfying synthesis results.

CHAPTER 4

Computer Testbeds and Experimental Procedures for Studying Human Search and Discrimination

4.1 XPET – The X-Based Perceptual Experiment Testbed

4.1.1 Overview of XPET

A testbed was developed to perform experiments in human texture discrimination using a Silicon Graphics workstation running Unix and X-windows with a color monitor. The software that was developed to handle the sequential display of the image stimulus pairs and the registration and recording of subject responses is the X-based Perceptual Experiment Testbed (XPET) [CTM96]. XPET features a graphical user interface which allows the user to choose between set-up mode and experiment mode. In set-up mode, the experimenter can specify the order and location in which the image stimuli will be presented to the observer. In experiment mode, the observer makes his responses to the stimulus pairs by simply using the mouse to point and click on his selections. When the experiment is finished, XPET saves the observer's responses to a file. The responses of all observers are combined to form one file of composite data, from which it is possible to estimate the appropriate scale values for the stimuli. Figure 4.1 shows an example of the XPET display in set-up mode. The presentation order list in the lower-left corner allows the experimenter to verify that

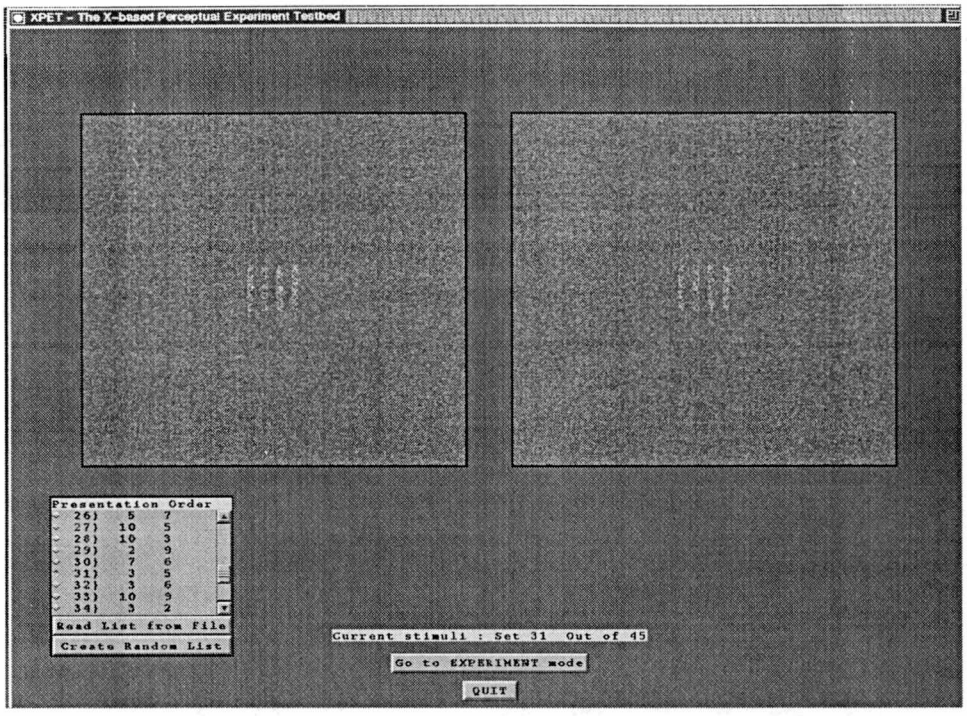


Figure 4.1: An example of the XPET display in set-up mode.

the experiment is configured to his liking, by stepping back and forth through the list and checking that the proper image stimuli are being displayed. In experiment mode, this display of the presentation order is not visible. Figure 4.2 shows a test subject using XPET in experiment mode to register his responses.

Using XPET to automate the display of image stimuli and collection of response data has three key advantages over more conventional methods which have been used [VE84]. First, it allows the experimenter to maintain constant environmental conditions for every observer. This is necessary to assure the scientific merit of the experiment. Second, XPET is flexible enough to allow a change in image stimuli or presentation order quickly and easily. Thus, a minimal amount of time is required to try various experiment configurations. Third, XPET's graphical interface is user-friendly so that a minimum of explanation is required for each human observer. This

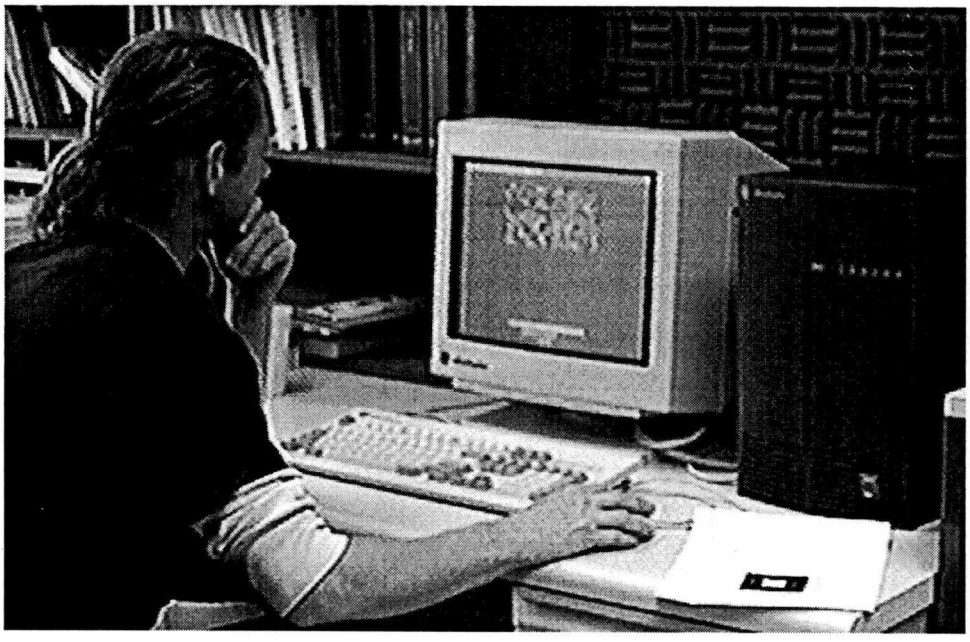


Figure 4.2: A test subject using XPET to register his responses.

simplicity also serves to minimize errors during data collection.

4.1.2 Conducting an Experiment with XPET

Before starting an XPET session, the experimenter must previously have created three files. These three files MUST be named 'XPET_images', 'XPET_order', and 'XPET_default.pgm', and MUST exist within the directory in which the xpet session is to be conducted. The file 'XPET_images' must contain a list of the filenames of the stored images to be used as perceptual stimuli. Each of these stored images must be in portable gray-map (pgm) format.

The file 'XPET_order' must contain groups of integers describing each of the sets of stimuli to be presented. Each integer refers to one of the image filenames provided in the 'XPET_images' file. For example, if the 'XPET_images' file looks

like this:

image_a.pgm

image_b.pgm

image_c.pgm

image_d.pgm

then the integer '1' will be associated with the image 'image_a.pgm', the integer '2' with the image 'image_b.pgm', and so on. Thus, if the 'XPET_order' file looks like this:

1 2

3 4

1 4

3 2

then the first stimulus set will be image_a compared to image_b, the second stimulus set will be image_c compared to image_d, and so on, for a total of four stimulus sets altogether.

'XPET_default.pgm' will be the image that is displayed in place of stimuli in experiment mode during the time between the observer's response and the display of the next stimulus set. This brief pause is necessary to give XPET time to read in the next two stimulus images.

When all three of the required files have been created in the experiment directory, the XPET session can be initiated by typing: **xpet (ENTER)**. XPET always begins in set-up mode. Figure 4.1 shows an example of the XPET display in set-up mode. In set-up mode, the experimenter can position the images as he wishes

and verify that the presentation order is acceptable. To position the images, press the left mouse button over one of the images and drag. The two images will stay aligned and centered. In addition, the display size of the images can be changed. Double-clicking the LEFT mouse button over one of the images will INCREASE the size of the displayed images, while double-clicking the RIGHT mouse button will DECREASE the size of the displayed images.

The presentation order is initially the one that was loaded from the file "XPET_order." This presentation order appears in a scrollable list object. This list object can be moved by pressing the left mouse button over the greenish title bar which reads "Presentation Order," and dragging. It is possible to either load another order from a file, or create a new randomized order. To load a new order from a file, press the "Read List from File" button at the bottom of the Presentation Order object. A file browser will pop up so the experimenter can load a different order file. To create a new randomized order, press the "Create Random List" button. The Experimenter will be prompted for a seed number for the random number generator. The newly created random order will be created based upon the number of stimuli being used, and will be suitable for a paired-comparison experiment.

After an acceptable presentation order is established, the experimenter can scroll up and down through the list, selecting various stimulus sets. When he presses the left mouse button on a particular stimulus set in the list, the images will change to the appropriate ones. Also, the greenish information display near the bottom center of the XPET window will indicate which set is showing. In this manner, the experimenter can verify that the images are updating as desired.

When the experimenter is ready to collect data from a test subject, he presses the "Go to EXPERIMENT mode" button. The presentation order list will disappear, the images will freeze in position, and the first stimulus set will be presented. The first human test subject now begins providing responses to the stimuli. For each pair of images, he merely clicks a mouse button over the image he selects, according to the perceptual quality the experimenter wishes him to judge. The image that was selected will be highlighted with a red square and both images will change to 'XPET_default.pgm' until the next set of images are displayed. When the test subject makes his selection for the last set of stimuli, XPET will indicate that the results have been saved in a file called "XPET_result_1," and revert back to set-up mode. When the next test subject is ready, the experimenter presses "Go to EXPERIMENT mode" again. The results from the second trial will be stored in "XPET_result_2," and so on.

The raw data is output in columnar format with each row containing the data for one response. The first column is the index number for that stimulus pair. The second column contains a '1' if the observer selected the image displayed on the left and a '2' if the right image was chosen. The third column contains the approximate time, in seconds, required for the observer to make his selection. This is measured from the moment that the images were displayed to the moment the observer clicks the mouse button over one of the images. The fourth column contains the (x,y) location of the mouse click within the selected image. These coordinates are normalized so that x varies from 0.0 to 1.0 left to right, and y varies from 0.0 to 1.0 bottom to top.

The typical instructions that are given to each test subject before a texture discrimination experiment are as follows:

> During this perceptual experiment, you will be presented with a set of images. Each image consists of a target pattern located in the center of a background pattern. The images will be presented in pairs, with one image on the left and another image on the right. As you are presented with each pair of images, you are to choose the image in which you think the target pattern is more DISTINCT from its background. To register your response, simply click a mouse button over the image you choose. You must choose one of the two images. In some cases the decision may be a difficult one. If so, just make your best guess. There are no right or wrong answers, as every person may respond differently. The amount of time you require to make each decision will be analyzed, so please use only enough time to look at the two images and be reasonably sure of your answer. You may position yourself however that you are comfortable and can see the images clearly. When you have registered your final response, XPET will state that your results have been saved to file. At that time, please leave the XPET program running and you are free to leave. Thank you for your participation in this perceptual experiment!

Note that no specific instructions are typically given as to what perceptual cues the test subjects were to use when deciding which target pattern was more distinct.

It is important to note that the Silicon Graphics display should be linearized

before an experiment is conducted. This is accomplished by setting the Unix gamma parameter to 1.0, which establishes a linear relationship between pixel intensity and display luminance. This is necessary to validate the analysis which follows to relate statistical measures based on pixel intensities to human perception of the visual appearance of the images on a computer display. Other test parameters (display brightness, contrast, etc.) are set appropriately and maintained constant for all observers.

4.1.3 Comparison Experiments to Investigate Human Texture Discrimination

Two experimental methods for studying human texture discrimination are possible using XPET. In the method of paired comparisons, each stimulus is paired with every other stimulus exactly once [Tor58, Ges84]. For each paired comparison, the human observer is required to choose which of the two stimuli is greater in the attribute being judged. With image stimuli that consist of target texture patterns against background texture patterns, the attribute being judged is "target distinctness." The raw data is used to estimate scale values for the stimuli, as explained in the next section.

The second type of experiment implements the complete method of triads, to investigate the perceptual distance from each of a set of textures to each of the others. In the complete method of triads, the stimuli are presented to the observer in sets of three. Every possible combination of three stimuli is presented to the observer three times, with a different one of the three used as the standard each

time. For example, consider a particular triad composed of stimuli A, B, and C. The first time the triad is presented, the observer is is required to judge whether A is more similar to B or to C. The second time, he judges whether B is more similar to A or to C. The final time, he judges whether C is more like A or B. Using data from such an experiment, it is possible to derive the comparative distances between all pairs of stimuli. This process is explained using the data from an actual triadic comparison experiment later in Section 7.2.3.

It is beneficial to consider the the relationship between the complete method of triads and the method of paired comparisons. In a paired comparison experiment, we are essentially investigating the perceptual distance between each of the patterns from a **common** point. This is accomplished by designating one pattern to serve as either the common target or the common background for all of the image stimuli. Using the more exhaustive complete method of triads, we can study the more fundamental issue of the perceptual distance from **each** of a set of "prototype" textures to **each** of the others. XPET is also suited for triadic comparisons of textures. We can do this by still presenting images two at a time, if each of the images consists of a target pattern of one texture against a background pattern that is a different texture. For example, we can consider the question of whether texture A is more similar to B or to C by comparing an image that is texture A on texture B against an image that is texture A on texture C. The observer must only decide which of the two images has a target pattern that is more distinct.

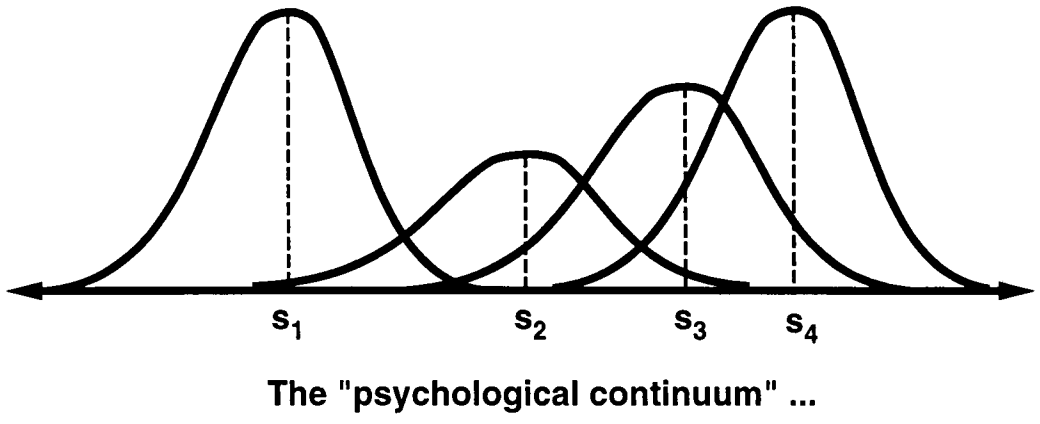


Figure 4.3: The distributions of discriminal processes for four possible stimuli.

Psychological Scaling and the Law of Comparative Judgment (LCJ)

Psychological scaling is any procedure used to assign relative scale values to a set of stimuli based upon human subjects' responses. The law of comparative judgment (LCJ), introduced by Thurstone [Thu27], is a set of equations relating the proportion of times any given stimulus k is judged greater on a given attribute than some other stimulus j to a scale ranking on the "psychological continuum" [Tor58, Fal86]. The LCJ is based on the postulate that if a stimulus is presented to a human subject, it excites a discriminal process which has some value on the psychological continuum. It is also assumed that this value will not be exactly the same each time the same stimulus is presented, but rather these values will form a normal distribution on the psychological continuum. Figure 4.3 shows the distributions of discriminal processes for four possible stimuli. The mean and standard deviation of each normal distribution are taken as the scale value and discriminal dispersion associated with that stimulus. Letting s_j and s_k represent the scale values of two

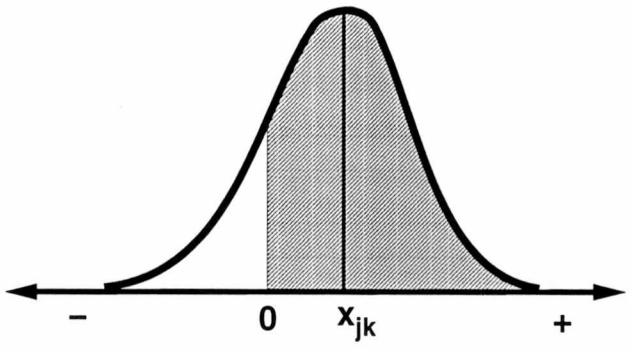


Figure 4.4: A possible distribution of discriminal differences $(d_k - d_j)$, where the shaded area corresponds to the proportion p_{jk}.

stimuli and σ_j and σ_k represent their discriminal dispersions, the law gives us:

$$s_k - s_j = x_{jk}(\sigma_j^2 + \sigma_k^2 - 2r_{jk}\sigma_j\sigma_k)^{\frac{1}{2}}, \tag{4.1}$$

where r_{jk} is the correlation between the pair of normal discriminal processes d_j and d_k. Equation 4.1 describes the theoretical distance between the scale values of the two stimuli, and is obtained by recognizing that the difference of two normal distributions is also a normal distribution with variance $\sigma_j^2 + \sigma_k^2 - 2r_{jk}\sigma_j\sigma_k$. The term x_{jk} is the unit normal deviate corresponding to p_{jk}, the theoretical proportion of times stimulus k is judged greater than stimulus j:

$$p_{jk} = \int_{-\infty}^{x_{jk}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx. \tag{4.2}$$

Figure 4.4 shows a possible distribution of discriminal differences $(d_k - d_j)$, where the shaded area corresponds to the proportion p_{jk}. The appropriate x_{jk} can be obtained for each p_{jk} from a table of areas under the unit normal curve or by numerical integration.

Hence the law forms a system of equations, one for every pair of stimuli. Equation 4.1 is the general form of the LCJ, but it is not solvable in this form, since there are always more unknown quantities than observation equations. It is therefore necessary to make additional assumptions about the statistical properties of the stimuli to reduce the law to a solvable form. One way to do this is to assume that the standard deviation of the distribution of discriminal differences is constant for all pairs of stimuli. This assumption is usually referred to as simplifying condition C for the LCJ. Denoting the constant standard deviation by c, Equation 4.1 reduces to:

$$s_k - s_j = cx_{jk} \tag{4.3}$$

The traditional method for solving condition C of the LCJ is given by Torgerson [Tor58]. In this method, it is required that each stimulus has been compared to each other stimulus numerous times. The resulting data will be the proportion of times a stimulus is judged greater than each of the other stimuli, in the quality or attribute being studied. For each proportion p_{jk}, the corresponding unit normal deviate x_{jk} is found. If the unit of measurement is chosen so that the constant c in Equation 4.3 is equal to unity, then the unit normal deviate x_{jk} is an estimate of the perceptual "distance" between stimuli j and k. A problem arises when the proportion p_{jk} is equal to 0.0 or 1.0, meaning that every observer gave the same response when presented with stimuli j and k. In this case, referred to as incomplete data, the corresponding unit normal deviate will be infinite, since a normal distribution never reaches zero in either direction. The traditional method disregards such infinite elements and estimates the distances only from the finite (known) elements.

Equation 4.3 is a solvable form of the LCJ. But the raw data from the experimental procedure must still be converted into a matrix format suitable for application to the LCJ. Our raw data consists of the number of times each stimulus was judged greater than every other stimulus. These numbers can be arranged into an $n \times n$ square matrix $\mathbf{F}$, where n is the total number of stimuli and each element f'_{jk} denotes the observed number of times stimulus k was judged greater than stimulus j. Diagonal cells are normally left vacant. Matrix $\mathbf{P}$ is derived from matrix $\mathbf{F}$ by dividing each element by the number of trials for that pair, so that each element p'_{jk} denotes the **proportion** of times stimulus k was judged greater than stimulus j.

Matrix $\mathbf{X}$, the basic transformation matrix, is constructed from matrix $\mathbf{P}$. Each element x'_{jk} is the unit normal deviate corresponding to the element p'_{jk}, and may be obtained by integrating the unit normal curve. The matrix $\mathbf{X}$ is skew-symmetric, since $x'_{jk} = -x'_{kj}$. The elements of this basic transformation matrix are the normal deviates x_{jk} in the LCJ equations.

Some important observations about the generated scale values should be noted. First, the scale values describe only the **relative** positions among the stimuli, on the scale of perceptual distinctness of their target patterns. For this reason, the location of the absolute zero point on the scale is unknown. In the case of complete data, the procedure generates scales such that their mean value is zero. For incomplete data, the least selected stimulus is assigned a scale value of zero, and the scale is constructed upwards from that point. But if there are still not enough finite elements to estimate a scale for a particular stimulus, then the procedure assigns that stimulus a value of infinity.



Figure 4.5: The author collecting eye tracking data using ITEMS.

4.2 ITEMS – The Integrated Testbed for Eye Movement Studies

4.2.1 Overview and Utility of ITEMS

This section discusses the design and implementation of ITEMS – the Integrated Testbed for Eye Movement Studies. This unique eye tracking testbed consists of an integrated system of both hardware and software which allows an experimenter to present an observer with an image displayed on a high resolution monitor and have the observer perform a visual task. Figure 4.5 shows the author collecting eye tracking data using ITEMS. Afterward, the experimenter can study the data by playing back the experiment, this time with a cursor which moves about the images indicating the observer's fixation points over time. The experimenter can

also analyze the data by computing various fixation time statistics for each image.

The primary hardware components of ITEMS are a Silicon Graphics Indy computer workstation with high resolution color monitor, a Sony CCD black and white video camera fitted with a 50mm lens and 5mm lens spacer, a Datacube MaxTD image processing system containing a MaxVideo200 pipeline processor and MVME-167 CPU system controller, and an adaptable yet sturdy apparatus to which the camera is mounted as well as a helmet for restricting observer head movements. The software components of ITEMS include an X-Windows application to handle the image scene display and observer response registration for the Indy workstation, pupil centroid tracking and registration for the Datacube MaxTD, a program for fixation point estimation and head movement adjustments, a program for spatial calibration and error interpolation, and another for calculation of target fixation and identification statistics.

ITEMS presents an unique environment for conducting psychophysical experiments, in which one can study not only an observer's judgments of various image stimuli, but also some information about the process by which the observer arrived at his decisions. For instance, while XPET allows an experimenter to study the relative amounts of target distinctness among a set of stimulus images, it provides no physical evidence as to what perceptual **cues** were used by the observers in making their judgments. The experimenter can only look at the decisions made by the observers and guess at the features in the image which attracted the observers' attention. With ITEMS, not only will one know the decisions made by the observers, but also exactly where in the images they fixated and for how long they fixated

these regions.

ITEMS also gives an experimenter the ability to study the process of human visual search for targets within a complex scene. Not only can we determine whether or not a particular target was identified by an observer, but also whether or not the target was ever fixated by the observer (if it was not identified as being a target), how long did it take before the target was first fixated, how long the target was studied before it was identified, what search path the observer took on the way to the target, and any number of other aspects of the process of visual search.

4.2.2 Conducting a Visual Search Experiment Using ITEMS

Conducting a visual search experiment using ITEMS requires cooperative and patient observers. Each data collection session begins with an explanation to the observer of what is required of him. Then the observer sits at the testbed and positions himself comfortably. He inserts his head into the helmet, and the reference mark is affixed just below his left eye. The experimenter then makes necessary adjustments to the video camera and illumination until adequate video of the observer's eye and reference mark is obtained. The software for image scene display and observer response registration is started on the Silicon Graphics workstation and the software for pupil centroid tracking and registration is started on the Datacube MaxTD system. Once both programs are running, the observer is asked to commence displaying the images and identifying targets. This procedure is described in detail later in Section 4.2.4.

All required images are created by the experimenter beforehand. This includes

a zero point image, several calibration images, and all desired experiment image scenes. The procedure is that the zero point image is first, then all calibration images in succession, the zero point image again, and then any number of sessions of experiment images, each pair of sessions separated by another presentation of the zero point image. The number of experiment sessions and number of images in each session can vary, but it has been found that five images per session with one calibration session and two experiment sessions results in a moderate ten minutes of data collection for each observer. This is the chronological scheme of image presentation illustrated in Figure 4.6.

The zero point image is an image with one target that is located such that it is approximately directly ahead of the observer's left eye when displayed on the monitor. The target consists of a square region of uniform gray level against a background of a different gray level. This image is used to establish a reference point to which all eye movements can be related and also to measure periodically the change in fixation point estimates that is due to small head movements accumulating over time. This procedure is described later in Section 4.2.4.

Each calibration image consists of a row of square targets. The targets in all the calibration images taken together constitute an array of evenly spaced points, which are used as ground truth with which to measure the error in fixation point estimates due to measurement and modeling error, which does not vary over time. As these errors will vary over different spatial locations in the display image, a number of samples are taken and then adjustments are made in fixation point estimates from an interpolation of the calibration samples from the vicinity of each estimate. This

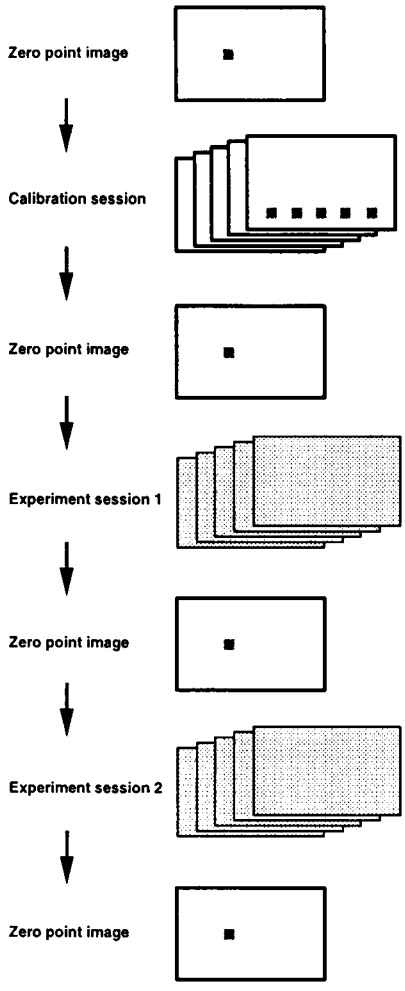


Figure 4.6: The chronological scheme of image presentation for ITEMS data collection.

procedure is described later in Section 4.2.4.

4.2.3 ITEMS Hardware Configuration

Figure 4.7 shows the interconnectedness of the various hardware components of ITEMS. Briefly, the Silicon Graphics INDY workstation is used to load each image scene stimulus from disk and display it on its high resolution color monitor. The observer uses mouse buttons to indicate when to display each image, when he wishes

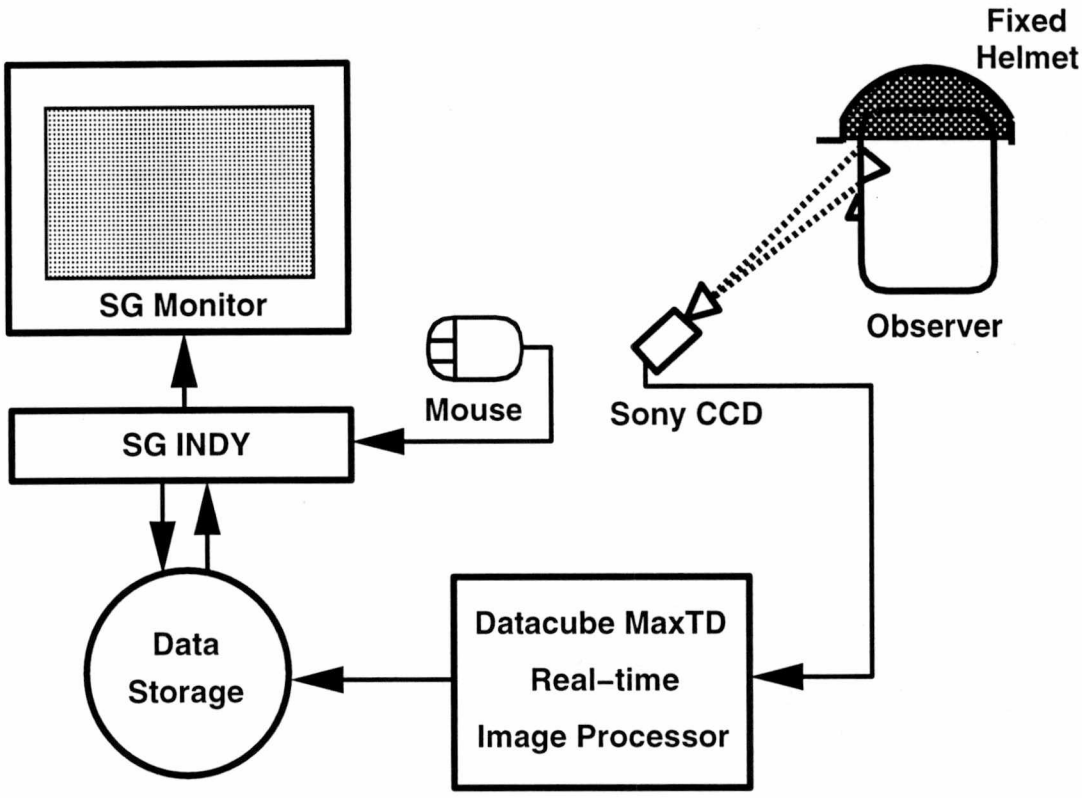


Figure 4.7: The interconnectedness of the various hardware components of ITEMS.

to identify a suspected target, and when he is finished searching a particular scene. The Sony CCD video camera sends continuous video to the Datacube MaxTD image processor, which locates, tracks, and records the pupil centroid location of the observer's left eye. The observer's head movement is restricted using a baseball batting helmet which is rigidly mounted to the table, using adjustable aluminum extrusion material. This material allows the helmet to be raised or lowered to accommodate different observers and also to be locked into place when the appropriate position is found. Head movement is further restricted by a chin rest stand (not shown).

The camera is mounted to a pan-tilt unit which can be manually adjusted so as to provide an adequate image of the observer's left eye and a small reference mark affixed just below the eye. This reference is a small, glossy black, paper circle,

used to distinguish eye movements from small head movements. The pan-tilt unit is mounted to the same aluminum extrusion material to which the helmet is mounted. This allowed the camera to be easily moved to various viewing positions to determine the most successful, which was finally decided to be directly in front of and below the observer, just below the Silicon Graphics monitor, looking upward at the observer's left eye. This position is necessary in order to obtain a clear view of the pupil even for observers with especially long or obscuring upper eyelashes. The observer's face is illuminated with a small portable flashlight as necessary to segregate the pupil and reference mark together from the rest of the video image. The Datacube MaxTD also has a small terminal screen which allows the experimenter to monitor the status of the image processor's eye tracking, and the video from the CCD camera is simultaneously displayed on a small monitor for the same purpose. The entire testbed sits on a single table, in a small room separated from the main laboratory to eliminate distractions.

4.2.4 ITEMS Software Configuration

Image Scene Display and Observer Response Registration

Image scene display and observer target identification response registration for ITEMS is handled by the Silicon Graphics INDY workstation. The X-Windows application created for this purpose is called I_SPY. I_SPY is used to load each image scene stimulus from disk and display it on the high resolution color monitor. In experiment mode, the observer uses mouse buttons to indicate when to display each image, when he wishes to identify a suspected target, and when he is finished

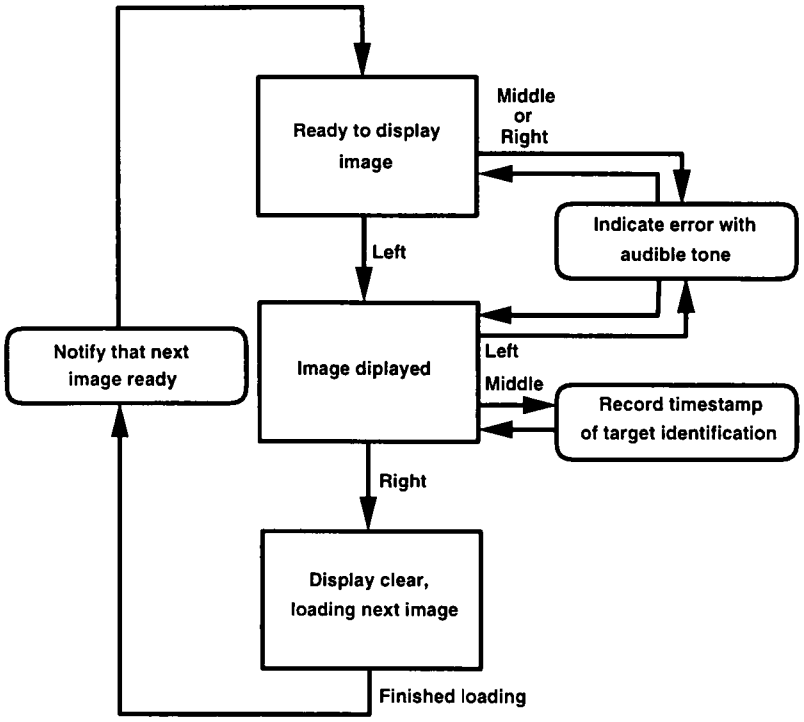


Figure 4.8: The observer control sequence for I_SPY in experiment mode.

searching a particular scene. In playback mode, I_SPY allows the experimenter to study the data by displaying the image scene stimuli with a cursor which moves about the images indicating the observer's fixation points over time.

The observer is in complete control of the image scene display during an eye tracking experiment. This means the observer can rest as needed between image sessions and display the next scene only when he is ready, allowing him to concentrate fully on the visual search task. This control is executed through the use of the three separate mouse buttons. Figure 4.8 shows the observer control sequence for I_SPY in experiment mode. From Figure 4.8, we see that the program can be in any one of three states, indicated by the rectangles. The "left," "middle," and "right" labels represent a mouse button pressed by the observer. The oval shapes

represent actions taken by the program. The top rectangle represents the state when the display is clear but and image is loaded into main memory and ready to display. The image will not display until the observer clicks the left mouse button to indicate that he is ready. If he clicks the middle or right button at this time, I_SPY simply makes a "beep" sound and continues to wait for the proper button.

Once the stimulus image is displayed, the observer uses the middle mouse button to identify each suspected target, by clicking the button at the moment he is fixating the area. Only the mouse button is used for this, as there is no mouse pointer icon on the screen. I_SPY records the timestamp of this target identification, and the observer's fixation point at the moment of the timestamp is used later to determine the location of the suspected target. If the left button is pressed during this time, I_SPY will again make a beep sound. The observer presses the right button when he is finished conducting his search of the scene, at which time the display is cleared. If the scene is the final image stimulus of the experiment, then the data is saved to disk and the experiment is concluded. Otherwise, I_SPY begins loading the next image from disk. This usually requires about five seconds (for an 800×1200 image), after which a message is printed to tell the observer that the next image is ready.

For each image stimulus, the proper button sequence is left to initially display the image, middle to identify each suspected target, and then right to end the display of the current image and to begin loading the next image from disk. This left-middle-right sequence is constant for both calibration and experiment images, and is quickly learned by the observer.

The file names of the images to be displayed as stimuli during an experiment are

Table 4.1: An example of the data resulting from a short I_SPY experiment session.

BUTTON	TIME (SEC)	TIME (USEC)
left	837728950	220974
middle	837728953	727284
middle	837728966	534179
middle	837728971	246923
right	837728982	866686
left	837728995	39562
middle	837728997	646466
middle	837729000	206188
middle	837729002	904575
middle	837729007	356105
middle	837729010	279462
middle	837729012	926056
middle	837729019	789389
middle	837729021	95531
middle	837729023	926969
right	837729025	835767
left	837729037	86907
middle	837729038	885492
middle	837729040	246572
middle	837729041	666052
middle	837729042	715741
middle	837729044	485167
middle	837729046	249585
right	837729057	705060

read from a file named 'I_SPY_experiment_images'. Each file name must appear on a separate line, each image must be in portable gray map or "pgm" format, and all images must be of the same dimensions. The data resulting from a single session is stored in a file named 'I_SPY_result'. Sample data from a short I_SPY experiment session is given in Table 4.1. The data consists solely of the timestamps of all

(non-error) mouse button presses during the session. Each "left" press indicates the time at which an image was displayed, each "middle" indicates the time at which the observer identified a suspected target, and each "right" is the time that the observer finished with the image. This information will later be corresponded to the fixation point data. Each timestamp consists of two long integers indicating the number of seconds and microseconds since January 1, 1970, the information obtained using the "gettimeofday()" function in Unix.

Pupil Centroid Tracking and Registration

The tracking and registration of pupil centroid position is handled by the Datacube MaxTD image processing system. The program created to accomplish this is called "I_TRACK". This program makes use of a library of code distributed with the MaxTD system called "Sill." Sill includes a very useful function for performing connectivity, or "blob," analysis of a binary image. Figure 4.9 shows the general flow of I_TRACK. The procedure is to first grab a single image frame from the incoming video. This image is subjected to a thresholding operation, in which all pixels with a gray level less than a threshold value t are set to black and all others are set to white. The threshold value t is input by the experimenter and is chosen such that the both the observer's pupil and the small, glossy black, paper circle affixed just below his eye and used as a reference mark appear as black circular blobs in the image. Often several different threshold values must be tried, along with adjustments of illumination and camera aperture, to obtain suitable image video. These settings will be different for every observer and can be obtained only

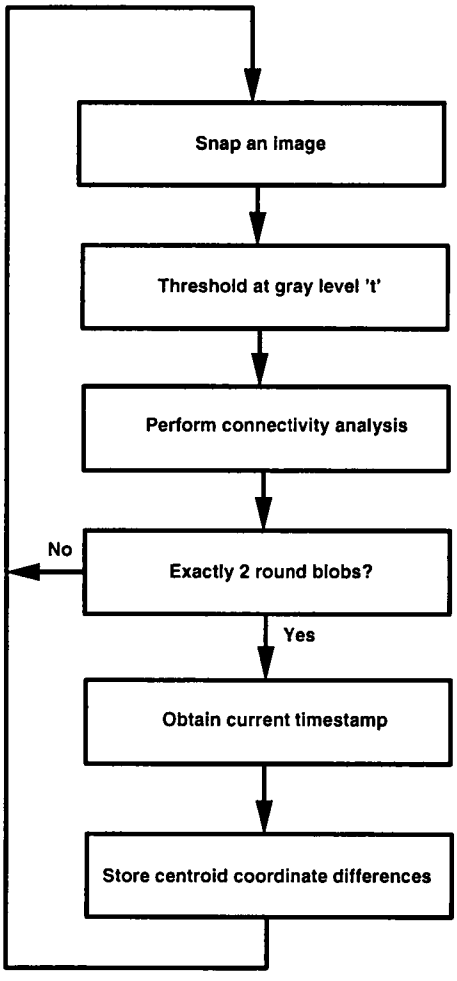


Figure 4.9: The general flow of the "I_TRACK" program.

by some amount of trial and error.

The binary image obtained in this manner is then subjected to Sill's connectivity analysis routine. This routine returns the number of blobs that were found in the image, along with several other pieces of information about the blobs. As there may well be various shadows and other artifacts in the image which produce black blobs in addition to the two blobs representing the pupil and reference mark, all the blobs are tested to find the two of interest. This is accomplished by finding a best-fit ellipse for each blob, and calculating the ratio of the length of the minor

axis to the length of the major axis of this ellipse. This results in a measure of roundness for each blob, a value between 0.0 and 1.0, where 1.0 would indicate a perfect circle. In practice, roundness values of 0.8 to 0.9 are obtained for the blobs representing the pupil and reference mark in the best conditions. Occasionally, due to various shadows, such as from eyeglasses or contact lenses, and other artifacts such as eyelashes that partially obscure the pupil, the roundness values for the two blobs of interest are much lower. For this reason, a roundness value of 0.4 is used to determine if a blob is sufficiently round. This value has been found to be acceptably effective at separating the pupil and reference mark blobs from various shadow artifacts, which generally do not appear as round blobs at all.

If, as a result of the connectivity analysis and roundness tests, there are determined to be any number of round blobs other than two, then the program immediately proceeds to snap another image frame. This results in a missed data sample, but it is more desirable to loose a sample than to allow an erroneous one. Most often, this case occurs when the image frame was obtained at a moment when the observer's eye was closed due to blinking, in which case only the reference mark blob is visible. As the observer cannot be fixating any part of the image at the moment of a blink, we certainly wish to ignore this sample and proceed to obtain another.

If the roundness tests yield exactly two blobs, then the current timestamp is obtained using the function "gettimeofday()". Then the pixel coordinates of the centroids of the best-fit ellipses for the two blobs are obtained from the connectivity analysis. The values that are stored are the values of the differences in both x-coordinates and y-coordinates between the upper blob (the pupil) and the lower

blob (the reference mark). Thus it is only movement of the pupil relative to the reference mark that is tracked and registered. In this way, movements of the eye can be distinguished from small head movements. That is, a small head movement will result in a change of position of both the pupil and the reference mark in the camera image. Depending on the type of movement, i.e., a translation or a rotation in any of the various directions, the relative positions of the two features will change only slightly. If the pupil alone was tracked, then even a small head movement will be interpreted absolutely as eye movement, and the estimate of the observer's fixation point in the image displayed on the monitor can change by a rather large amount. This phenomenon will be explained in greater detail when the geometry of the problem is presented in the next section. Although a helmet mounted in a fixed position and a chin rest are used to restrict observer head movement, in practice there is still a bit of a small head movement even in the most cooperative observers, due to breathing, heartbeats, etc.

So I_TRACK repeatedly cycles through this process, obtaining an image from the incoming video and trying to locate the two round blobs representing the pupil and the reference mark. The processing required for each frame prevents the cycle from matching the rate of the incoming video, which is 30 frames/sec. Most of the processing time required is due to the connectivity analysis. Although the MaxTD image processing system is capable of performing many types of operations, such as image filtering, at frame rate, geometrical analysis of the various features in the video is more temporally challenging. In practice a sampling rate of about 7.5 frames per second is obtained, meaning we get estimates of the observer's fixation point

Table 4.2: An example of the data resulting from an I_TRACK session.

TIME (SECS)	TIME (USECS)	X DIFFERENCE	Y DIFFERENCE
839976788	150000	12.748628	-267.083477
839976788	280000	12.547692	-267.373825
839976788	420000	12.705235	-267.140801
839976788	550000	12.844859	-267.514080
839976788	690000	15.139511	-269.483062
839976788	820000	15.045757	-269.250315
839976788	950000	14.303707	-268.815163
839976789	90000	14.386379	-268.644339
839976789	220000	14.263808	-268.995103
839976789	350000	14.302111	-268.862772
839976789	490000	14.228385	-268.973487
839976789	620000	10.829218	-265.168341
839976789	750000	10.917554	-265.078341
839976789	890000	11.207165	-265.215847
839976790	20000	11.388494	-265.344565
839976790	150000	11.314533	-265.348253
839976790	290000	13.102178	-267.944246
839976790	420000	13.075749	-267.794912
839976790	550000	13.084578	-267.752992
839976790	690000	12.663160	-266.888320
839976790	820000	12.426472	-266.729049
839976790	950000	14.227137	-264.292227

about every 0.13 seconds. No estimate is obtained only when an analyzed frame is of an instant when the observer is blinking. A small portion (for only three seconds!) of the data that is collected and stored by I_TRACK is given in Table 4.2.

Fixation Point Estimation and Head Movement Adjustments

Eye Tracking Geometry The program to estimate the observer's fixation points in the display image is executed after the eye tracking session is completed. This

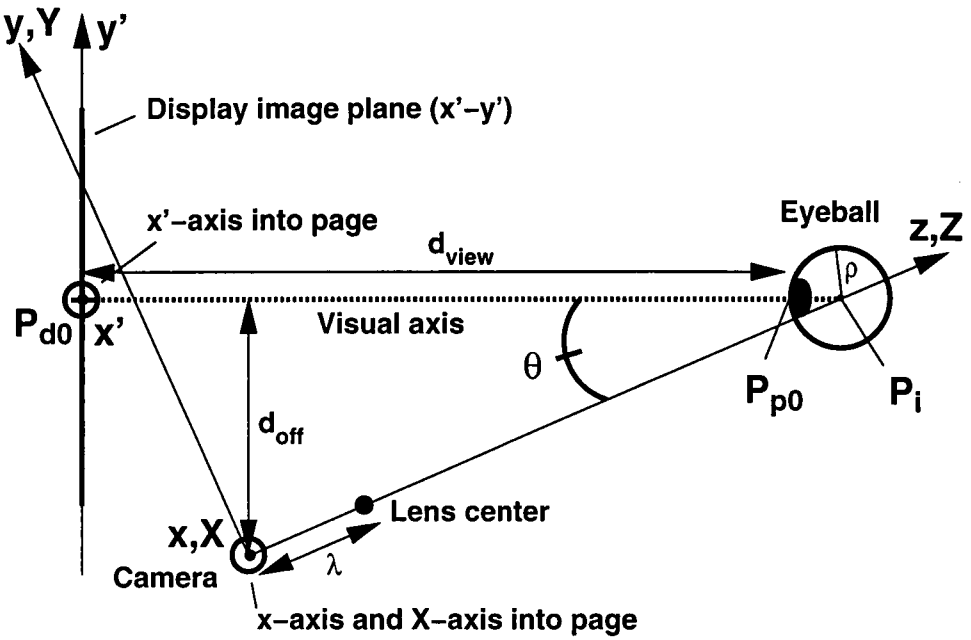


Figure 4.10: The geometrical model of the eye tracking problem.

program, called "I_WHERE," takes as inputs the data file from I_SPY (such as the example data in Table 4.1), the data file from I_TRACK (such as the example data in Table 4.2), and a third data file which contains the various measurement data necessary to estimate the point at which the observer's visual axis intersects the display image. The geometrical model of the problem is shown in Figure 4.10. The state depicted in the figure is that when the observer is looking straight ahead, such that his visual axis is perpendicular to the display image. This situation will be referred to as the zero state.

The camera coordinate system (x, y, z) is aligned with the world coordinate system (X, Y, Z), such the x-axis and X-axis represent the horizontal axis of the camera image, the y-axis and Y-axis represent the vertical axis, and the z-axis

and Z-axis extend outward down the camera view. The camera is located directly beneath the observer's visual axis when he is looking straight ahead. The center of the camera lens is located at $(0, 0, \lambda)$. The observer's eyeball is modeled as a sphere with radius ρ, centered at point P_i located at $(0, 0, Z_i)$. The display image is modeled as a plane. The display zero point, P_{d0}, is the point in the display image which is fixated in the zero state. The point P_{d0} is also considered the origin of the display image coordinate axes, x' and y'. The angle θ is the angle formed between the Z-axis and the observer's visual axis in the zero state. The visual axis is presumed to pass through both the center of the pupil and the eyeball center P_i. The center of the pupil in the zero state is P_{p0}, located at $(0, \rho \sin\theta, Z_i - \rho \cos\theta)$.

Location of the Display Zero Point The location of the display zero point P_{d0} in world coordinates is most easily found using vectors. Figure 4.11 shows the same geometry as in Figure 4.10, but with only the world coordinate axes and the vectors of interest shown. Once again, the world coordinate axes are aligned with the camera image. The vector $\mathbf{v}$ is the vector from the world origin to the center of the eyeball P_i. Thus, $\mathbf{v} = \langle 0, 0, Z_i \rangle$. The vector $\mathbf{w_0}$ is the vector from the eyeball center P_i through the pupil center P_{p0} to the display zero point P_{d0}. The length of $\mathbf{w_0}$ is evident from Figure 4.10, $\|\mathbf{w_0}\| = d_{view} + \rho$. If we translate $\mathbf{w_0}$ so that its initial point is at the origin, the components of $\mathbf{w_0}$ become evident. The x-component of $\mathbf{w_0}$ is zero, since the camera image center, eyeball center P_i, and display zero point P_{d0} all lie on the Y-Z plane. The angle that the translated $\mathbf{w_0}$ forms with the z-axis is still θ, so the y-component of $\mathbf{w_0}$ is $(d_{view} + \rho) \sin\theta$. Likewise, the z-component of $\mathbf{w_0}$ is $(d_{view} + \rho) \cos\theta$. Thus we have $\mathbf{w_0} = \langle 0, (d_{view} + \rho) \sin\theta, -(d_{view} + \rho) \cos\theta \rangle$.

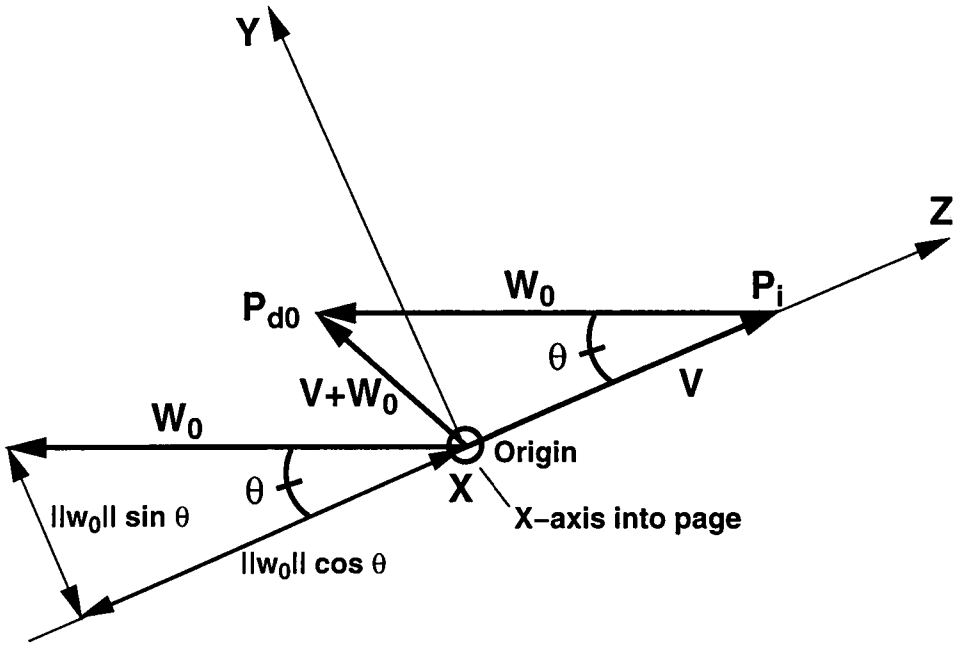


Figure 4.11: The vectors of interest in the eye tracking problem.

The location of P_{d0} is found by adding $\mathbf{v}$ and $\mathbf{w_0}$, which gives us:

$$P_{d0} = (0, (d_{view} + \rho)\sin\theta, Z_i - (d_{view} + \rho)\cos\theta). \tag{4.4}$$

Display Image Plane and Visual Axis Equations We now need an equation for the display image plane. We know that the plane passes through the point P_{d0} and that the vector $\mathbf{w_0}$ is a normal to the plane. The point-normal form of the equation of a plane passing through the point $P_0(X_0, Y_0, Z_0)$ and perpendicular to the nonzero vector $\mathbf{n} = \langle a, b, c\rangle$ is [Ant84]:

$$a(X - X_0) + b(Y - Y_0) + c(Z - Z_0) = 0. \tag{4.5}$$

Thus, the equation of the display image plane in our model is:

$$\|\mathbf{w_0}\| \sin\theta(Y - \|\mathbf{w_0}\| \sin\theta) + (-\|\mathbf{w_0}\| \cos\theta(Z - Z_i + \|\mathbf{w_0}\| \cos\theta) = 0, \qquad (4.6)$$

where $\|\mathbf{w_0}\| = d_{view} + \rho$.

Now suppose the observer's eyeball rotates so as to fixate another point in the display image. The center of the pupil moves from P_{p0} to a new point, which we will call P_{p1}, located at (X_{p1}, Y_{p1}, Z_{p1}). With the center of the eyeball still at P_i, the observer's visual axis now lies along a vector from P_i to P_{p1}. The vector from P_i to P_{p1} will be designated $\mathbf{w_1}$. Since the components of any vector are the coordinates of its terminal point minus the coordinates of its initial point [Ant84], we have:

$$\mathbf{w_1} = \langle X_{p1}, Y_{p1}, Z_{p1} - Z_i \rangle . \qquad (4.7)$$

The line passing through (X_0, Y_0, Z_0) and parallel to the nonzero vector $\langle a, b, c \rangle$ has parametric equations [Ant84]:

$$X = X_0 + at \qquad (4.8)$$

$$Y = Y_0 + bt \qquad (4.9)$$

$$Z = Z_0 + ct. \qquad (4.10)$$

Thus the line representing the new visual axis, passing through P_i and parallel to the vector $\mathbf{w_1}$, has parametric equations:

$$X = X_{p1} + X_{p1}t = X_{p1}(t+1) \qquad (4.11)$$

$$Y = Y_{p1} + Y_{p1}t = Y_{p1}(t+1) \qquad (4.12)$$

$$Z = Z_{p1} + (Z_{p1} - Z_i)t = Z_{p1}(t+1) - Z_i t. \qquad (4.13)$$

Finding the Intersection Point Now we have an equation for the display image plane and parametric equations describing the line along which the visual axis lies. The task now is to find the point at which the line intersects the plane, which we will designate P_{d1}. The coordinates of this point satisfy the equation of the plane and the equations for the line. Let the value of t for which this true be represented by t_1. We now set the value of t in Equation 4.11 through Equation 4.13 equal to t_1, and then insert these three equations for X, Y, and Z into Equation 4.6:

$$\|\mathbf{w_0}\| \sin\theta(Y_{p1}(t_1+1)-\|\mathbf{w_0}\| \sin\theta)+(-\|\mathbf{w_0}\| \cos\theta(Z_{p1}(t+1)-Z_i t_1-Z_i+\|\mathbf{w_0}\| \cos\theta) = 0. \tag{4.14}$$

Now we need to solve for t_1:

$$\begin{aligned} t_1 &= \frac{-Y_{p1}\sin\theta + \|\mathbf{w_0}\|\sin^2\theta + Z_{p1}\cos\theta - Z_i\cos\theta + \|\mathbf{w_0}\|\cos^2\theta}{Y_{p1}\sin\theta - Z_{p1}\cos\theta + Z_i\cos\theta} \\ &= -1 + \frac{\|\mathbf{w_0}\|}{Y_{p1}\sin\theta - Z_{p1}\cos\theta + Z_i\cos\theta}. \end{aligned} \tag{4.15}$$

Equation 4.15 along with Equation 4.11 through Equation 4.13 give us the world coordinates of P_{d1}, the new intersection point of the observer's visual axis with the display image plane. The remaining task is to find the coordinates of this intersection point in terms of the axes representing the display image plane, x' and y'. This is accomplished by finding the relative differences $\Delta x'$ and $\Delta y'$ between the new point P_{d1} and the display image zero point, P_{d0}, given by Equation 4.4. The point P_{d0} is also the origin of the x' and y' axes. Since the x'-axis is parallel to the x-axis (see Figure 4.10), we have:

$$\Delta x' = X_{d1} - X_{d0} = X_{d1}. \tag{4.16}$$

The difference in y' coordinates must therefore be the distance between the two points in the y-z plane:

$$\Delta y' = \pm\sqrt{(Y_{d1} - Y_{d0})^2 + (Z_{d1} - Z_{d0})^2}, \tag{4.17}$$

where $\Delta y'$ is positive if $Y_{d1} > Y_{d0}$.

Location of Pupil Center Point We now are able to obtain an estimate of the observer's fixation point in the display image for a given location in world coordinates of the pupil center. Now we must obtain the relationships necessary to yield the pupil center location in 3-D space from its location the camera image. In our model, the world coordinate system is aligned with the camera coordinate system. The equations relating the location of a point in the camera image to its location in 3-D space in this situation are [GW92]:

$$x = \frac{\lambda X}{\lambda - Z} \tag{4.18}$$

$$y = \frac{\lambda Y}{\lambda - Z}. \tag{4.19}$$

These equations constitute a perspective transformation, where the point (X, Y, Z) in the world is projected to point (x, y) in the camera image, and λ is the distance from the origin to the center of the lens. In our problem, the image location of the point is known and it is the world location of the corresponding point that is desired. Since there are three unknown coordinates and only two equations, we need one more relationship from our model to solve the problem. This relationship is that the pupil center point P_{p1} is located on the sphere representing the eyeball, with radius ρ and center at point $P_i(0, 0, z_i)$:

$$\rho^2 = X_{p1}^2 + Y_{p1}^2 + (Z_{p1} - Z_i)^2. \tag{4.20}$$

The procedure is to solve for X and Y in Equation 4.18 and Equation 4.19, substitute these expressions as X_{p1} and Y_{p1} in Equation 4.20 and solve for Z_{p1}. This is a quadratic relationship, so two solutions are obtained. The geometric meaning of this is that the straight line of points in 3-D space that are mapped to the single image point by the perspective transformation passes through the sphere representing the eyeball in two points. The pupil will always lie on the front portion of the eyeball, so the lesser of the two solutions is taken as the true value of Z_{p1}. This value is then substituted back in Equation 4.18 and Equation 4.19 to yield X_{p1} and Y_{p1}.

Using Pupil and Reference Point Coordinate Differences Now we can estimate the observer's fixation point in the image from the location of the pupil center in the camera image. But the data from I_TRACK gives only the coordinate differences between the pupil centroid and the reference point centroid at each sample time, such as in Table 4.2. The procedure is to scan the data from I_SPY (such as the example data in Table 4.1) for the time of the first click of the middle button. This is the moment at which the observer was fixating the center of the display zero point target in the first stimulus image, which means the pupil center was at the point P_{p0} in our model. The corresponding location of the pupil center in the camera image is computed using Equation 4.18 and Equation 4.19. The timestamp at this point in the I_SPY data is used to find the closest sample in the I_TRACK data. The coordinate differences at this zero point constitute the reference to which every future sample will be compared. That is, the location of the pupil center in the camera image at any sample time is considered to be the current coordinate differences, minus the coordinate differences at the zero point time, plus the location

Table 4.3: The measurement data used by I_WHERE.

1200	display_image_width_pixels
800	display_image_height_pixels
333.0	display_image_width_mm
220.0	display_image_height_mm
525	display_zero_point_x_pixels
400	display_zero_point_y_pixels
768	camera_image_width_pixels
494	camera_image_height_pixels
6.45	camera_image_width_mm
4.84	camera_image_height_mm
55.0	λ_mm
455.0	Z_i_mm
10.0	ρ_mm
490.0	d_{view}_mm
185.0	d_{off}_mm

of the pupil center at the zero point time.

Measurement Data I_WHERE needs the values of various measurements describing the geometry of the problem in order to compute the fixation point estimates. These measurements are given in Table 4.3. The measurements are in millimeters (mm) except for those indicated in image pixels. Care must be taken to convert between pixels and millimeters as necessary to estimate the fixation points.

The value of λ, the distance from the camera image to the lens center, was measured empirically by capturing an image of a ruler with millimeter markings and finding the corresponding distance in image pixels. The value of ρ is set at 10mm, since the human eye has an average diameter of approximately 20mm [GW92].

Head Movement Adjustments The observer's head position is fixed by the use of the helmet and chin rest. In addition, the use of the reference mark just below the observer's eye helps to distinguish between eye movements and small head movements. But even with these efforts, any head movements, regardless of how small, will cause error in the fixation point estimates. This error is measured by having the observer periodically identify the zero point target. As explained previously in Section 4.2.2, data collection consists of a calibration session and one or more experiment sessions. As illustrated previously in Figure 4.6, the zero point image appears at the beginning, at the end, and between each pair of sessions. The sample at the moment the observer identifies the first zero point is taken as the reference for all later samples, so the resulting fixation estimate at this moment is exactly the zero point, i.e., the center of the zero point target in the display image. The error due to head movement over time can be taken as the difference in the fixation estimate for every subsequent zero point identification from the true location of the zero point.

The result is that for every session, we have a measurement of the head movement error at the beginning and at the end of the session. There is no way of knowing exactly when and in what direction the head movements occurred, but we can compensate a bit for this error by considering the error during any particular session

to be the average of the error at the beginning and at the end of the session. This compensation is accomplished by subtracting this average session error from every fixation estimate for the session.

Review of Procedure and Example Data To review, the steps necessary to obtain the fixation estimate for each data sample are:

1. Extract from the I_TRACK data the values for the difference in x- and y-coordinates between the pupil centroid and the reference point centroid.

2. Compare these values to the same values from the moment the observer identified the first zero point. The change is taken to be the movement of the pupil center in the camera image from the zero state.

3. From the location of the pupil center in the camera image, its location in world coordinates is found according to Equation 4.18 and Equation 4.19, subject to the constraint that the point also lies on the surface of the sphere representing the eyeball, described by Equation 4.20.

4. Based on the location of the pupil center in world coordinates, Equation 4.15 is used to find the value of t at the point of intersection of the visual axis and the display image plane, and this value is substituted into Equation 4.11 through Equation 4.13 to yield the coordinates of this intersection point.

5. The location of the intersection point in the display image plane in terms of the display image axes is found using Equation 4.16 and Equation 4.17, which give the coordinates of the point relative to the display zero point.

Table 4.4: An example of the fixation point data generated by I_WHERE.

TIME (SECS)	TIME (USECS)	X FIXATION	Y FIXATION	BUTTON
839976788	150000	0.438019	0.626717	NONE
839976788	280000	0.439989	0.631141	NONE
839976788	420000	0.438442	0.627590	NONE
839976788	550000	0.437226	0.633276	NONE
839976788	690000	0.416251	0.663312	NONE
839976788	820000	0.417056	0.659756	NONE
839976788	950000	0.423881	0.653110	NONE
839976789	90000	0.423054	0.650503	NONE
839976789	220000	0.424311	0.655857	NONE
839976789	350000	0.423911	0.653837	NONE
839976789	490000	0.424637	0.655527	NONE
839976789	620000	0.455705	0.597602	NONE
839976789	750000	0.454850	0.596233	NONE
839976789	890000	0.452137	0.598319	NONE
839976790	20000	0.450449	0.600273	NONE
839976790	150000	0.451149	0.600330	NONE
839976790	290000	0.434922	0.639830	NONE
839976790	420000	0.435129	0.637554	NONE
839976790	550000	0.435034	0.636915	NONE
839976790	690000	0.438774	0.623746	NONE
839976790	820000	0.440965	0.621323	NONE
839976790	950000	0.423532	0.584267	NONE

6. The fixation point estimate is adjusted for small head movements by subtracting the average of the error at the beginning of the session and at the end of the session.

An example of the data generated by I_WHERE is given in Table 4.4. This example data is for the same data samples as in the I_TRACK example data presented previously in Table 4.2. The values of x-fixation and y-fixation are represented as

fractional components of the entire display image, where 0.0 corresponds to the left boundary for x-fixation and the bottom boundary for y-fixation, 1.0 corresponds to the right and top boundaries. The last column of the data indicates the mouse button that the observer was pressing at that particular moment. This information is extracted from the I_SPY data file.

Spatial Calibration and Error Interpolation

Measurement and Modeling Error The geometrical model used to obtain fixation estimates is a simple one, which varies from the real world in several ways. The observer's eyeball is probably not a perfect sphere. The pupil, which is actually an opening in the iris, does not lie on the outer surface of the eye, but is actually receded a bit, behind the cornea. The visual axis may not necessarily pass through the center of the pupil and the center of the eye. The monitor upon which the stimulus image is displayed is not a perfect plane. The centroid of the blob in the camera image representing the pupil is not precisely the center of the pupil, since the camera is viewing the eye from an angle. All of these factors combine to produce a certain amount of modeling error.

In addition to modeling error, there is a certain amount of error in each of the various measurements that are needed. The modeling and measurement error will result in fixation estimates that differ from the true fixation points. This error will also likely vary over different spatial locations in the display image. However, unlike the error due to small head movements over time for which I_WHERE attempts to compensate, the error due to modeling and measurement will at least be constant

over time. For this reason, a calibration session is conducted for each observer prior to the experiment sessions, to sample the modeling and measurement error over various locations in the display image so that fixation point estimates can be adjusted accordingly.

Finding the Appropriate Calibration Region As illustrated previously in Figure 4.6 and discussed in Section 4.2.2, the currently employed method is to use five images in the calibration session. Each image has a single row of eight targets, where each target consists of a square region of uniform gray level against a background of a different gray level. The targets in all five calibration images taken together constitute an array of 40 evenly spaced points, as illustrated in Figure 4.12. The procedure that is used is to adjust every fixation estimate according to an interpolation of the error within its calibration region. A calibration region is a quadrangle bounded by the fixation point estimates for four calibration targets. Our grid of 40 calibration targets define 28 calibration regions, as illustrated in Figure 4.12. The regions along the edges and corners are taken to extend outward indefinitely.

In actual practice, the calibration regions are defined by the calibration point fixation estimates rather than the true target locations, so the regions will not form perfectly rectangular regions. An illustration of this concept is given in Figure 4.13. The square objects represent four adjacent calibration targets. The circles represent the location of the fixation point estimate at the moment the observer identified each calibration target. The arrows represent the error at each sample point, computed as the fixation point estimate minus the true location of the target. This error has

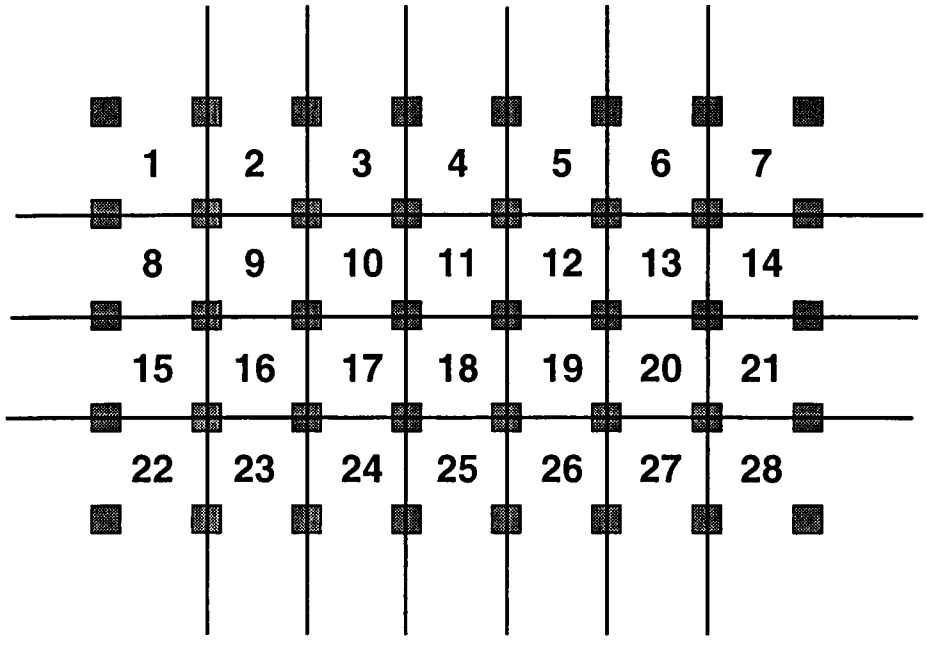


Figure 4.12: An illustration of the 28 calibration regions defined by the 40 calibration targets.

both x- and y-components. For a particular fixation point estimate $P_i(x_i, y_i)$ during the experiment, an initial guess at the appropriate region is made, based upon the rectangular regions defined by the calibration targets as in Figure 4.12. Then this initial guess is checked against the actual quadrangle region bounded by the four calibration point fixation estimates to see if this is in fact the correct region or if perhaps the point lies within another region. Equations for the four lines that form the quadrangle are derived, and the point P_i is checked to see if it lies on the proper side of each line. If so, the fixation estimate P_i adjusted by an interpolation of the error in this region. If P_i does not lie within this quadrangle, then an adjacent region is checked, according to which of the boundaries were violated. This process continues until the appropriate region for P_i is found. Again, the regions along the

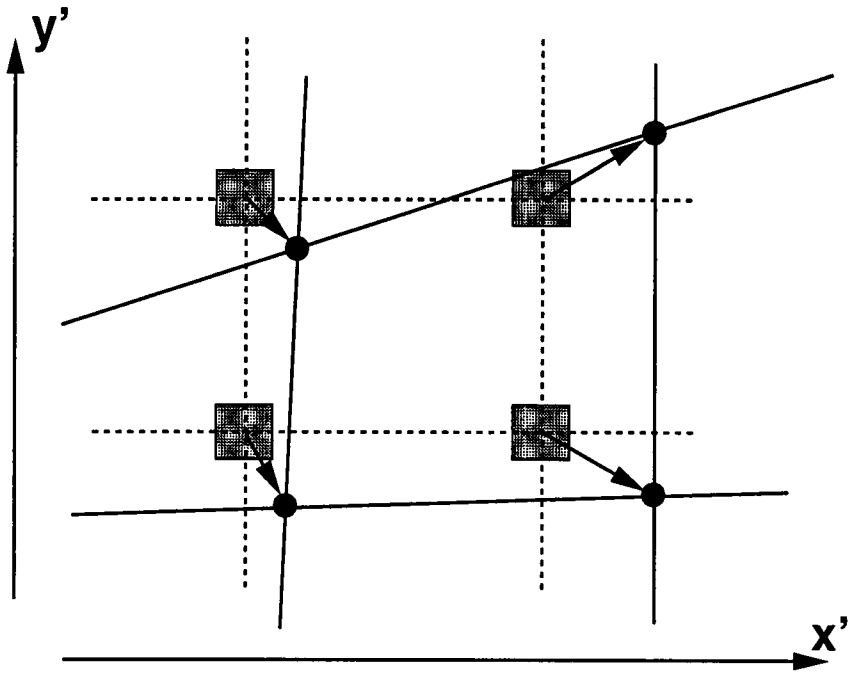


Figure 4.13: An illustration of four possible fixation point estimates (circles) along with the true locations of the targets (squares). The arrows represent the error vectors (estimate minus truth) for the estimates.

edges and corners are taken to extend outward indefinitely.

Interpolating the Error in a Region Once the appropriate calibration region is determined, both the x- and y-coordinates of the fixation point estimate are adjusted. This adjustment is made based upon planar interpolation of the error in the region. The parameters necessary for this interpolation are obtained from bivariate regression. Essentially, this treats the error as a bivariate function varying over the region, which has been sampled at $N = 4$ points. In this regression, there are $q = 3$ independent variables: x' and y' location, and one dummy variable to account for an additive constant. There are also $p = 2$ dependent variables: x-error

and y-error. This model can be expressed in the general form:

$$\mathbf{E} = \beta \mathrm{L} + \epsilon, \tag{4.21}$$

where $\mathbf{E}$ is a $p \times N$ matrix of x-error and y-error, $\mathbf{L}$ is a $q \times N$ matrix of spatial locations, β is the $p \times q$ matrix of unknown regression coefficients, and ϵ is the matrix of random errors which are presumed to be bivariate normal with a mean of zero. For a given set of observations, the least squares estimate of β is given by [SC83]:

$$\hat{\beta} = \mathbf{EL}'(\mathbf{LL}')^{-1}. \tag{4.22}$$

The β parameter estimates can now be used to obtain a planar interpolation of the x-error and y-error at any point in the region, based upon the four sample locations for which the error is known.

CHAPTER 5

First-order Image Statistics and Human Perception

5.1 Purpose of the Experiment

First-order statistics (i.e., mean and standard deviation) can be derived solely from image histogram while it is higher-order statistics, which are based on the spatial relationships of pixels, that characterize the more general concept of image **texture**. Although we must study both first-order and higher-order image statistics if we should hope to more fully understand human perception of targets and backgrounds in real scenery, the focus the psychophysical scaling experiment reported in this chapter was restricted to first-order image statistics. First-order statistics are inherently simpler than higher-order statistics, and we want to first establish our experimental procedure and verify its scientific validity before looking at more complex problems.

5.2 Creation of the Image Stimuli

To study only first-order statistics, we desire image stimuli in which the first-order statistics are carefully controlled but the higher-order statistics (based upon spatial relationships) are equal. For this reason, the images used as stimuli consisted of (spatially) uncorrelated Gaussian noise, where the pixel intensities (histogram)

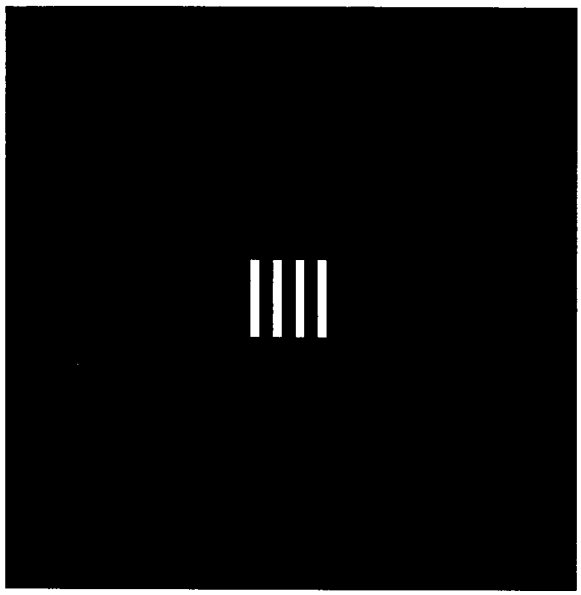


Figure 5.1: The 4-bar pattern that was used when superimposing the targets on the background.

form a Gaussian distribution with a certain mean and standard deviation. The pixels located in a vertical four-bar pattern in the center of each stimulus image were considered target pixels, while all other pixels were considered background pixels. This vertical four-bar target mask is illustrated in Figure 5.1. In every stimulus, the first-order statistics of the target pixels differed from those of the background pixels by a controlled amount. Thirteen different patterns were generated, one to serve as the constant background pattern and the other twelve to serve as the target patterns. Figure 5.2 shows the mean and standard deviation parameters of the twelve target patterns as well as the background pattern. The number (1-12) of each target pattern corresponds to the stimulus number for which that pattern served as the target. The twelve image stimuli are shown in Figure 5.3. Since there

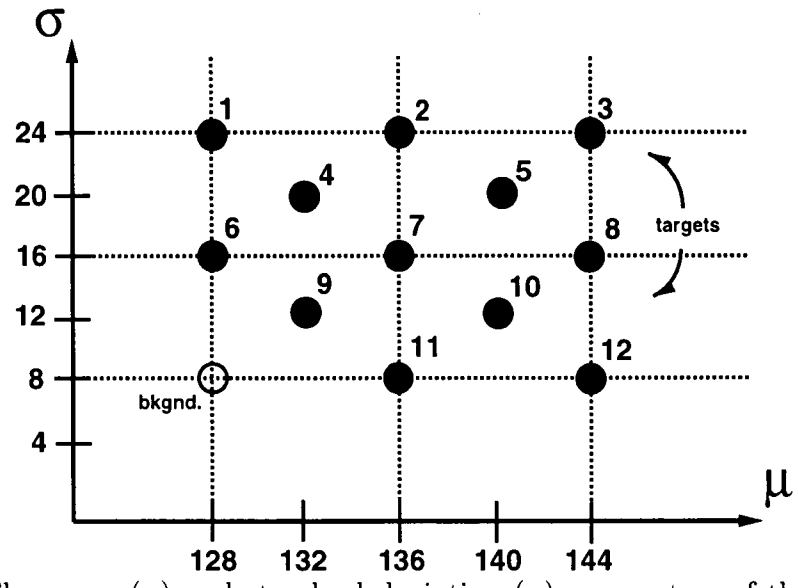
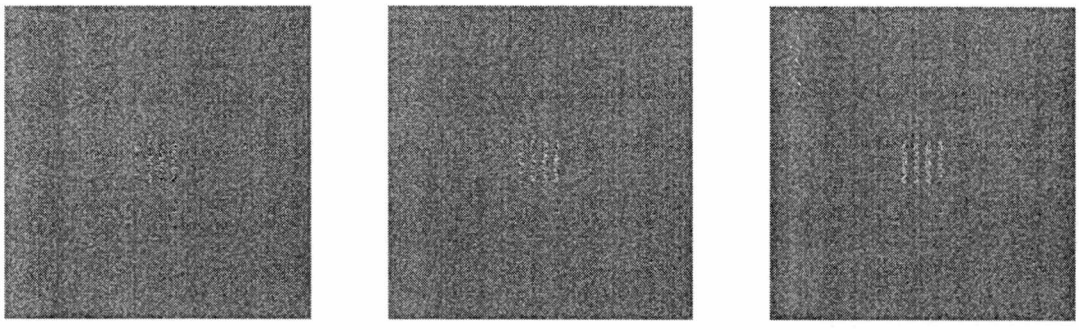


Figure 5.2: The mean (μ) and standard deviation (σ) parameters of the 12 target patterns as well as the background pattern.

were n=12 stimuli used, there were a total of $\frac{n(n-1)}{2} = \frac{(12)(11)}{2} = 66$ unique pairs of stimuli.

5.3 Data Collection

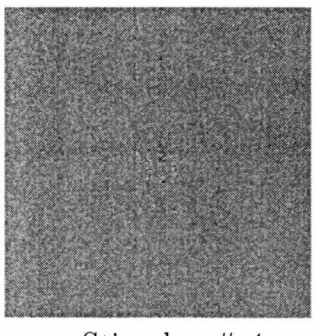
Each observer was given the exact same instructions before beginning the experiment. They were told that each image stimulus consists of a four-bar pattern of one pattern superimposed on the center of another pattern. They were shown the binary mask (Figure 5.1) that indicates the relative size and location of the four-bar pattern. They were instructed to choose, for each image pair, the image in which they thought the four-bar pattern is more distinct. Thus, the attribute being judged in this experiment was the "perceptual distinctness" of the target from the



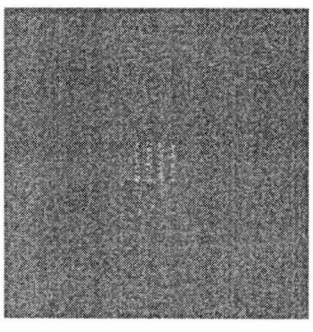
Stimulus # 1

Stimulus # 2

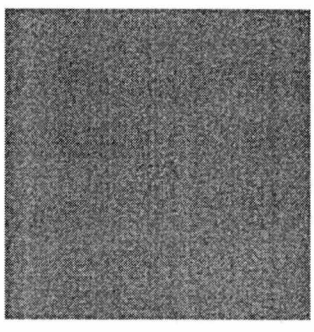
Stimulus # 3



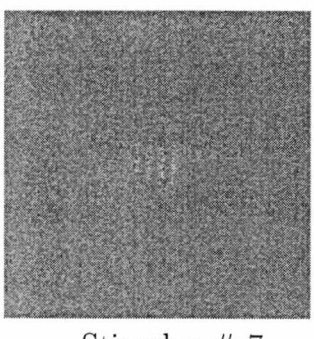
Stimulus # 4



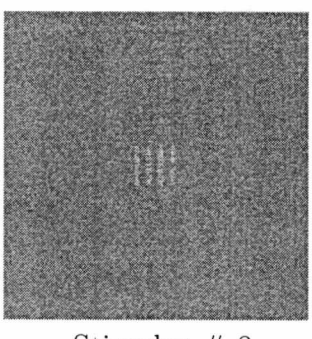
Stimulus # 5



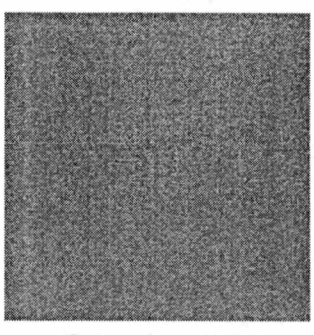
Stimulus # 6



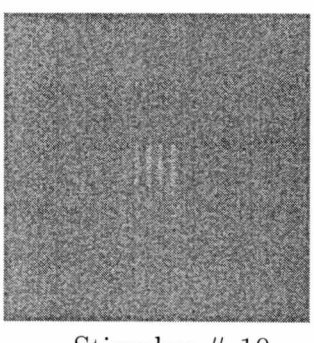
Stimulus # 7



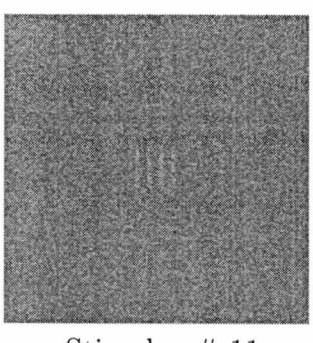
Stimulus # 8



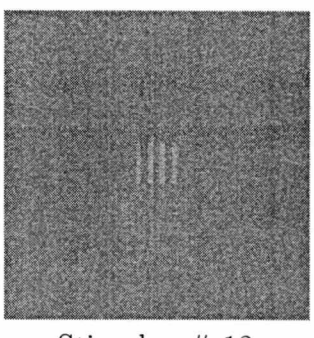
Stimulus # 9



Stimulus # 10



Stimulus # 11



Stimulus # 12

Figure 5.3: The twelve image stimuli used in the pilot experiment.

background. No specific instructions were given on what perceptual cues the test subjects were to use when deciding which target pattern was more distinct.

Nine observers were used in the experiment. Two XPET sessions were conducted for each observer on separate days, so there was a total of eighteen observations for each pair of stimuli. The presentation order of the 66 sets of stimuli used during each observer's first session was the same randomly chosen order. The presentation order used during the second session was the exact opposite: the stimulus pairs were presented in reverse order, and the left-right positioning of the stimuli in each pair was flipped. This procedure was used to reduce the effects of observer bias in the resulting data.

5.4 LCJ Analysis of the Data

Table 5.1 contains the raw data obtained during the pilot experiment. The element in row j and column k of the table is the number of times (out of 18 observations) that stimulus k was judged greater than stimulus j. Notice that all the elements along the main diagonal are zeros, since a stimulus was never compared to itself. This data was used in an LCJ analysis to estimate the scale values for the 12 stimuli, using the Case 3 assumption (Equation 4.3) and procedure explained previously in Section 4.1.3. This analysis, however, was unable to estimate the psychological "distance" from stimuli #6 and #9 to the other 10 (refer to Figure 5.2). From column 9 of Table 5.1, we see that stimulus #9 was never chosen over any other stimulus except stimulus #6. From column 6, we see that stimulus #6 was chosen only when paired with stimulus #9, except for one lone judgment over stimulus #10.

Table 5.1: The raw data for the 12 stimuli used in the LCJ analysis. The element in row j and column k is the number of times (out of 18 observations) that stimulus k was judged greater than stimulus j.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	15	18	1	18	0	7	16	0	12	4	18
2	3	0	18	0	11	0	0	15	0	10	0	15
3	0	0	0	0	0	0	0	3	0	1	0	4
4	17	18	18	0	17	0	11	18	0	18	7	18
5	0	7	18	1	0	0	0	13	0	1	0	13
6	18	18	18	18	18	0	18	18	8	17	18	18
7	11	18	18	7	18	0	0	17	0	15	3	18
8	2	3	15	0	5	0	1	0	0	0	1	11
9	18	18	18	18	18	10	18	18	0	18	18	18
10	6	8	17	0	17	1	3	18	0	0	0	17
11	14	18	18	11	18	0	15	17	0	18	0	17
12	0	3	14	0	5	0	0	7	0	1	1	0

In order to estimate the appropriate location on the "psychological continuum" for each stimulus, we use the approximate distance from the stimulus to each of its overlapping neighbors. This "psychological distance" between two stimuli can be estimated by the proportion of times each was chosen over the other. But a proportion of either 0.0 or 1.0 gives us no information about the this relative distance. In that case, we can think of the normal curves representing the distributions of discriminal processes along the psychological continuum for the two stimuli as having zero (or practically zero) overlap. Hence, stimuli #6 and #9 have practically zero overlap with the other 10 stimuli, so we can't accurately estimate their locations on the continuum. It was impossible to foresee this situation when the stimuli for the experiment were chosen.

Table 5.2: Computed values for selected metrics and their correlation coefficients with respect to the LCJ scale values.

Stimulus #	LCJ Scale	ΔT	ΔT_{rss}	ΔT_{rss4}	ΔT_{suma}	ΔT_{Sum}	Doyle
1	1.1817	0.0000	24.0000	48.0000	16.0000	24.0000	16.0000
2	2.2872	8.0000	25.2982	48.6621	24.0000	32.0000	17.8885
3	4.1282	16.0000	28.8444	50.5964	32.0000	40.0000	22.6274
4	0.1075	4.0000	20.3961	40.1995	16.0000	24.0000	12.6491
5	2.9032	12.0000	23.3238	41.7612	24.0000	32.0000	16.9706
7	0.7500	8.0000	17.8885	32.9848	16.0000	24.0000	11.3137
8	3.4432	16.0000	22.6274	35.7771	24.0000	32.0000	17.8885
10	1.8492	12.0000	16.9706	26.8328	16.0000	24.0000	12.6491
11	0.0000	8.0000	11.3137	17.8885	8.0000	16.0000	8.0000
12	3.5732	16.0000	17.8885	22.6274	16.0000	24.0000	16.0000
Corr. Coeff. (r) :		0.5979	0.7996	0.2766	0.7815	0.7815	0.8532

There was, however, plenty of "perceptual overlap" among the other 10 stimuli, so the LCJ analysis was successful when only these 10 were analyzed. In all of the analysis that follows, stimulus #6 and stimulus #9 have been dropped and only the other 10 stimuli are used. The second column of Table 5.2 shows the LCJ scale values that resulted from analysis of the data for these 10 stimuli. The scale values only describe **relative** positions among the stimuli on the scale of perceptual distinctness of their four-bar patterns. The least distinct stimulus, stimulus #11, is assigned a scale value of zero and the scale is constructed from that point.

5.5 Evaluation of Selected Target Signature Metrics

The last six columns of Table 5.2 show the values computed for several metrics which attempt to model the visibility of a target against a background. At the

bottom of each of these columns is shown the correlation coefficient of the metric values in that column with respect to the LCJ psychological scale values in column 2. This correlation coefficient (r) is defined for 2 vectors X and Y as:

$$r = \frac{Cov(X,Y)}{\sqrt{\sigma_x^2 \sigma_y^2}} \tag{5.1}$$

where Cov(X,Y) denotes the covariance of X and Y.

The six metrics evaluated here were some of the most promising of a larger set that was evaluated in a perceptual experiment conducted at the Army's Night Vision and Electronic Sensors Directorate in February, 1993 [OWD93]. The metrics used here were Delta T, RSS Delta T, RSS4 Delta T, Suma Delta T, Sum Delta T, and the Doyle metric. These metrics are defined as follows:

$$\Delta T = |\mu_t - \mu_b| \tag{5.2}$$

$$\Delta T_{rss} = \sqrt{(\mu_t - \mu_b)^2 + \sigma_t^2} \tag{5.3}$$

$$\Delta T_{rss4} = \sqrt{(\mu_t - \mu_b)^2 + 4\sigma_t^2} \tag{5.4}$$

$$\Delta T_{suma} = |\mu_t - \mu_b| + |\sigma_t - \sigma_b| \tag{5.5}$$

$$\Delta T_{sum} = |\mu_t - \mu_b| + \sigma_t \tag{5.6}$$

$$Doyle = \sqrt{(\mu_t - \mu_b)^2 + (\sigma_t - \sigma_b)^2} \tag{5.7}$$

where

- μ_t and μ_b are the computed mean gray level of the target and background fields,
- σ_t and σ_b are the computed standard deviations of the target and background fields.

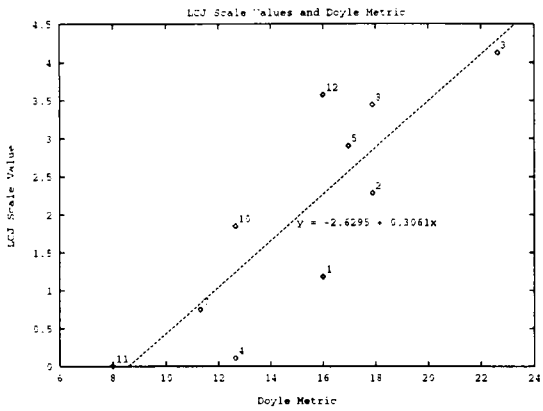
Table 5.3: Computed values for some more recently developed metrics and their correlation coefficients with respect to the LCJ scale values.

Stimulus #	LCJ Scale	Mod. Doyle	Log Doyle	Hybrid Doyle	REDT
1	1.1817	10.2702	0.0759	0.0767	5.9150
2	2.2872	13.0183	0.0971	0.0978	9.3920
3	4.1282	19.0126	0.1401	0.1406	13.8820
4	0.1075	8.6793	0.0704	0.0653	5.6640
5	2.9032	14.2594	0.1097	0.1065	10.3940
7	0.7500	9.5063	0.0773	0.0717	6.6740
8	3.4432	16.8038	0.1271	0.1239	12.8820
10	1.8492	12.2716	0.0939	0.0916	9.5460
11	0.0000	8.0000	0.0606	0.0606	6.2120
12	3.5732	16.0000	0.1178	0.1178	13.7360
Corr. Coeff. (r) :		0.9878	0.9813	0.9881	0.9634

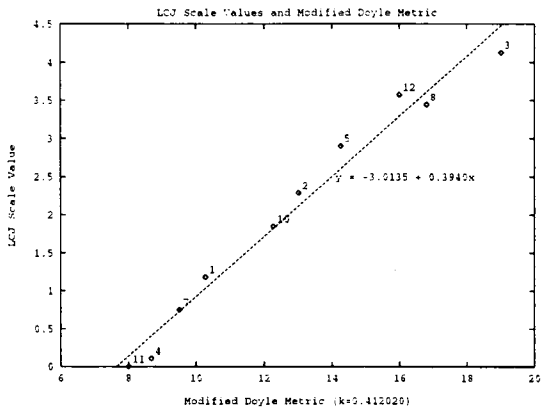
Among these six metrics, the Doyle metric yielded the highest correlation with the LCJ scale values (r=0.8532). Figure 5.4(a) shows a scatter plot of the ten LCJ scale values against the values of the Doyle metric for the ten stimuli. Also shown is the best fitting line from a linear regression analysis.

5.6 Doyle Metric Derivatives

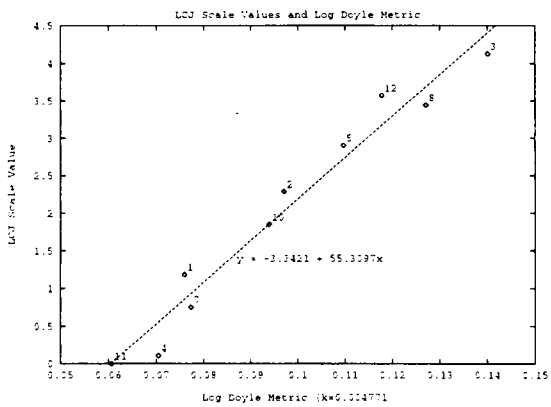
We can also consider some modifications to the Doyle metric, to see if improved correlations could be obtained between such a modified Doyle metric and the experimentally derived scale values. Table 5.3 summarizes the results. Figure 5.4(b-d) shows scatter plots of the values of three such modified Doyle metrics versus the LCJ scale values for the ten stimuli that were analyzed. This section will address the motivation behind the development of these new metrics.



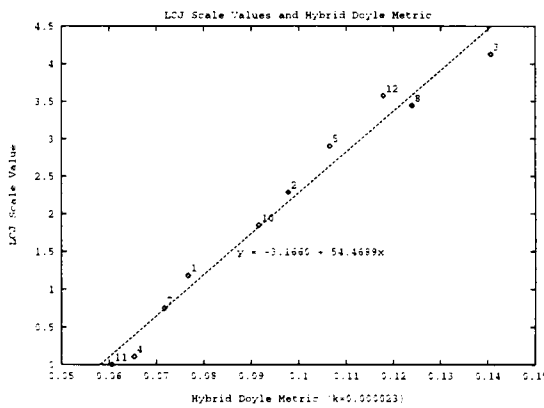
(a) LCJ scale values and the Doyle metric.



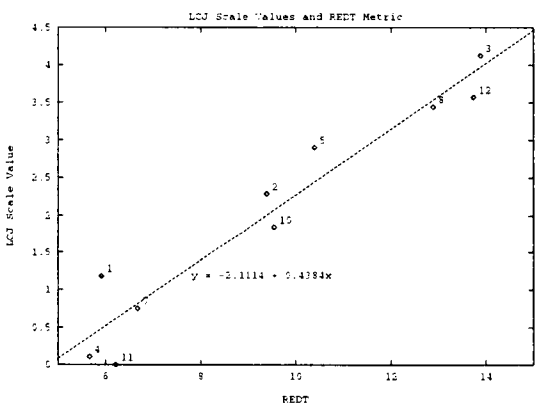
(b) LCJ scales and Modified Doyle.



(c) LCJ scales and Log Doyle.



(d) LCJ scales and Hybrid Doyle.



(e) LCJ scales and REDT metric.

Figure 5.4: The stimuli plotted with their LCJ scale values and computed values for various metrics.

Examining Equation 5.7, one will see that the Doyle metric is essentially the Euclidean distance between the target and the background, when viewed as points in the μ-σ plane (as in Figure 5.2). If this model is accurate, then there is a psychophysical equivalence between distance along the μ-axis and distance along the σ-axis. However, what if perceptual distinctness is better modeled if one of the axes is scaled by a constant factor? This would yield the Modified Doyle metric [MWD93]:

$$Doyle_{Mod} = \sqrt{(\mu_t - \mu_b)^2 + k(\sigma_t - \sigma_b)^2} \tag{5.8}$$

An iterative scheme was used to find the value for k in Equation 5.8 that would yield the maximum value for the correlation coefficient when $Doyle_{Mod}$ is compared to the LCJ scale values already determined. It was found that a value of k=0.412020 resulted in a maximum correlation of r=0.987843, a significant improvement. The values listed in the "Mod. Doyle" column of Table 5.3 were calculated using k=0.412020.

Both the Doyle and the Modified Doyle metrics are functions of the absolute differences between the μ and σ parameters of the target and background; but what if the human visual system is better modeled by **ratios** of the parameters? In this case, the human visual response would be proportional to the logarithms of the parameter values, yielding the new Log Doyle metric:

$$Doyle_{Log} = \sqrt{(\log \mu_t - \log \mu_b)^2 + k(\log \sigma_t - \log \sigma_b)^2} \tag{5.9}$$

Once again, an iterative scheme was used to find a value for k to maximize the correlation coefficient r. It was found that a value of k=0.004771 resulted in a maximum correlation of r=0.981340. The values listed in the "Log Doyle" column

of Table 5.3 are for k=0.004771. The correlation coefficient using the Log Doyle metric was slightly lower than that using the Modified Doyle metric. This tends to indicate that using ratios of the two parameters is less effective than using absolute differences of the two parameters.

Among the remaining possibilities, it may be that absolute differences best model the response of the human visual system to one of the parameters while ratios are best for the other. For example, if one assumes that ratios are the best model for contrast (μ) while absolute differences are the best model for variability (σ), one obtains the new Hybrid Doyle metric:

$$Doyle_{Hybrid} = \sqrt{(\log \mu_t - \log \mu_b)^2 + k(\sigma_t - \sigma_b)^2} \tag{5.10}$$

The values for "Hybrid Doyle" in Table 5.3 were obtained using this equation. This time the iterative scheme arrived at a solution of k=0.000023, for a resulting maximum correlation of r=0.988075. This is slightly higher than the correlation for the Modified Doyle metric.

5.7 A Discrimination Threshold Metric

Another first-order signature metric has been proposed by McManamey et al. [MWD93, MW94]. The value of this metric (called Response Equivalent Delta T, or REDT) is defined to be the minimum distance between a threshold response surface and a signature point defined by the target and background parameters. In [MW94], an approximate equation of this response surface is given as

$$-0.065\sigma_B^2 + 0.174\sigma_B\sigma_T - 0.086\sigma_T^2 = \kappa \tag{5.11}$$

where σ_B is the standard deviation of the background, σ_T is the standard deviation of the target, and κ is the contrast as defined by the ΔT metric in Equation 5.2. While there is not believed to be any closed-form solution for this distance, it can be found by iterative methods as discussed in [MW94]. The last column of Table 5.3 gives values of REDT that have been provided for each of the stimuli, and Figure 5.4(e) shows a scatter plot of the ten LCJ scale values against the values of the REDT metric for the ten stimuli. The correlation between the LCJ scale values and REDT was r=0.9634.

5.8 Conclusions

From the success of the pilot experiment, it is evident that the experimental procedure consisting of psychophysical paired-comparison judgments, psychological scaling based on the LCJ, and statistical correlation, is a powerful tool for investigating human perception of visual stimuli. The procedure is especially suited for the development and evaluation of quantitative models and metrics for target distinctness, human texture discrimination, and image clutter. XPET, the computer environment created to automate the presentation of stimuli and recording of data, has proved to be an effective testbed for conducting such experiments. The pilot experiment has also shown that "perceptual overlap" is an important consideration when selecting the images to be used as stimuli.

The analysis of the data from the pilot experiment has affirmed that the Doyle derivatives and REDT are promising first-order image metrics. The especially high correlation values (as high as 0.9881) for the Doyle derivatives should be regarded

with caution, since these metrics have constants that were "tuned" to the data. Further studies must be conducted to determine if these metrics are robust enough to be useful. The REDT metric, however, was not tuned to the data, so it may prove to be more broadly applicable. It is important to realize that the first-order measures evaluated in this study can only be applied directly to cases where image texture is not a variable, which is typically not true of real scenery. Future experiments will necessarily include target and background patterns with controlled second-order statistics, so that the role of image texture in target distinctness can be studied.

CHAPTER 6

Second-order Image Statistics and Human Perception

The purpose of these next three psychophysical experiments was to verify that our texture synthesis process, described earlier in Section 3.1, produces patterns that continuously approach a more visually pleasing result. Specifically, we wished to test whether GLC error (such as in Equations 3.1 and 3.2) that is minimized during texture synthesis corresponds closely to the texture pattern's perceptual distance from the desired result.

6.1 Creation of the Image Stimuli

In each of the three experiments, a different type of texture pattern was used as the background in the images that served as stimuli. To generate the stimuli to be used in the experiments, target patterns were synthesized in a 96x96 square in the center of each of the three 256x256 background textures, according to the binary mask shown in Figure 6.1. Ten different target patterns were created for each of the three experiments. The stimuli used in the pigskin, raffia, and evergreen experiments are shown in Figures 6.2, 6.3, and 6.4, respectively. The procedure used to generate the image stimuli was as follows. First, the GLC model was computed from each original background texture. All possible displacements of up to $nx = ny = 8$ pixels were considered, and the images were quantized to 8 gray levels. This results in a

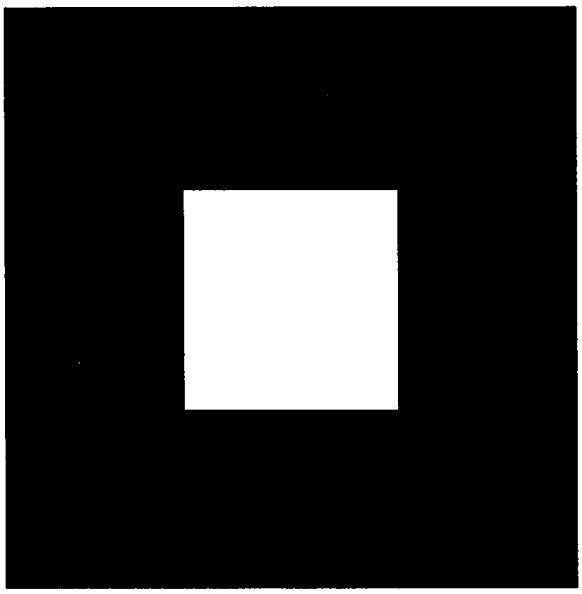
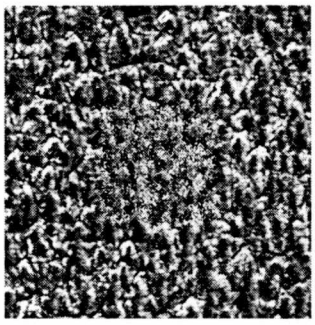
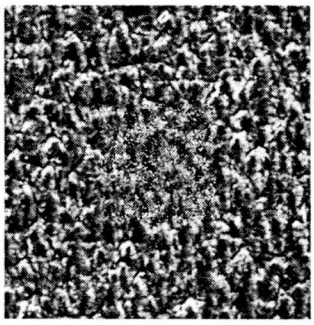


Figure 6.1: The binary mask showing the relative location of the 96x96 target field within the 256x256 background textures.

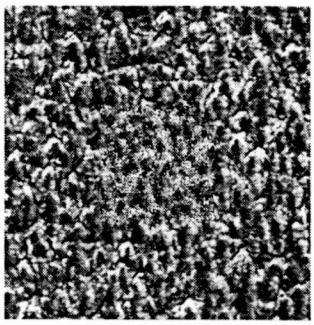
GLC model consisting of $(2 \times nx \times ny) + nx + ny = 144$ 8×8 matrices. Then, using the spin-exchange texture synthesis method with boundary blending that was described previously in Section 3.1, a new pattern was synthesized in the 96×96 target area. In each case, the GLC model used to constrain the synthesis of the targets was the model that was computed from the background. However, the synthesis process for each target was halted at a different point, yielding targets that look similar to the background but differing by various amounts in their GLC joint probabilities. In addition, the random seed used as input to the synthesis procedure was different for every stimulus. This random seed is used in generating the initial distribution of uncorrelated pixel intensities, with which the spin-exchange process then begins to modify. Different random seeds were used to ensure that each synthesized pattern



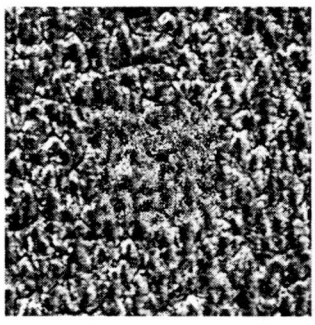
Stimulus # 1



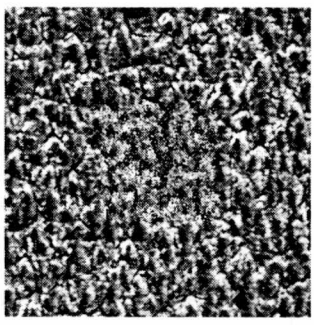
Stimulus # 2



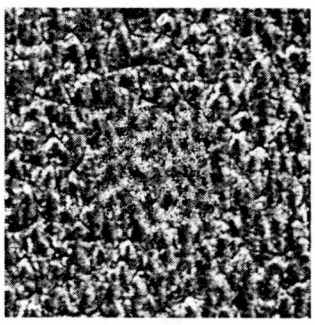
Stimulus # 3



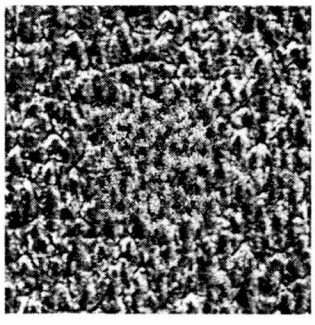
Stimulus # 4



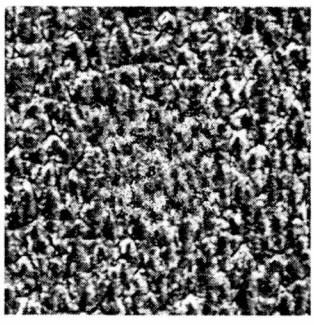
Stimulus # 5



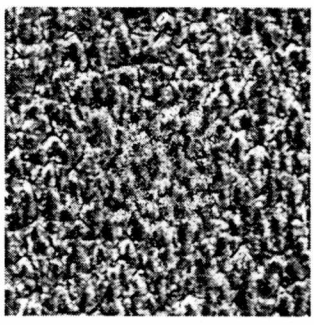
Stimulus # 6



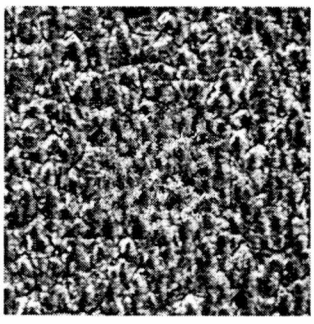
Stimulus # 7



Stimulus # 8

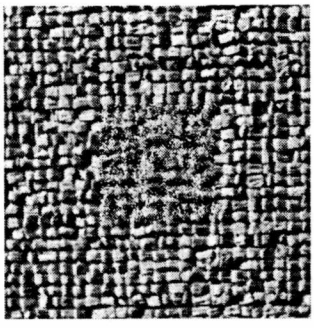


Stimulus # 9

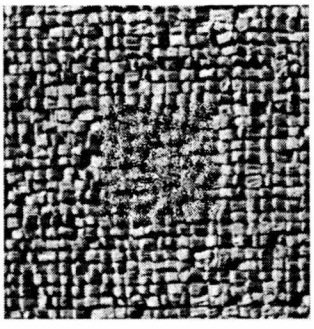


Stimulus # 10

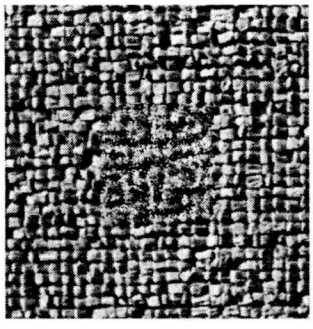
Figure 6.2: The stimuli used in the pigskin experiment.



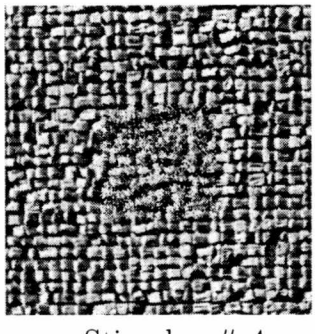
Stimulus # 1



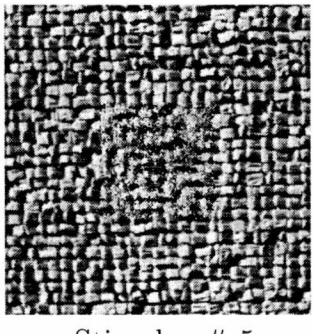
Stimulus # 2



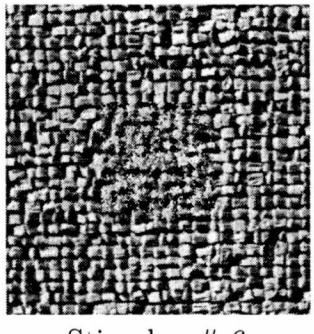
Stimulus # 3



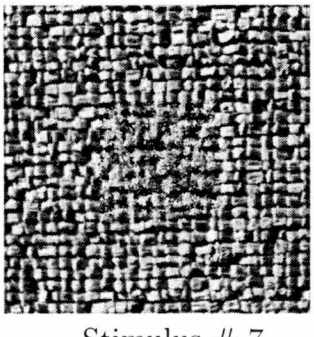
Stimulus # 4



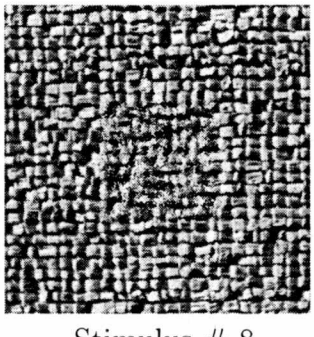
Stimulus # 5



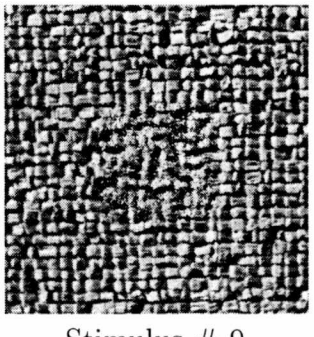
Stimulus # 6



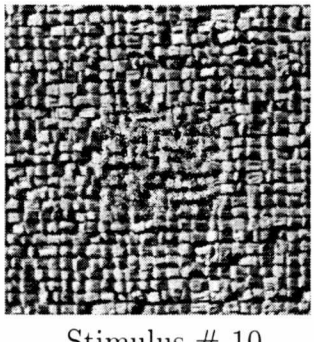
Stimulus # 7



Stimulus # 8

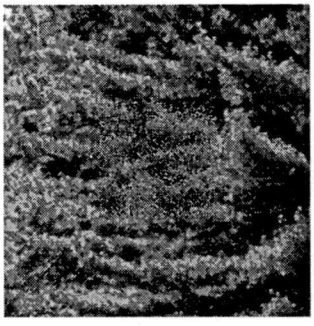


Stimulus # 9

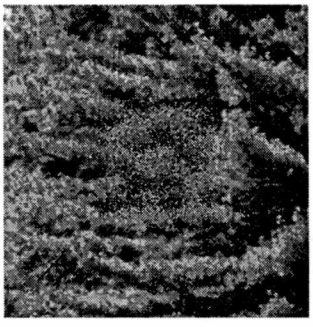


Stimulus # 10

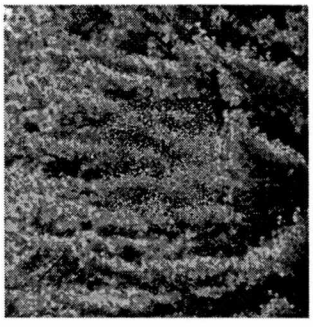
Figure 6.3: The stimuli used in the raffia experiment.



Stimulus # 1



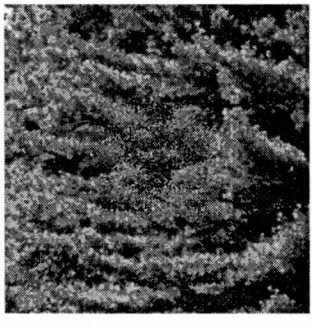
Stimulus # 2



Stimulus # 3



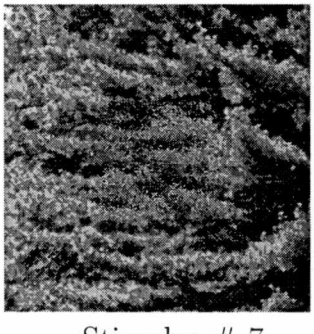
Stimulus # 4



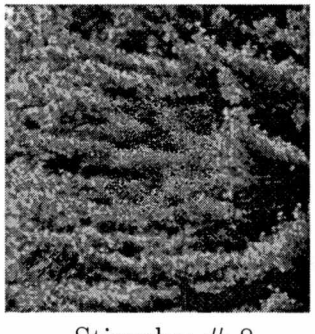
Stimulus # 5



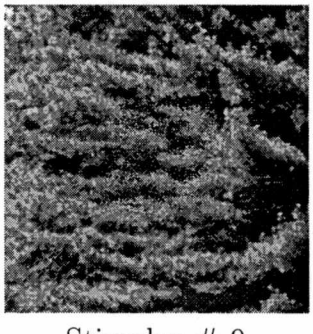
Stimulus # 6



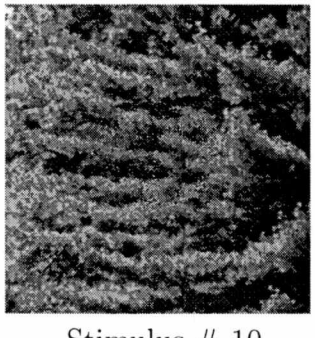
Stimulus # 7



Stimulus # 8



Stimulus # 9



Stimulus # 10

Figure 6.4: The stimuli used in the evergreen experiment.

is a different realization of the desired texture, rather than the same exact pattern only at different points in the synthesis process.

6.2 Design and Conduct of the Experiments

A separate paired-comparison experiment was conducted using the stimuli created for each of the three background textures. In each of these experiments there were $n = 10$ stimuli, yielding a total of $\frac{n(n-1)}{2} = \frac{(10)(9)}{2} = 45$ possible pairs of stimuli. For each paired comparison, the human observer was required to choose the image in which he thought the target pattern is more distinct from its background. Thus, the attribute being judged in this experiment is the "perceptual distinctness" of the target from the background. Again, no specific instructions were given on what perceptual cues the test subjects were to use when deciding which target pattern was more distinct.

A total of 13 observers were used in the experiment. The XPET testbed was used to automate the sequential display of the image stimulus pairs and the registration and recording of subject responses. Two sessions were conducted for each observer on separate days, so there were a total of 26 observations for each pair of stimuli. The presentation order of the 45 sets of stimuli used during each observer's first session was completely random. The presentation order used during the second session was the exact opposite: the stimulus pairs were presented in reverse order, and the left-right positioning of the stimuli in each pair was flipped. Observers were instructed to position themselves so that they could comfortably see the display, which was usually about 18 inches away. There was no time limit imposed on the

observers to make their decisions, but they were urged not to study the images for more than a few seconds.

6.3 Results of the Experiments

The data that was collected from each experiment consists of the number of times, out of 26 observations, that each of the stimuli was judged greater than each of the other stimuli, i.e., its target pattern was judged more distinct. This raw data was used in an LCJ analysis to solve for the scale values for each stimulus, as explained previously in Section 4.1.3. The scale values only describe **relative** positions among the stimuli, on the scale of perceptual distinctness of their target patterns. The least distinct stimulus in each case is assigned a scale value of zero and the scale is constructed from that point. Table 6.1 shows the LCJ scale values that resulted from this analysis on the 10 stimuli in each of the three experiments.

These psychological scale values were compared to the ACE measure which was used in the synthesis of the stimuli, and also to five other cooccurrence error measures. The six measures that were compared to the scales will be referred to as ACE (average cooccurrence error), WACE (weighted average cooccurrence error), AWCE (average weighted cooccurrence error), ASCE (average squared cooccurrence error), WASCE (weighted average squared cooccurrence error), and AWSCE (average weighted squared cooccurrence error). The equations for these six computational measures are given in Equation 6.1 through Equation 6.6. (ACE is given again for

Table 6.1: The LCJ scale values for the 10 stimuli in each of the three texture synthesis experiments.

Stimulus #	Pigskin	Raffia	Evergreen
1	4.35	3.37	4.16
2	3.17	2.94	4.55
3	2.91	3.24	3.82
4	2.80	2.09	3.21
5	2.30	1.91	2.43
6	2.20	1.67	2.95
7	1.22	1.05	2.30
8	1.08	0.89	1.36
9	0.57	0.81	0.92
10	0.00	0.00	0.00

comparison.)

$$ACE = \frac{1}{D}\sum_{d=1}^{D}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}|P_t(i,j|d) - P_b(i,j|d)| \tag{6.1}$$

$$WACE = \frac{1}{D}\sum_{d=1}^{D}\left[\frac{1}{\sqrt{x^2+y^2}}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}|P_t(i,j|d) - P_b(i,j|d)|\right] \tag{6.2}$$

$$AWCE = \frac{1}{D}\sum_{d=1}^{D}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{|P_t(i,j|d) - P_b(i,j|d)|}{|i-j|+1} \tag{6.3}$$

$$ASCE = \frac{1}{D}\sum_{d=1}^{D}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(P_t(i,j|d) - P_b(i,j|d))^2 \tag{6.4}$$

$$WASCE = \frac{1}{D}\sum_{d=1}^{D}\left[\frac{1}{\sqrt{x^2+y^2}}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(P_t(i,j|d) - P_b(i,j|d))^2\right] \tag{6.5}$$

$$AWSCE = \frac{1}{D}\sum_{d=1}^{D}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{(P_t(i,j|d) - P_b(i,j|d))^2}{|i-j|+1} \tag{6.6}$$

where

- $D = (2 \times nx \times ny) + nx + ny$ is the total number of displacement vectors considered,

- G is the number of possible gray levels,

- $P_t(i, j|d)$ is the joint probability of a pixel of gray level i and a pixel of gray level j given the displacement vector d=[x y] for the target pattern,

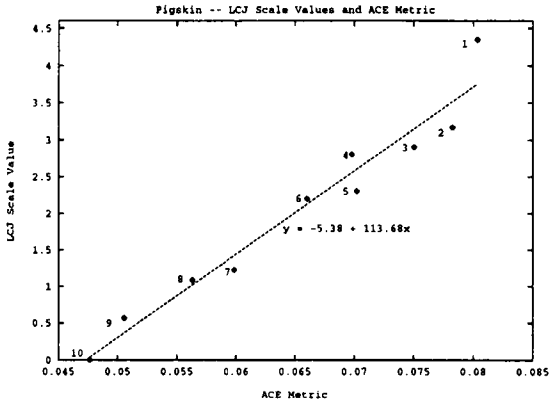
- $P_b(i, j|d)$ is the joint probability of a pixel of gray level i and a pixel of gray level j given the displacement vector d=[x y] for the background pattern.

These measures are all similar in that they compute GLC error between a target GLC model and a background GLC model in terms of the differences between corresponding matrix elements, or joint probability terms. For ACE, WACE, and AWCE, the absolute value of the differences are used; for ASCE, WASCE, and AWSCE, the differences are squared. In WACE and WASCE, the differences are weighted inversely by the Euclidean length of the displacement vector of each GLC matrix. This type of weighting considers the errors in the GLC matrices for short displacements more important than the errors for longer displacements. In AWCE and AWSCE, the differences are weighted inversely by the city-block distance of the matrix element from the diagonal of the GLC matrix. This type of weighting considers the errors in the GLC matrix elements closer to the diagonal to be more important. Since the diagonal elements represent the likelihood of nearby pixels having the same gray level, the strength of these elements represents the tile structure in the texture. These particular variations are considered due to the observations from the texture experiments discussed in Chapter 3.

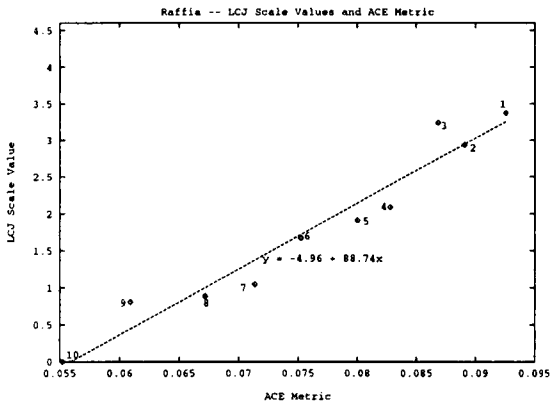
Table 6.2: The correlation coefficients for the six GLC computational measures considered, computed relative to the LCJ scale values for the texture synthesis experiments.

Measure	Pigskin	Raffia	Evergreen
ACE	0.9752	0.9712	0.9776
WACE	0.9784	0.9717	0.9817
AWCE	0.9726	0.9679	0.9802
ASCE	0.9766	0.9764	0.9628
WASCE	0.9768	0.9760	0.9743
AWSCE	0.9748	0.9732	0.9652

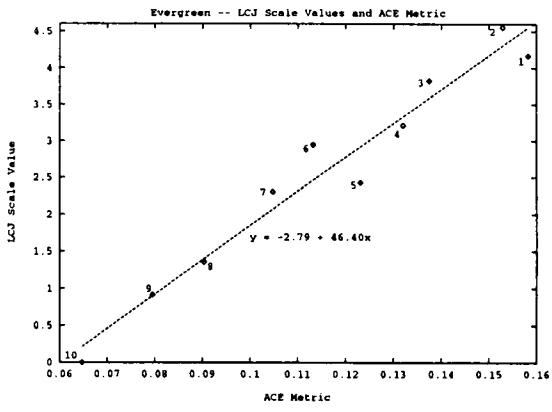
Table 6.2 shows the correlation coefficients for the six GLC computational measures considered, computed relative to the LCJ scale values. Figure 6.5 shows the 10 stimuli plotted using their LCJ scale values and the ACE measure computations for each of the three experiments. Also shown in each case is the best fitting line from a linear regression analysis. From the computed correlation values in Table 6.2, it is evident that all six metrics yielded a relatively high correlation with the empirical scale values. There is not a single metric which performed best for all three experiments. The results of the two different weighting schemes are inconclusive. In some cases our weighting methods raised the correlations slightly, while in other cases they seemed to lower them.



(a) Pigskin



(b) Raffia



(c) Evergreen

Figure 6.5: The 10 stimuli plotted using their LCJ scale values and the ACE measure computations for each of the three texture synthesis experiments.

6.4 Conclusions

These three texture synthesis experiments have demonstrated that measures of GLC error can correspond closely to differences in the visual appearance of texture patterns. This result has been shown for a disordered, an ordered, and a real-world texture pattern. However, in these experiments, the test subjects were asked only to discriminate between a texture and another pattern with similar GLC matrices. A

raffia pattern was only compared to other raffia patterns, an evergreen pattern was only compared to other evergreen patterns, etc. This means that any conclusions from these experiments about the performance of the GLC error metrics in modeling human texture discrimination can only apply to such patterns with similar GLC matrices.

CHAPTER 7

Hybrid Textures and Human Perception

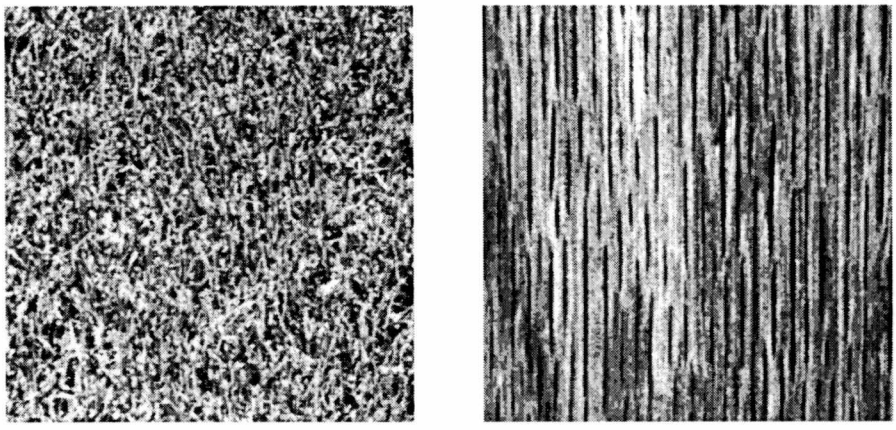
7.1 Paired Comparisons of Hybrid Textures

The purpose of this next experiment was to assess the ability of GLC-based error metrics to model the perceptual distance between textured patterns, by considering the case where the texture patterns are hybrids of two significantly different textures.

7.1.1 Creation of the Image Stimuli

The creation of the stimuli to be used in the experiment was accomplished as follows. Two different 256×256 texture patterns were selected. After experimenting with several different pairs of textures, we settled on a wood grain texture and a grass texture, both from the Brodatz album of textures [Bro66]. These two texture patterns are shown in Figure 7.1. The grass texture is considered disordered, while the wood grain texture is weakly ordered [Rao90]

These texture images were first subjected to a histogram specification, which forces them to have identical first-order statistics. This histogram specification was done so that the first-order statistics would be identical and known for all images to be studied, ensuring that any visual differences between two textures can be explained by considering their higher-order statistics only. Note that this condition was satisfied during the three synthesis error experiments without histogram speci-



(a) grass (b) wood

Figure 7.1: The original texture images for the texture hybridization experiment.

fication, since the spin-exchange synthesis method does not alter the histogram and the texture patterns from each model were only compared to other patterns from the same model. The histogram to which the two textures were specified has eight gray levels (0 to 7) and a discrete version of a normal, or Gaussian, distribution with a mean of 3.5 and a standard deviation of 1.5 gray levels. This histogram is shown in Figure 7.2, where the number of pixels (from a total of 65,536 in a 256×256 image) at each of the eight gray levels are indicated. A Gaussian histogram was again chosen over a uniform, or equal, histogram because the process of histogram equalization can force some gray level changes which change the visual appearance of the image, often creating an unnaturally high level of contrast. Images usually ap-

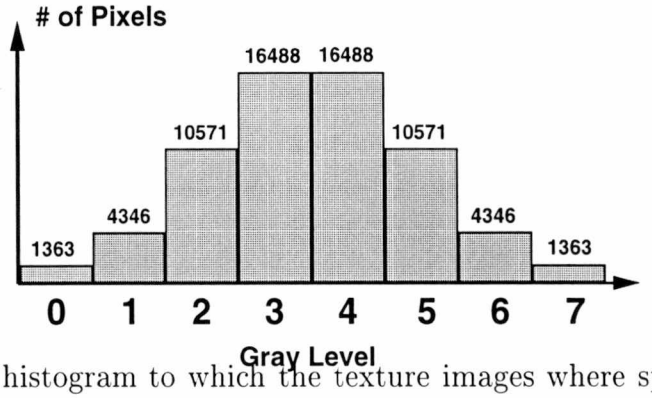


Figure 7.2: The histogram to which the texture images where specified. This histogram is normal with a mean of 3.5 and a standard deviation of 1.5 gray levels.

pear more natural when a smooth, Gaussian-shaped histogram is used. The images shown in Figure 7.1 are the results of this histogram normalization.

The GLC models were then computed from the two histogram-specified texture images, again considering all possible displacements of up to $nx = ny = 8$ pixels. From these two GLC models, nine new synthetic models were created by linearly combining the two real models using nine different weighting schemes. This linear combination was performed by simply taking each element of each GLC matrix for texture model A, multiplying it by the weight w_A, and adding it to the corresponding matrix element from texture model B multiplied by weight w_B. The weights were chosen such that they always summed to unity, meaning that the resulting synthetic model represents a proper probability distribution in that the elements in each GLC matrix also sum to unity. This combination of two real GLC models to form a synthetic model is described by Equation 7.1:

$$P_{(w_A,w_B)}(i,j|d) = w_A \cdot P_A(i,j|d) + w_B \cdot P_B(i,j|d) \qquad \forall i, \forall j, \forall d \qquad (7.1)$$

$$w_A + w_B = 1$$

A synthetic texture model created in this way from a texture A and a texture B can be described by the two weights, w_A and w_B. These models will be referred to as hybrid texture models.

Designating the grass texture model as A and the wood grain model as B, the nine hybrid texture models for our experiment were created using the weighting schemes (w_A,w_B) as follows: (0.9,0.1), (0.8,0.2), (0.7,0.3), (0.6,0.4), (0.5,0.5), (0.4,0.6), (0.3,0.7), (0.2,0.8), and (0.1,0.9). Together with the two original models, A (1,0) and B (0,1), we now have a total of eleven GLC models. Next, eleven different 256 × 256 textures were synthesized with the spin-exchange method, using the eleven GLC models as inputs. The synthesis process in every case was continued until no further improvement in the results was evident. Also, a different initial seed for the random number generator was used for each of the eleven, resulting in an unique initial spatial distribution of pixel intensities. These eleven different synthesized textures would serve as the background patterns for the experiment stimuli.

Target patterns were again synthesized in a 96 × 96 square in the center of each of the 256 × 256 background textures, according to the binary mask shown earlier in Figure 6.1. The GLC model from which the target pattern was synthesized in every case was the model computed from the original grass texture (the (1,0) model). Thus, the stimuli all had the same target texture model and only the background textures were different. The spin-exchange algorithm with blending was used for the synthesis, in an attempt to minimize the edge effects at the boundary between target and background. Once again, the synthesis process in every case was continued until

no further improvement in the results was evident, and a different initial random number seed was used for each target. The eleven stimuli used in the experiment are shown in Figure 7.3. Stimulus #1 should theoretically have no target visible, since the original grass (1,0) model was used for both target and background synthesis. Stimulus #11 should theoretically have the most visible target, since it consists of a texture synthesized from the grass (1,0) model against a background synthesized from the wood (0,1) model. The nine stimuli between these two extremes have targets synthesized from the grass (1,0) model against backgrounds synthesized from the hybrid texture models.

7.1.2 Design and Conduct of the Experiment

A paired-comparison experiment was conducted using the synthesized stimuli. Since there were $n = 11$ stimuli used in the experiment, there were a total of $\frac{n(n-1)}{2} = \frac{(11)(10)}{2} = 55$ possible pairs of stimuli. For each paired comparison, the human observer was required as before to choose the image in which he thought the target pattern is more distinct from its background. So once again, the attribute being judged in the experiment is the "perceptual distinctness" of the target from the background, but no specific instructions were given on what perceptual cues the test subjects were to use when making their decisions. The XPET testbed was again used to automate the sequential display of the image stimulus pairs and the registration and recording of subject responses. Ten observers were used, with one session for each. All other test parameters were the same as described previously for the synthesis error experiments.

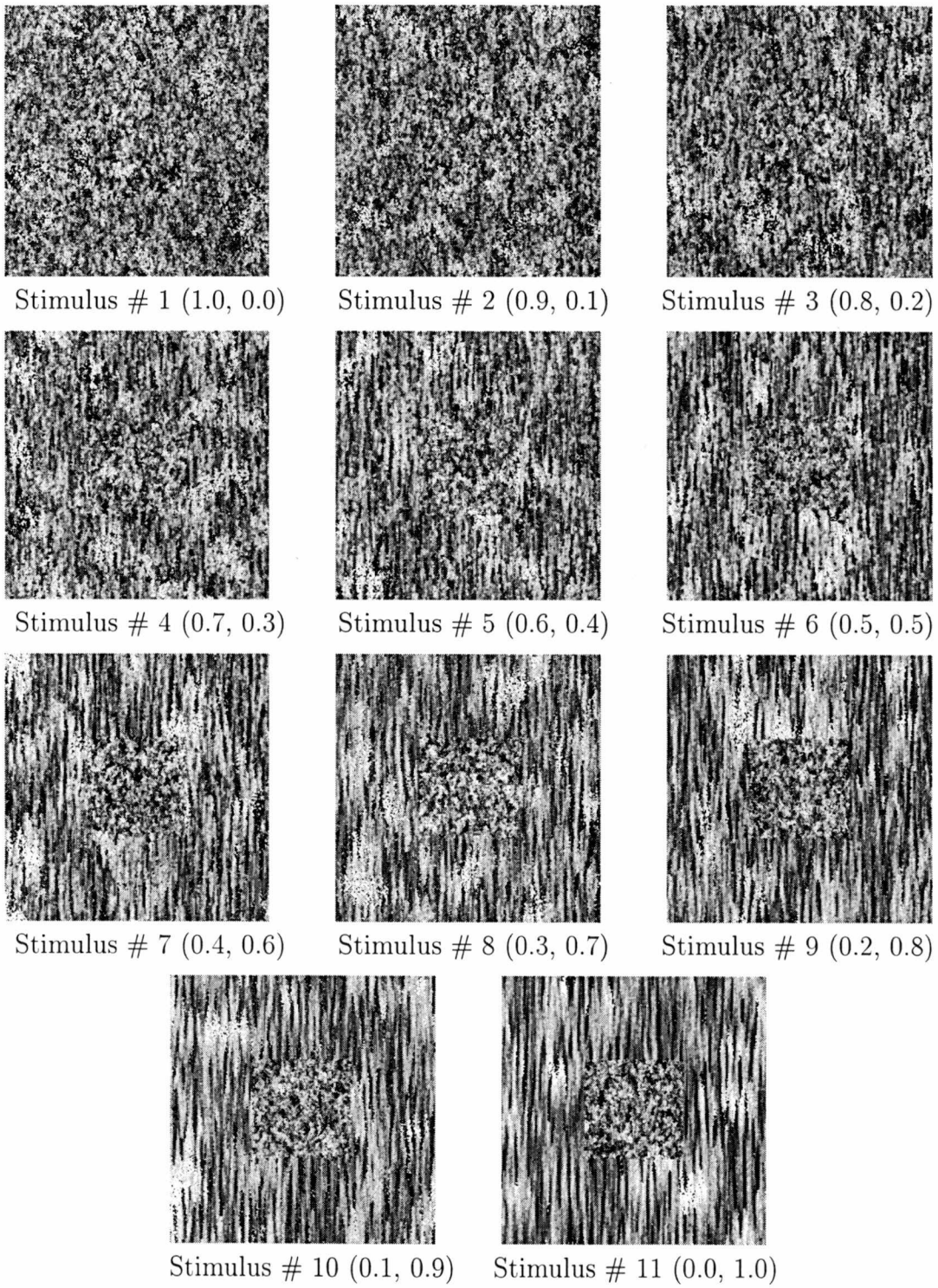


Figure 7.3: The stimuli used in the texture hybridization experiment.

Table 7.1: The raw data from the texture hybridization experiment. The element in row j and column k is the number of times that stimulus k was judged greater than stimulus j. The last line contains the sums over each column.

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	8	10	10	10	10	10	10	10	10
2	9	0	10	10	10	10	10	10	10	10	10
3	2	0	0	5	10	10	10	10	10	10	10
4	0	0	5	0	3	10	10	10	10	10	10
5	0	0	0	7	0	9	10	10	10	10	10
6	0	0	0	0	1	0	6	10	10	10	10
7	0	0	0	0	0	4	0	7	10	9	10
8	0	0	0	0	0	0	3	0	10	7	10
9	0	0	0	0	0	0	0	0	0	2	10
10	0	0	0	0	0	0	1	3	8	0	10
11	0	0	0	0	0	0	0	0	0	0	0
	11	1	23	32	34	53	60	70	88	78	100

7.1.3 Results of the Experiment

The data that was collected from the experiment consists of the number of times, out of 10 observations, that each of the stimuli was judged greater than each of the other stimuli, i.e., its target pattern was judged more distinct. The raw data for the texture hybridization experiment is given in Table 7.1. The element in row j and column k is the number of times (out of 10 observations) that stimulus k was judged greater than stimulus j. Notice that all diagonal elements are zeros, since a stimulus was never compared to itself. The last line of the table contains the sums over each column.

This raw data was used in an LCJ analysis to solve for the scale values for each stimulus, as explained previously in Section 4.1.3. The scale values resulting from

Table 7.2: The traditional and adjusted scale values for the 11 stimuli in the texture hybridization experiment.

Stimulus #	Target Model	Bkgnd. Model	Traditional Scale	Adjusted Scale
1	(1,0)	(1,0)	1.29	1.29
2	(1,0)	(.9,.1)	0.00	0.00
3	(1,0)	(.8,.2)	2.14	2.14
4	(1,0)	(.7,.3)	2.14	2.14
5	(1,0)	(.6,.4)	1.61	2.14
6	(1,0)	(.5,.5)	2.90	3.43
7	(1,0)	(.4,.6)	3.16	3.69
8	(1,0)	(.3,.7)	3.77	4.30
9	(1,0)	(.2,.8)	5.22	5.75
10	(1,0)	(.1,.9)	4.37	4.90
11	(1,0)	(0,1)	∞	∞

this traditional method are given in the fourth column of Table 7.2. Note that stimulus #11 could not be assigned a relative scale because it was chosen over every other stimulus by every observer. In this case, we can think of the normal curve representing the distribution of discriminal processes for stimulus #11 as being so far separated from the distributions of the other stimuli as to have practically zero overlap, so we can't accurately estimate its relative location on the continuum. It was impossible to foresee this situation when the stimuli for the experiment were created. There was, however, plenty of "perceptual overlap" among the other 10 stimuli, so the LCJ analysis was successful when only these 10 were analyzed.

Theoretically, stimulus #1 should have the least distinct target, since the synthesis of both the target and the background was constrained by the same GLC model. And if the same types of GLC-based errors used previously can adequately model human perception of the hybrid textures, then stimuli #2 through #11 should be

in exact order of increasing target distinctness. However, the rank order of the traditional scales does not follow this trend. For instance, stimulus #1 was judged greater than stimulus #2. We can see from the column totals in Table 7.1 that the target in stimulus #2 was judged more distinct only one time (out of a total of 100 comparisons in the experiment), while the target in stimulus #1 was judged more distinct eleven times, including nine times over stimulus #2. Observing stimulus #1 and stimulus #2 closely (see Figure 7.3), we see that in stimulus #1, there is a distinct dark area very close to where the lower right-hand corner of the target box would be located. In stimulus #2 there are no such features. Since the test subject knows the exact location of the square target box during the experiment, it is very likely that he will look to those areas near the border of target and background to make his decision about the distinctness of the target. So the distinct dark area near the lower right-hand corner of the target box could be the reason why the test subjects chose stimulus #1 over stimulus #2. Despite the blending which minimizes the edge effects between target and background, the randomness of the synthesis process will occasionally result in a pattern with such features.

A more disturbing deviation from the expected rank order is that stimulus #5 is ranked below stimuli #3 and #4, which are ranked exactly equal. The reason that this is more disturbing is that from observing the column totals in Table 7.1, it seems that the test subjects generally ranked these in the expected relative order. Stimulus #3 was picked a total of 23 times, stimulus #4 received 32 "votes," while stimulus #5 received 34. So why did our method for estimating the LCJ scale values place stimulus #5 **below** stimuli #3 and #4? The reason is related to the peculiar,

but not necessarily erroneous, way the test subjects judged this particular trio of stimuli. From Table 7.1, we can see that the ten test subjects split their judgments of stimuli #3 and #4 evenly: each was chosen over the other exactly five times. We also see that stimulus #5 was chosen over stimulus #3 all ten times, but stimulus #4 was chosen over stimulus #5 seven out of ten times. The problem is again that it is impossible to convert a proportion of either 0.0 or 1.0 into a corresponding unit normal deviate. The traditional method of solving for the scale values essentially ignores the data in cases of proportions of 0.0 or 1.0. So for our experiment, the fact that stimulus #5 was chosen over stimulus #3 all ten times has no bearing on the final scales that are produced. This explains why stimulus #5 was placed **below** stimuli #3 and #4, even though it was chosen more total times than either of them.

Has our scaling method failed us? In general, it is not possible to produce a set of numerical scale values for these three stimuli that will simultaneously satisfy the conditions $3 = 4$, $5 > 3$, and $4 > 5$. The seemingly paradoxical data from the experiment could mean either that the test subjects were not consistent, there were not enough test subjects to comprise an adequate statistical sample, or simply that the perceptual cues used by the test subjects to make their judgments varied for different pairs of stimuli. The last of these possibilities could be the result again of the randomness of the synthesis process, which allows the formation of images which have individual peculiarities that could be used as perceptual cues.

Regardless of the cause of this apparent inconsistency, we still need to estimate a set of numerical scale values which reflect the data as closely as possible. We would prefer the rank order to correspond closely to the column totals in Table 7.1. In a

sense, the paired-comparison experiment is similar to a "round robin" format for an athletic contest, where each contestant (or team) is matched in turn against every other contestant. In this format, the final win-lose records are a good indication of relative strength. For our experiment, we could first establish the rank order of the stimuli using only the column totals in Table 7.1. Then, we could assign numerical scales to the stimuli using the appropriate normal deviates and the same LCJ theory as before, but with the requirement that the established rank order not be modified. When an apparently inconsistent set of data is encountered, such as with stimuli #3, #4, and #5 in our experiment, we should probably assign them all a numerical scale that places the group a proper distance from its closest neighbors, attempting no relative scaling **among** the inconsistent stimuli. Using this method, the scales in the last column of Table 7.2 are obtained. Comparing the traditional and adjusted scale values in Table 7.2, we see that the only difference between the two scalings is that the scale for stimulus #5 has been increased by 0.53 so that it is equivalent to the scales for stimuli #3 and #4, and the stimuli above this point have also been increased by 0.53 to maintain the same distance between stimulus #5 and stimulus #6. Stimuli #3, #4, and #5 are assigned the same scale because it is impossible to estimate their relative strengths due to the inconsistency in the data.

The same six computational measures based on GLC models used previously (Equation 6.1 through Equation 6.6) were compared to both the traditional and the adjusted LCJ scalings of the stimuli. The GLC models used in calculating the metrics for each stimulus were not the ones that were used to direct the synthesis process, but rather ones computed directly from the targets and backgrounds after

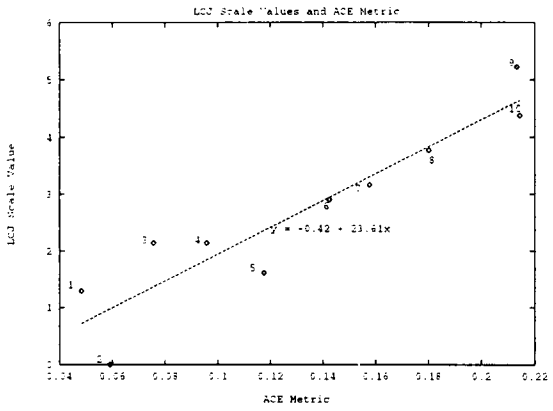
Table 7.3: The correlation coefficients for the six GLC computational measures considered, computed relative to the traditional and the adjusted LCJ scale values for the texture hybridizations experiment.

Measure	Traditional	Adjusted
ACE	0.9285	0.9546
WACE	0.9257	0.9537
AWCE	0.9280	0.9546
ASCE	0.9186	0.9350
WASCE	0.9163	0.9342
AWSCE	0.9166	0.9334

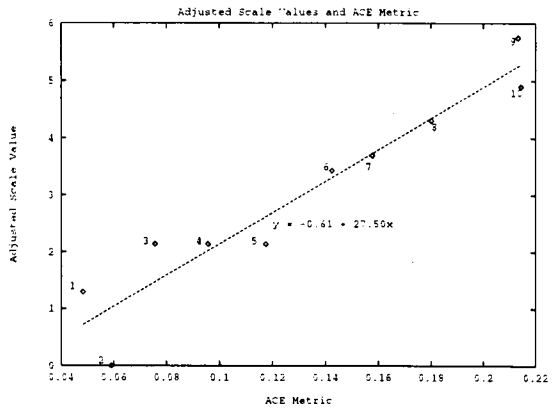
synthesis. Table 7.3 shows the correlation coefficients for the six GLC computational measures considered, computed relative to the traditional and the adjusted LCJ scale values. For these computations, stimulus #11 was not considered. Figure 7.4 shows the 10 finite stimuli plotted using the ACE metric computations and both the traditional and adjusted LCJ scale values. Also shown in each case is the best fitting line from a linear regression analysis.

7.1.4 Conclusions

We can see from Table 7.3 that the correlation coefficients for the adjusted scales are a bit higher than those for the traditional scales. We also see that the three metrics in which the differences between corresponding matrix elements are not squared (ACE, WACE, and AWCE) seem to be the best performers. Although these correlation values are not quite as high as the ones obtained in the synthesis



(a) ACE and Traditional Scales



(b) ACE and Adjusted Scales

Figure 7.4: The 10 finite stimuli in the texture hybridization experiment plotted using the ACE metric computations and both the traditional and adjusted LCJ scale values.

error experiments, they do indicate that the GLC error metrics are indeed capable of modeling human perception of textures with **different** GLC models.

7.2 Triadic Comparisons of Hybrid Textures

During the previous experiment, synthetic GLC texture models were created by linearly combining the GLC models computed from two actual texture images, a grass texture and a wood grain texture. By using different weighting schemes for the two models, we were able to create "hybrid" textures with various levels of grass and wood grain characteristics. Since each stimulus in the experiment had the grass texture model as its target pattern and only the background texture was varied, we call this a **common-target** experiment. Thus, we were investigating the relative

perceptual "distance" from each of the eleven textures to the common grass texture.

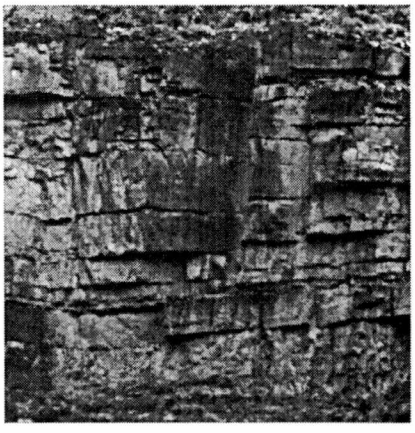
7.2.1 Creation of the Image Stimuli

For this experiment, we selected three examples of homogeneous, natural textures. These examples were extracted from natural scenes available as part of the VisTex image database available from the MIT Media Lab (http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html). Gray-level versions of the three original uncropped images are shown in Figure 7.5. From each these images, a 256x256 portion containing a relatively homogeneous texture pattern was extracted, and then these cropped images were subjected to a histogram specification, which forces them to have identical first-order statistics. This again ensures that any visual differences between the two textures can be explained by considering their second-order statistics only. The histogram to which the two textures were specified has eight gray levels (0 to 7) and a normal distribution with a mean of 3.5 and a standard deviation of 1.5 gray levels. This histogram is the same as was used during the previous experiment, and was illustrated in Figure 7.2. The resulting 256x256 images are shown in Figure 7.6.

The GLC models were then computed from the three histogram-normalized texture images, considering all possible displacements of up to 8 pixels in both vertical and horizontal directions ($nx = ny = 8$). From these three GLC models, four new synthetic models were created by linearly combining the three real models in various ways. The scheme that was used in creating the synthetic models from the real models is illustrated in Figure 7.7. In this diagram, texture models #1, #2,



(a) "Leaves.0007"



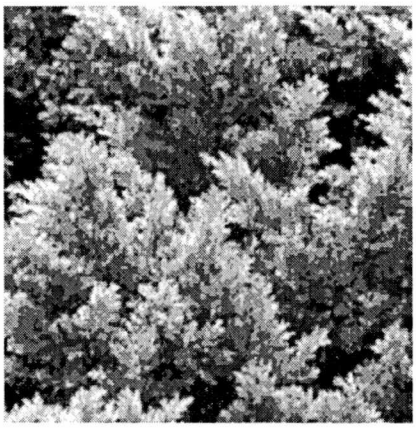
(b) "Stone.0003"



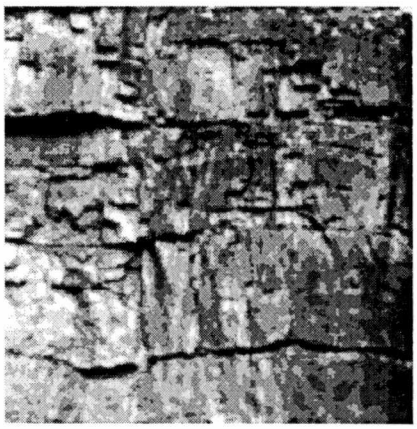
(c) "Water.0000"

Figure 7.5: The three images selected from the VisTex database.

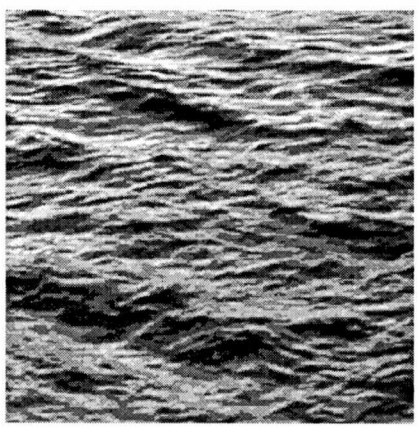
and #3 are the three models that were computed from the real textures. Texture models #4, #5, and #6, the ones along the sides of the triangle shape, were each created by linearly combining the two real models at the neighboring corners. This linear combination method is the same as used during the previous hybrid texture experiment, where each element of the synthetic texture model is the weighted sum of each of the corresponding elements of the contributing real models. In these cases,



(a) Leaves



(b) Stone



(c) Water

Figure 7.6: The three 256x256 texture samples after a histogram normalization.

each of the two corners were weighted with $\frac{1}{2}$ weight. Texture model #7 was created by combining all three of the models, each with $\frac{1}{3}$ weight.

The relative spatial locations of the seven texture models in the diagram of Figure 7.7 are used only to illustrate the scheme for creating the four synthetic models. This diagram does not necessarily represent the **perceptual** relationships that may exist between any patterns that could be synthesized using the models.

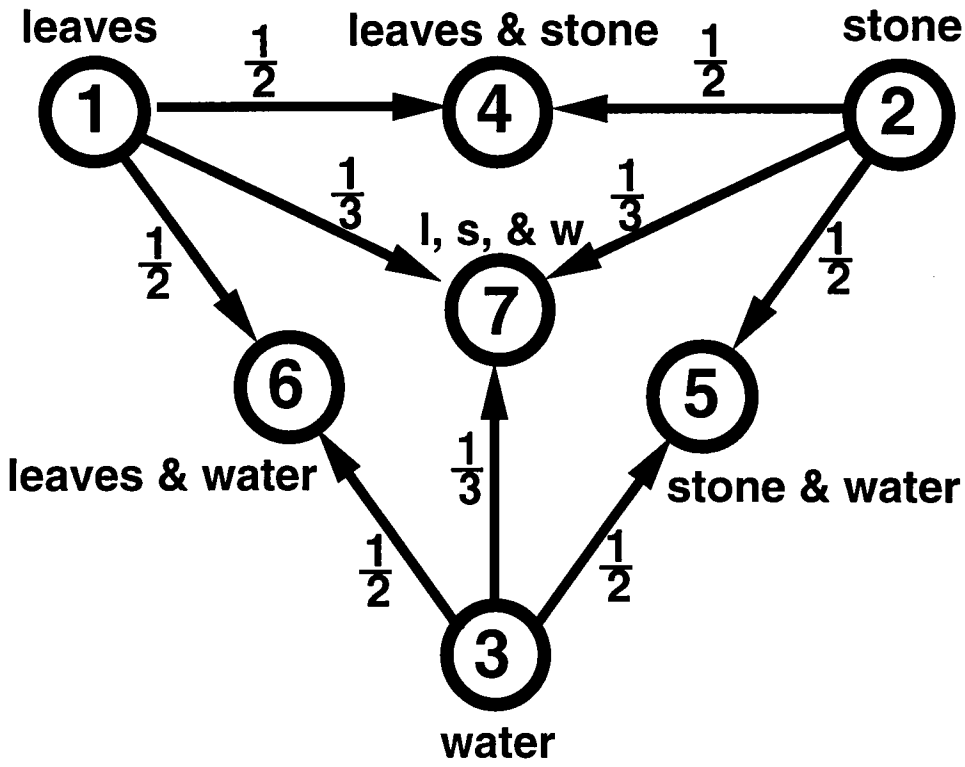


Figure 7.7: The scheme that was used in creating the synthetic models from the real models. The texture models on the corners of the triangle shape were computed from the real textures. Texture models #4, #5, and #6 were created by combining the two real models at the neighboring corners and model #7 was created by combining all three.

In fact, it is the purpose of our psychophysical experiment to study the perceptual **distance** from each of the seven textures to each of the others.

In order to utilize the complete method of triads, we will need to have two images prepared for each possible pair of texture models. For example, for two textures A and B, we will need an image where A is the target pattern against B and also one where B is against A. With $n = 7$ texture models, there are $\frac{n(n-1)}{2} = \frac{(7)(6)}{2} = 21$

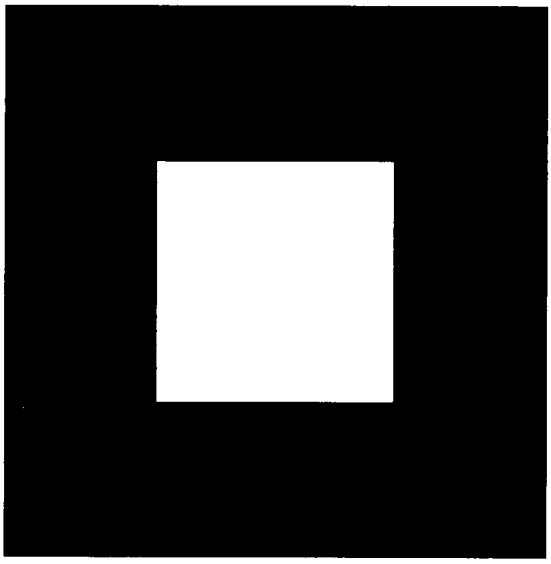


Figure 7.8: The binary mask showing the relative location of the 224 × 224 target field within the 512 × 512 background textures.

possible pairs of texture models. Thus, we need a total of 42 images prepared.

The stimulus creation started by synthesizing six 512 × 512 patterns for each model. The spin-exchange synthesis algorithm was used, which ensures that the first-order statistics of all the textures remain equal, just as described previously in Section 3.1. The synthesis process in every case was continued until virtually no improvement in the results was evident. During the synthesis of the images, a different initial random number seed was used in the synthesis of each pattern. Thus, for any one of the seven texture models, six different backgrounds patterns, or realizations, are present in the experiment, each one a **sample** of all the (infinite) patterns that could be synthesized using that model. Then the appropriate target patterns were synthesized in a 224×224 square in the center of each of the 512×512 background textures, as shown in Figure 7.8. Note that the both the background

patterns and the square target area within the backgrounds are significantly larger than those used in the hybrid textures experiment conducted previously. This was done to allow a larger sample of each texture for the test subjects to compare. The synthesis of the target patterns within the background patterns was accomplished using spin-exchange with blending, in an attempt to minimize the edge effects at the boundary between target and background. The final synthesized image stimuli are shown in Appendix A.

It is extremely important to note another major difference from the hybrid texture paired comparison experiment conducted previously. In the previous experiment we included only one texture pattern for each model (except for the grass texture, which served as the target pattern in all of the image stimuli). Now we have six different background patterns as well as six different target patterns for a total of twelve samples, or realizations, of each texture. Including a greater number of samples of each texture should help reduce an undesirable effect noted after the hybrid textures experiment. Since the observer knows beforehand the precise location of the target in each image stimulus, it is likely that he will look closely at the areas near the border of target and background to make his decision about the distinctness of the target. So if the randomized synthesis process creates a background pattern that contains any distinct features in that particular area of the image, then this may cause a target that is blended into that background to appear more distinct than in a different background pattern of the same texture but with no such features. Although this situation may still occur in some of the background patterns synthesized for a given texture, it is unlikely that it will occur in all of

Table 7.4: The five texture models used in creating the training session.

Texture	Weighting (W_g, W_w)
A	(1,0) [all grass]
B	(0.75,0.25)
C	(0.5,0.5)
D	(0.25,0.75)
E	(0,1) [all wood]

the six backgrounds for that texture, thus the experimental data should be far less biased by such effects.

A training session was also designed, to ensure that the test subjects for the experiment understand what is required of them. For this training, the grass and wood textures from the previous hybrid texture experiment were used. Three synthetic hybrid textures were created from the two natural ones, yielding a total of five texture models. The weighting for each of the five models is shown in Table 7.4, where W_g and W_w are the weights used in the hybridization process for the grass and wood GLC models, respectively. Four stimuli were created from the five models, as shown in Table 7.5. The fourth column of Table 7.5 shows the absolute difference between the hybrid weights of the target and the background. This difference is shown in the format $(|W_{g(t)} - W_{g(b)}|, |W_{w(t)} - W_{w(b)}|)$, where the hybrid weighting of the target texture is $(W_{g(t)}, W_{w(t)})$ and the weighting of the background texture is $(W_{g(b)}, W_{w(b)})$. Based on the results of the hybrid texture experiment, these weighting differences should describe the relative amount of target distinctness for each stimulus. The target and background textures for the four stimuli were selected in-

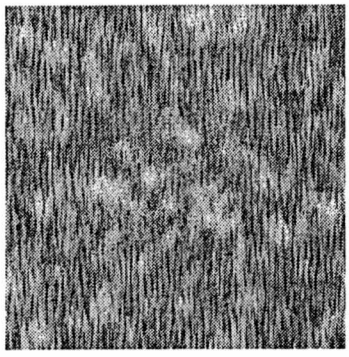
Table 7.5: The scheme used to create the four stimuli for the training session.

Stimulus	Target	Background	Weighting Difference
1	C	D	(0.25,0.25)
2	A	C	(0.5,0.5)
3	B	E	(0.75,0.75)
4	E	A	(1,1)

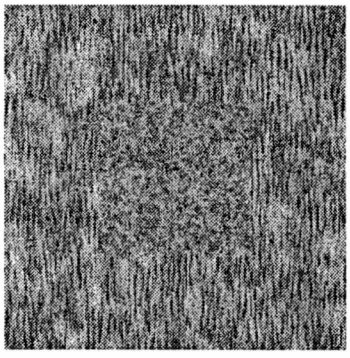
tentionally so as to create stimuli that have weighting differences that are different for each stimulus, by a large enough amount that it is expected that most observers will agree on which target is more distinct when any two of the stimuli are presented together. The training session for each observer consisted of the presentation of all six possible pairings of the four stimuli, and it was decided that an observer must make the "expected" selection in at least five out of the six pairings to be allowed to participate in the full experiment. The four 512×512 image stimuli used in the training session are shown in Figure 7.9.

7.2.2 Design and Conduct of the Experiment

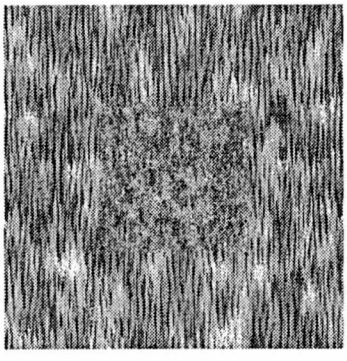
Since there are $N = 7$ textures being considered, there are $\binom{N}{3} = \frac{N!}{(N-3)!3!} = \frac{N(N-1)(N-2)}{3!} = \frac{(7)(6)(5)}{6} = 35$ possible triads of textures. Since the complete method of triads requires that each triad be presented three times, there are a total of $35 \times 3 = 105$ decisions to be made by each observer. Another way of looking at this is that we are essentially conducting seven different paired-comparison experiments simultaneously. Each of these paired-comparison experiments has one of the seven textures as the common target, which is being compared to $N = 6$ other textures,



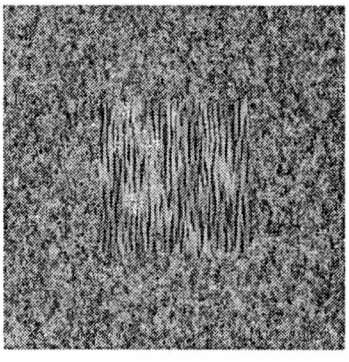
(a) Training stimulus C_on_D



(b) Training stimulus A_on_C



(c) Training stimulus B_on_E



(d) Training stimulus E_on_A

Figure 7.9: The four image stimuli used in the training session.

yielding $\binom{N}{2} = 15$ paired comparisons. This also gives $15 \times 7 = 105$ total decisions. For each comparison, the test subject was once again required to choose which of the two images in which he thought the target pattern is more distinct from its background.

7.2.3 Results and Analysis of the Data

A total of 15 test subjects participated in the experiment. During the training sessions, one test subject failed to make the "expected" choice in only one of the six

pairings, while the other 14 subjects made the expected choice in all six. Thus, all 15 test subjects were allowed to participate in the full experiment. For an experiment implementing the complete method of triads for N stimuli, the raw decision data can be represented by N $N \times N$ matrices ${}_kF_{ij}$, where the element ${}_kf_{ij}$ is the number of times (frequency) that stimulus k was judged more similar to stimulus i than to j. The resulting $N = 7$ F_{ij} matrices for this experiment are:

$$ {}_1F_{ij} = \begin{bmatrix} - & - & - & - & - & - & - \\ - & - & 15 & 2 & 14 & 14 & 11 \\ - & 0 & - & 0 & 0 & 0 & 0 \\ - & 13 & 15 & - & 15 & 14 & 15 \\ - & 1 & 15 & 0 & - & 3 & 4 \\ - & 1 & 15 & 1 & 12 & - & 7 \\ - & 4 & 15 & 0 & 11 & 8 & - \end{bmatrix} \tag{7.2} $$

$$ {}_2F_{ij} = \begin{bmatrix} - & - & 15 & 9 & 15 & 15 & 11 \\ - & - & - & - & - & - & - \\ 0 & - & - & 0 & 0 & 0 & 0 \\ 6 & - & 15 & - & 15 & 14 & 12 \\ 0 & - & 15 & 0 & - & 7 & 1 \\ 0 & - & 15 & 1 & 8 & - & 1 \\ 4 & - & 15 & 3 & 14 & 14 & - \end{bmatrix} \tag{7.3} $$

$$
{}_3F_{ij} = \begin{bmatrix}
- & 0 & - & 6 & 0 & 0 & 0 \\
15 & - & - & 14 & 0 & 0 & 0 \\
- & - & - & - & - & - & - \\
9 & 1 & - & - & 0 & 0 & 0 \\
15 & 15 & - & 15 & - & 3 & 11 \\
15 & 15 & - & 15 & 12 & - & 15 \\
15 & 15 & - & 15 & 4 & 0 & -
\end{bmatrix} \tag{7.4}
$$

$$
{}_4F_{ij} = \begin{bmatrix}
- & 11 & 15 & - & 15 & 15 & 15 \\
4 & - & 15 & - & 15 & 15 & 14 \\
0 & 0 & - & - & 0 & 0 & 0 \\
- & - & - & - & - & - & - \\
0 & 0 & 15 & - & - & 10 & 0 \\
0 & 0 & 15 & - & 5 & - & 0 \\
0 & 1 & 15 & - & 15 & 15 & -
\end{bmatrix} \tag{7.5}
$$

$$
{}_5F_{ij} = \begin{bmatrix}
- & 3 & 6 & 3 & - & 2 & 0 \\
12 & - & 13 & 4 & - & 2 & 1 \\
9 & 2 & - & 2 & - & 1 & 0 \\
12 & 11 & 13 & - & - & 1 & 1 \\
- & - & - & - & - & - & - \\
13 & 13 & 14 & 14 & - & - & 3 \\
15 & 14 & 15 & 14 & - & 12 & -
\end{bmatrix} \tag{7.6}
$$

$$
{}_6F_{ij} = \begin{bmatrix}
- & 1 & 12 & 2 & 0 & - & 0 \\
14 & - & 15 & 2 & 0 & - & 0 \\
3 & 0 & - & 0 & 0 & - & 0 \\
13 & 13 & 15 & - & 2 & - & 0 \\
15 & 15 & 15 & 13 & - & - & 2 \\
- & - & - & - & - & - & - \\
15 & 15 & 15 & 15 & 13 & - & -
\end{bmatrix} \tag{7.7}
$$

$$
{}_7F_{ij} = \begin{bmatrix}
- & 0 & 15 & 0 & 9 & 0 & - \\
15 & - & 15 & 7 & 15 & 11 & - \\
0 & 0 & - & 0 & 0 & 0 & - \\
15 & 8 & 15 & - & 15 & 10 & - \\
6 & - & 15 & 0 & - & 2 & - \\
15 & 4 & 15 & 5 & 13 & - & - \\
- & - & - & - & - & - & -
\end{bmatrix} \tag{7.8}
$$

Notice that the data matrices ${}_kF_{ij}$ have vacant cells in the main diagonal, and in the kth row and column. This is because no single target/background combination was ever compared to itself, and no single texture served as both target and background in the same image. These raw frequency (${}_kF_{ij}$) matrices were first converted to appropriate proportion (${}_kP_{ij}$) matrices by dividing each occupied cell by 15, the total number of observations. In condition C of the law of comparative judgment (LCJ), it is assumed that the standard deviation of the distribution of discriminal differences is constant for all pairs of stimuli. This is the simplifying condition used during the previous scaling experiments that reduces the LCJ to a solvable set of equations. For

triadic comparisons, the distances between stimuli on the "psychological continuum" can be considered to be scaled so that this constant standard deviation is equal to unity. If this is the case, each experimentally observed proportion ($_kp_{ij}$) can be replaced by the corresponding unit normal deviate ($_kx_{ij}$) to create 7 X_{ij} matrices, according to the equation:

$$_kp_{ij} = \int_{-\infty}^{kx_{ij}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)\, dx. \tag{7.9}$$

The appropriate $_kx_{ij}$ can be obtained for each $_kp_{ij}$ from a look-up table of areas under the unit normal curve created by numerical integration. We represent the psychological "distance" between two stimuli i and j as $d_{ij} = d_{ji}$. Each $_kx_{ij}$ element is an estimate of the difference in distance for two pairs of stimuli:

$$_kx_{ij} = d_{kj} - d_{ki} \tag{7.10}$$

The 7 X_{ij} matrices for this experiment are:

$$_1X_{ij} = \begin{bmatrix} -- & -- & -- & -- & -- & -- & -- \\ -- & -- & \infty & -1.11 & 1.50 & 1.50 & 0.62 \\ -- & -\infty & -- & -\infty & -\infty & -\infty & -\infty \\ -- & 1.11 & \infty & -- & (2.48) & 1.50 & (1.58) \\ -- & -1.50 & \infty & (-2.48) & -- & -0.84 & -0.62 \\ -- & -1.50 & \infty & -1.50 & 0.84 & -- & -0.08 \\ -- & -0.62 & \infty & (-1.58) & 0.62 & 0.08 & -- \end{bmatrix} \tag{7.11}$$

$$
{}_2X_{ij} = \begin{bmatrix}
-- & -- & \infty & 0.25 & (1.39) & (1.29) & 0.62 \\
-- & -- & -- & -- & -- & -- & -- \\
-\infty & -- & -- & -\infty & -\infty & -\infty & -\infty \\
-0.25 & -- & \infty & -- & (1.96) & 1.50 & 0.84 \\
(-1.39) & -- & \infty & (-1.96) & -- & -0.08 & -1.50 \\
(-1.29) & -- & \infty & -1.50 & 0.08 & -- & -1.50 \\
-0.62 & -- & \infty & -0.84 & 1.50 & 1.50 & --
\end{bmatrix} \tag{7.12}
$$

$$
{}_3X_{ij} = \begin{bmatrix}
-- & (-1.75) & -- & -0.25 & -\infty & -\infty & -\infty \\
(1.75) & -- & -- & 1.50 & -\infty & -\infty & -\infty \\
-- & -- & -- & -- & -- & -- & -- \\
0.25 & -1.50 & -- & -- & -\infty & -\infty & -\infty \\
\infty & \infty & -- & \infty & -- & -0.84 & 0.62 \\
\infty & \infty & -- & \infty & 0.84 & -- & (1.46) \\
\infty & \infty & -- & \infty & -0.62 & (-1.46) & --
\end{bmatrix} \tag{7.13}
$$

$$
{}_4X_{ij} = \begin{bmatrix}
-- & 0.62 & \infty & -- & \infty & (0.39) & (2.12) \\
-0.62 & -- & \infty & -- & \infty & (0.39) & 1.50 \\
-\infty & -\infty & -- & -- & -\infty & -\infty & -\infty \\
-- & -- & -- & -- & -- & -- & -- \\
-\infty & -\infty & \infty & -- & -- & 0.43 & \infty \\
(-0.39) & (-0.39) & \infty & -- & -0.43 & -- & \infty \\
(-2.12) & -1.50 & \infty & -- & \infty & \infty & --
\end{bmatrix} \tag{7.14}
$$

$$
{}_5X_{ij} = \begin{bmatrix}
-- & -0.84 & -0.25 & -0.84 & -- & -1.11 & (-1.75) \\
0.84 & -- & 1.11 & -0.62 & -- & -1.11 & -1.50 \\
0.25 & -1.11 & -- & -1.11 & -- & -1.50 & (-2.52) \\
0.84 & 0.62 & 1.11 & -- & -- & -1.50 & -1.50 \\
-- & -- & -- & -- & -- & -- & -- \\
1.11 & 1.11 & 1.50 & 1.50 & -- & -- & -0.84 \\
(1.75) & 1.50 & (2.52) & 1.50 & -- & 0.84 & --
\end{bmatrix} \tag{7.15}
$$

$$
{}_6X_{ij} = \begin{bmatrix}
-- & -1.50 & 0.84 & -1.11 & (-1.25) & -- & -\infty \\
1.50 & -- & (2.34) & -1.11 & (-1.63) & -- & (-0.88) \\
-0.84 & (-2.34) & -- & (-1.95) & (-0.66) & -- & -\infty \\
1.11 & 1.11 & (1.95) & -- & -1.11 & -- & (-2.22) \\
(1.25) & (1.63) & (0.66) & 1.11 & -- & -- & -1.11 \\
-- & -- & -- & -- & -- & -- & -- \\
\infty & (0.88) & \infty & (2.22) & 1.11 & -- & --
\end{bmatrix} \tag{7.16}
$$

$$
{}_7X_{ij} = \begin{bmatrix}
-- & (0.00) & \infty & -\infty & 0.25 & (-0.86) & -- \\
(0.00) & -- & \infty & -0.08 & (0.87) & 0.62 & -- \\
-\infty & -\infty & -- & -\infty & -\infty & -\infty & -- \\
\infty & 0.08 & \infty & -- & (1.54) & 0.43 & -- \\
-0.25 & (-0.87) & \infty & (-1.54) & -- & -1.11 & -- \\
(0.86) & -0.62 & \infty & -0.43 & 1.11 & -- & -- \\
-- & -- & -- & -- & -- & -- & --
\end{bmatrix} \tag{7.17}
$$

As always with the LCJ, there is the question of how to transform experimental proportions ${}_kp_{ij}$ of either 0.0 or 1.0 to appropriate ${}_kx_{ij}$ values. When using data

from triadic comparisons, we can estimate the value for such an apparently infinite element from $N-2$ other pairs of ${}_kx_{ij}$ values. Any ${}_kx_{ij}$ can be estimated from $N-3$ differences as in Equation 7.18 and also one other difference as in Equation 7.20:

$$ {}_kx_{ij} = {}_kx_{im} - {}_k\, x_{jm}, m \neq i, j, k \tag{7.18} $$

$$ (d_{kj} - d_{ki} = d_{km} - d_{ki} - d_{km} + d_{kj}) \tag{7.19} $$

$$ {}_kx_{ij} = {}_ix_{kj} - {}_j\, x_{ki} \tag{7.20} $$

$$ (d_{kj} - d_{ki} = d_{ij} - dik - d_{ji} + d_{jk}) \tag{7.21} $$

Of course, a number of these $N-2$ other equations may also contain infinite elements. The procedure used was to estimate infinite ${}_kx_{ij}$ elements with the average of the estimates from the $N-2$ equations that contained only finite elements. The elements shown in parentheses correspond to ${}_kx_{ij}$ values estimated in this manner. Elements shown as $\pm\infty$ still could not be estimated since all of the $N-2$ equations contain at least one infinite element.

What is desired is a set of values which describe the psychological comparative distance between each possible pair of the $N = 7$ texture stimuli. For a least-squares solution, the method is given by Torgerson [Tor58]. We first calculate a new matrix, ${}_kX_{.j}$, which contains as each row the average of the columns of the corresponding

${}_kX_{ij}$ matrix. This matrix is:

$$ {}_kX_{.j} = \begin{bmatrix} -- & -0.63 & \infty & -1.67 & 1.36 & 0.56 & 0.37 \\ -0.89 & -- & \infty & -1.01 & 1.23 & 1.05 & -0.39 \\ 1.00 & -1.63 & -- & 0.63 & 0.11 & -1.15 & 1.04 \\ -1.04 & -0.42 & \infty & -- & -0.43 & 0.40 & 1.81 \\ 0.96 & 0.26 & 1.20 & 0.09 & -- & -0.88 & -1.62 \\ 0.75 & -0.05 & 1.45 & -0.17 & -0.71 & -- & -1.40 \\ 0.20 & -0.35 & \infty & -0.68 & 0.94 & -0.23 & -- \end{bmatrix} \tag{7.22} $$

In computing the column averages, the remaining infinite ${}_kx_{ij}$ elements were ignored. If the only values of a particular column are infinite, then the average of that column is considered to be infinite. Next, the averages of the columns of the ${}_kX_{.j}$ matrix are calculated, once again ignoring the remaining infinite elements. This result is:

$$ {}_.X_{.j} = [0.16 \quad -0.47 \quad 1.32 \quad -0.47 \quad 0.42 \quad -0.04 \quad -0.03] \tag{7.23} $$

Then G_{kj} is computed, such that $g_{kj} =_k x_{.j} +_. x_{.k}$. G_{kj} is thus a 7×7 matrix, with elements that are the nonsymmetric psychological comparative distances between pairs of textures. With infinite elements of ${}_kX_{.j}$ carrying through, this result is:

$$ G = \begin{bmatrix} -- & -0.46 & \infty & -1.50 & 1.52 & 0.72 & 0.54 \\ -1.36 & -- & \infty & -1.48 & 0.76 & 0.58 & -0.85 \\ 2.32 & -0.30 & -- & 1.95 & 1.43 & 0.17 & 2.36 \\ -1.51 & -0.89 & \infty & -- & -0.90 & -0.07 & 1.34 \\ 1.38 & 0.67 & 1.62 & 0.50 & -- & -0.46 & -1.20 \\ 0.71 & -0.09 & 1.41 & -0.21 & -0.75 & -- & -1.44 \\ 0.17 & -0.38 & \infty & -0.71 & 0.91 & -0.26 & -- \end{bmatrix} \tag{7.24} $$

We desire inter-texture psychological comparative distances that are symmetrical ($d_{ij} = d_{ji}$). We finally compute H_{kj} such that $h_{kj} = \frac{1}{2}(g_{kj} + g_{jk})$. However, if either g_{kj} or g_{jk} is infinite, we consider h_{kj} to be infinite. This result is:

$$H = \begin{bmatrix} -- & -0.91 & \infty & -1.51 & 1.45 & 0.72 & 0.36 \\ -0.91 & -- & \infty & -1.19 & 0.72 & 0.25 & -0.62 \\ \infty & \infty & -- & \infty & 1.52 & 0.79 & \infty \\ -1.51 & -1.19 & \infty & -- & -0.20 & -0.14 & 0.31 \\ 1.45 & 0.72 & 1.52 & -0.20 & -- & -0.60 & -0.15 \\ 0.72 & 0.25 & 0.79 & -0.14 & -0.60 & -- & -0.85 \\ 0.36 & -0.62 & \infty & 0.31 & -0.15 & -0.85 & -- \end{bmatrix} \tag{7.25}$$

Out of 21 possible pairs of textures, we have an estimate of the psychological comparative distance for 17 of them. Some of these 17 distances are negative values. As these distances are only comparative, there is no established "zero" distance. This means that the distances do not correspond to any spatial model, and have been established only relative to each other. A more negative value simply indicates a smaller comparative distance. The absolute distances (d_{jk}), which would correspond to a spatial model, are related to the comparative distances by an unknown additive constant c:

$$d_{jk} = h_{jk} + c \tag{7.26}$$

For our purposes, however, the comparative distances are adequate. We can now compare these distances to the set of GLC-based metrics which was also compared to the results of the two previous experiments and defined in Equation 6.1 through Equation 6.6.

These six metrics were calculated for all 21 possible pairs of the 7 textures in the experiment. In every case, the GLC model used in the calculations is the "prototype" model from which the various patterns were synthesized. The resulting calculations are given in exponential notation in Table 7.6. The prototype texture models are numbered 1 through 7 as in Figure 7.7. Using only the 17 of the 21 texture pairs for which we successfully estimated a psychological comparative distance that is finite (the h_{kj} values in Equation 7.25), a correlation coefficient with the calculations for each of the six metrics was computed. The results are shown in Table 7.7.

The computed coefficients do indicate a strong positive correlation between the estimated psychological distances and the computed metrics. However, they are significantly lower than those computed during the previous two experiments. In consideration of this, we must keep in mind that the psychological distances are only estimates, based on the psychophysical data. During the estimation process, there were several instances when infinite values were simply ignored. The effect of this on the resulting scales has already been noted during the analysis of the data from the previous hybrid texture experiment. In addition, the underlying theory of the LCJ, upon which the entire estimation process was based, is that the discriminal processes associated with any stimulus are normally distributed, and also that the standard deviation of the distribution of discriminal differences is constant for all pairs of stimuli. This "normality assumption" might not be an appropriate one for our stimuli. These issues serve as several possible sources of error in the final psychological distances, even if the original raw data is free of error.

We can avert these potential errors by avoiding the LCJ analysis altogether and

Table 7.6: The six GLC metrics calculated for every possible pair of the 7 prototype texture models.

Model Pair	ACE	WACE	AWCE	ASCE	WASCE	AWSCE
1,2	1.9049e-01	3.5423e-02	1.0167e-01	1.3258e-03	2.5148e-04	8.2519e-04
1,3	3.2923e-01	6.3841e-02	1.8221e-01	4.7722e-03	1.0325e-03	3.2937e-03
1,4	9.5273e-02	1.7717e-02	5.0841e-02	3.3154e-04	6.2877e-05	2.0632e-04
1,5	2.2352e-01	4.1796e-02	1.1865e-01	1.9696e-03	3.8523e-04	1.2448e-03
1,6	1.6464e-01	3.1919e-02	9.1098e-02	1.1932e-03	2.5813e-04	8.2342e-04
1,7	1.4896e-01	2.7855e-02	7.9089e-02	8.7545e-04	1.7120e-04	5.5329e-04
2,3	2.9394e-01	5.8609e-02	1.6938e-01	4.3167e-03	1.0268e-03	3.2577e-03
2,4	9.5221e-02	1.7706e-02	5.0828e-02	3.3153e-04	6.2890e-05	2.0634e-04
2,5	1.4705e-01	2.9313e-02	8.4703e-02	1.0795e-03	2.5678e-04	8.1454e-04
2,6	1.9572e-01	3.8309e-02	1.1123e-01	1.6285e-03	3.8111e-04	1.2182e-03
2,7	1.3042e-01	2.5532e-02	7.4140e-02	7.2357e-04	1.6933e-04	5.4127e-04
3,4	2.9330e-01	5.8134e-02	1.6602e-01	4.2127e-03	9.6680e-04	3.0695e-03
3,5	1.4689e-01	2.9297e-02	8.4675e-02	1.0790e-03	2.5667e-04	8.1440e-04
3,6	1.6459e-01	3.1923e-02	9.1111e-02	1.1931e-03	2.5813e-04	8.2351e-04
3,7	1.9555e-01	3.8759e-02	1.1068e-01	1.8727e-03	4.2976e-04	1.3644e-03
4,5	1.6463e-01	3.1922e-02	9.1100e-02	1.1930e-03	2.5815e-04	8.2340e-04
4,6	1.4701e-01	2.9312e-02	8.4706e-02	1.0793e-03	2.5676e-04	8.1453e-04
4,7	9.7751e-02	1.9376e-02	5.5343e-02	4.6802e-04	1.0741e-04	3.4101e-04
5,6	9.5266e-02	1.7712e-02	5.0833e-02	3.3142e-04	6.2874e-05	2.0627e-04
5,7	7.4557e-02	1.3941e-02	3.9562e-02	2.1894e-04	4.2834e-05	1.3834e-04
6,7	6.5303e-02	1.2777e-02	3.7094e-02	1.8116e-04	4.2396e-05	1.3548e-04

Table 7.7: The correlation coefficients between the 17 finite psychological comparative distances and the corresponding metric calculations.

Metric	Correlation
ACE	0.5558
WACE	0.5853
AWCE	0.5802
ASCE	0.6314
WASCE	0.6585
AWSCE	0.6522

analyzing the raw decisions made by the observers. These decisions are listed in the table included as Appendix B. In this table, the first column gives the comparison number (1→105). The second column gives the texture model (numbered as in Figure 7.7) of the target pattern for each comparison. This texture served as the target texture in both the left and right stimuli. The third column gives the texture model number for the background pattern of the stimulus displayed on the left, and the fourth column gives the background texture number of the stimulus on the right. The fifth and sixth columns give the number of times (out of 15 observations) that the left and right stimuli were chosen, respectively. The seventh column gives the value of the ACE metric computed for the target and background models of the left stimulus minus the same metric computed for the right stimulus. Thus, a negative value in this column means that the ACE metric for the right stimulus was greater than for the left stimulus. Note that this value is given in exponential notation. The eighth column lists a code, either 1 for right or -1 for wrong, for whether the ACE

metric correctly yields a higher value for the stimulus selected by a majority of the observers. The final two columns of the table will be discussed later. As an example, consider the data for comparison #1 in the first row of the table in Appendix B. We see that the left stimulus consisted of texture #7 as the target compared to a background of texture #3. The right stimulus was texture #7 against texture #4. All 15 observers agreed that the target in the left stimulus was more distinct from its background than the target in the right stimulus. The ACE metric computed for textures #3 and #7 is greater than the metric computed for textures #4 and #7 by a difference of 9.7799e-02, thus the ACE metric correctly predicts the selection of the left stimulus by the observers.

This tabulated data provides a method for evaluating our GLC-based metrics without first applying an analysis based on the LCJ. One possible evaluation is the proportion of the 105 comparisons for which the metric correctly predicts the stimulus selected by a majority of the observers. Another method is to compute the same proportion but considering only the comparisons for which at least 80% or 90% of the observers agreed in their decision. Only 89 of the 105 comparisons satisfied 80% agreement, while only 72 satisfied 90% agreement. These computed proportions are given in Table 7.8. Although these computed proportions indicate that the GLC-based metrics are relatively successful at predicting the observers' judgments, there are still a significant number of stimulus comparisons for which the metrics disagreed with the observers. Upon a close examination of the individual comparisons which gave the metrics problems, an interesting trend was discovered. It seems that the observers, when making their judgments, usually used the difference in the "water"

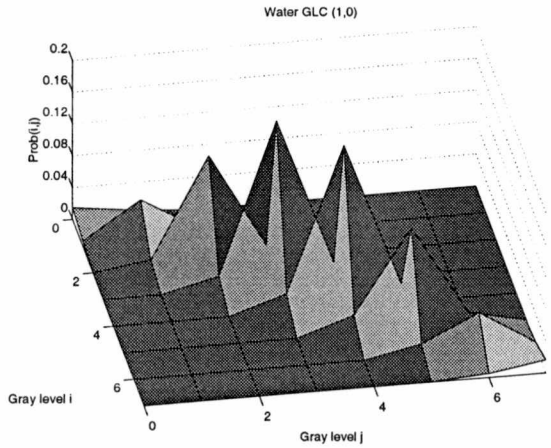
Table 7.8: The computed proportions of correct predictions for each of the six GLC-based metrics.

Metric	All	≥80% Agree	≥90% Agree
ACE	80.0%	87.6%	90.3%
WACE	81.0%	88.8%	90.3%
AWCE	81.0%	88.8%	90.3%
ASCE	80.0%	87.6%	90.3%
WASCE	82.9%	89.9%	93.1%
AWSCE	81.9%	88.8%	90.3%

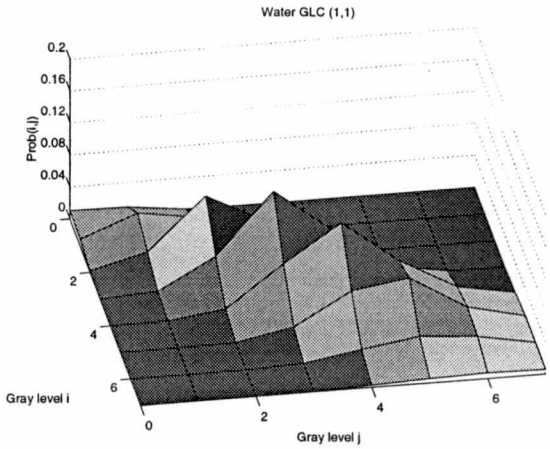
content between the target and background as the major criterion affecting the distinctness of the target. In addition, it seems that this criterion was non-linear, in that the perceptual difference was considered larger for textures with more water content. This phenomenon can be best described by examining a particular example from the data set. Consider comparison #73 from the table in Appendix B. In this comparison, texture #5 was compared both to texture #3 and to texture #2. From Figure 7.7, we see that texture model #5 was created by equally combining models #2 and #3. The result of this is that every element of every GLC matrix in model #5 is equidistant from the corresponding element of models #2 and #3. Our GLC-based metrics, which are all based on comparing these element-by-element distances, will therefore indicate that texture #5 is equidistant from textures #2 and #3 (except for a small amount of round-off error). This is evident in Table 7.6, where the metric calculations on the 2,5 line are virtually identical to those on the 3,5 line. However, in comparison #73, we see that 13 out of 15 observers judged

textures #5 and #3 to be more distinct than textures #5 and #2. The result of this is evident in the psychological comparative distances in Equation 7.25, where the distance from texture #2 to #5 is estimated to be 0.72, while the distance from texture #3 to #5 is estimated to be 1.52.

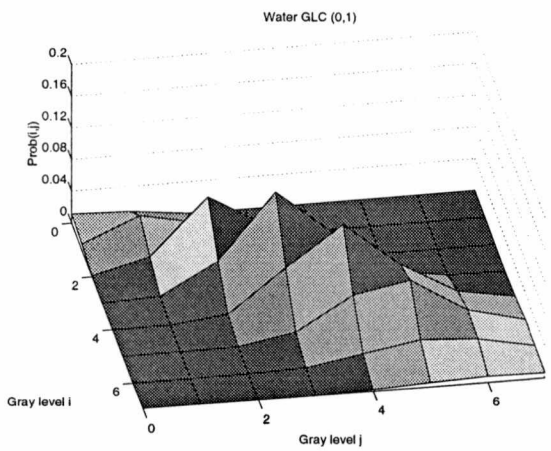
From an examination of the stimuli (Appendix A), it is clear that the great perceptual difference between the water texture and the leaves and stone textures is due to the strong horizontal directionality of the tiling structure of the water texture. This directionality can be viewed in a fundamental way by visualizing the GLC matrices of the water texture. The GLC matrices for the four smallest displacements are plotted in Figure 7.10. In (a) is the GLC matrix for displacement vector [1,0], the horizontal neighbor matrix. We can see that the horizontal tiling structure is evident in that this matrix has a much stronger main diagonal than the matrices for the other three directions. Because of this, we can surmise that the energy contained in this matrix is greater than the energy for the other three directions. We can compute the energy in a GLC matrix by simply summing the squares of the values at every element. This yields a scalar value for each displacement. If we plot these values indexed by the spatial coordinates, we can also get an idea of the tiling structure of the texture. This is done for the three original natural textures in Figure 7.11. From these plots, it is evident that the leaves and stone textures have energy that is relatively evenly spread in all directions, while the water texture has a large amount of energy horizontally but very little in the other directions. We can measure the amount of directionality in a texture by computing the variance of the energy (VOE) at different displacements. Then we can estimate the difference



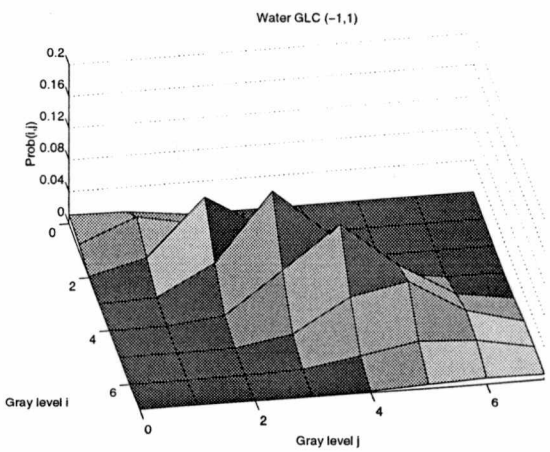
(a) Water GLC [1,0]



(b) Water GLC [1,1]

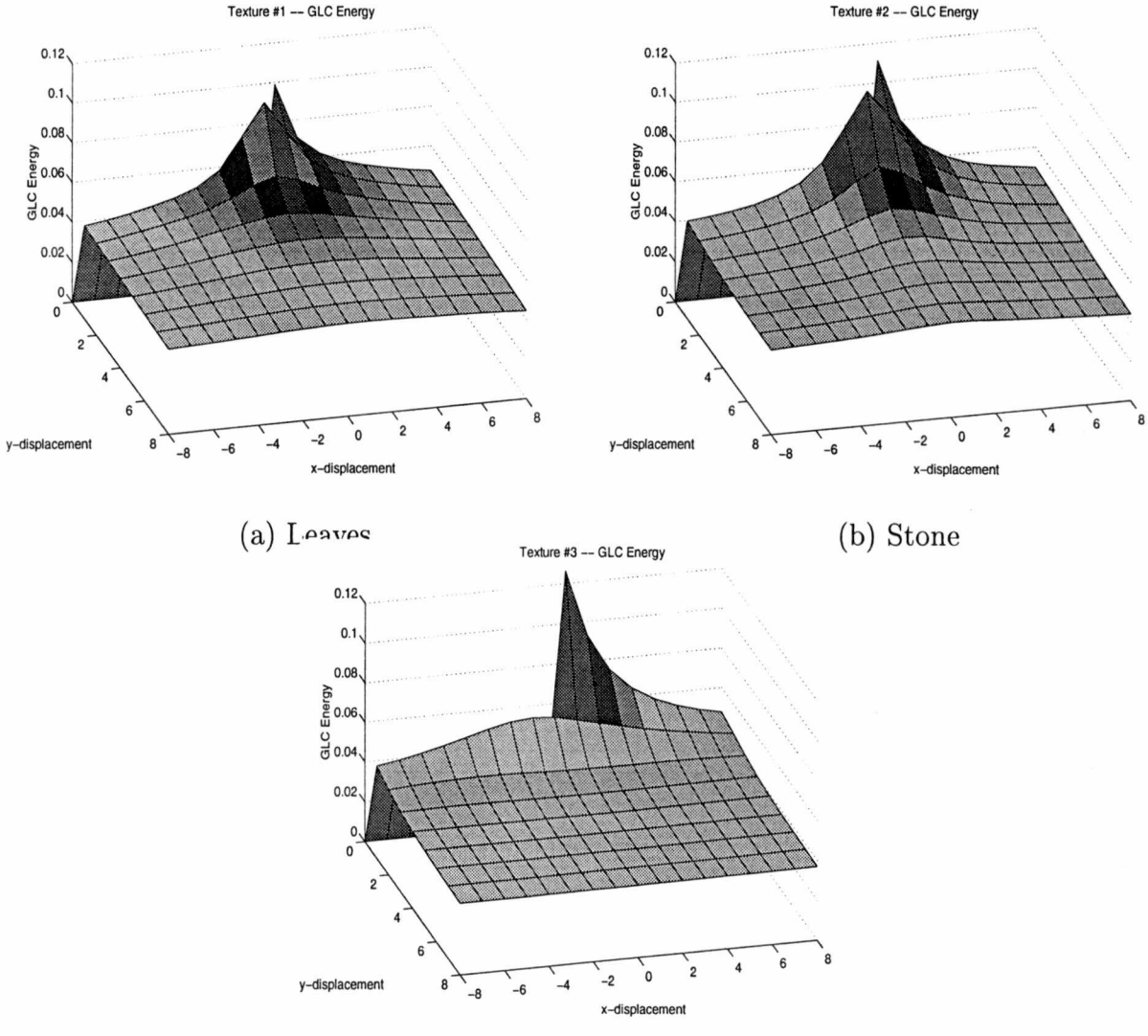


(c) Water GLC [0,1]



(d) Water GLC [-1,1]

Figure 7.10: The GLC matrices of the water texture for the four smallest displacements.



(a) Leaves

(b) Stone

(c) Water

Figure 7.11: Plots of the energy in the GLC matrices for every displacement for the three original natural textures.

in directionality between any two textures by simply taking the absolute difference between their VOE values. These calculations for all possible pairs of the seven

textures are given in matrix format in Equation 7.27.

$$\Delta\text{VOE} = \begin{bmatrix} -- & 3.2882e-05 & 7.3132e-04 & 1.4191e-05 & 2.9417e-04 & 2.2222e-04 & 1.4411e-04 \\ 3.2882e-05 & -- & 6.9844e-04 & 1.8692e-05 & 2.6129e-04 & 1.8934e-04 & 1.1123e-04 \\ 7.3132e-04 & 6.9844e-04 & -- & 7.1713e-04 & 4.3715e-04 & 5.0910e-04 & 5.8721e-04 \\ 1.4191e-05 & 1.8692e-05 & 7.1713e-04 & -- & 2.7998e-04 & 2.0803e-04 & 1.2992e-04 \\ 2.9417e-04 & 2.6129e-04 & 4.3715e-04 & 2.7998e-04 & -- & 7.1950e-05 & 1.5006e-04 \\ 2.2222e-04 & 1.8934e-04 & 5.0910e-04 & 2.0803e-04 & 7.1950e-05 & -- & 7.8108e-05 \\ 1.4411e-04 & 1.1123e-04 & 5.8721e-04 & 1.2992e-04 & 1.5006e-04 & 7.8108e-05 & -- \end{bmatrix} \tag{7.27}$$

For these calculations, the energy variance was taken over only the four shortest displacements. As variance is a square-order calculation rather than a linear one, we see the result that the difference between texture #3 and texture #5, from our previous example, is greater than the difference between textures #2 and #5.

The final two columns of the data in Appendix B give the VOE measure computed for the models of the left stimulus minus the same metric computed for the right stimulus, and the 1 or -1 code indicating whether the VOE metric correctly yields a higher value for the stimulus selected by a majority of the observers. The same proportions as calculated for the six metrics earlier (Table 7.8) were also calculated for the new VOE metric. The VOE metric correctly predicts the stimulus chosen by a majority of the observers in 89.5% of the 105 comparisons. The VOE metric correctly predicts 95.5% of the 89 comparisons for which there was 80% agreement, and 100% of the 72 for which there was 90% agreement. These proportions are significantly higher than those for the earlier metrics. In addition, the VOE metric calculations were correlated with the 17 finite distances that were estimated using the LCJ (the h_{kj} values in Equation 7.25). The resulting correlation coefficient is 0.8271, which is also significantly higher than the earlier correlations in Table 7.7.

7.2.4 Conclusions

The fact that the VOE metric does a better job than the earlier six metrics of explaining the results of the triadic comparisons experiment should be considered with caution. The VOE metric was developed in an *ad hoc* fashion, in that the data from the experiment was examined and then a metric was invented to explain the data. The VOE metric is intended to measure differences in the levels of directionality in textures, which is apparently the most significant perceptual cue that was used by the observers in this experiment.

This new metric was also calculated for the stimuli patterns of the previous hybrid texture experiment as well as the three earlier synthesis error experiments. For the hybrid texture experiment, the correlation coefficient for VOE was 0.9056 for the traditional scales, as compared to 0.9163→0.9285 for the six other metrics. For the adjusted scales, the correlation was 0.9245, as compared to 0.9334→0.9546. For the synthesis error experiments, the correlation was 0.9473 as compared to 0.9726→0.9784 for pigskin, 0.8814 as compared to 0.9473→0.9764 for raffia, and 0.7871 as compared to 0.9628→0.9817 for evergreen. It may be true that the directionality cue was not as important in these earlier experiments as other cues. It is likely that if human texture perception is to be modeled precisely, a greater understanding of the various perceptual cues is needed. It may also be necessary to possess a set of several different metrics, perhaps one for every cue that is being utilized by the observer.

CHAPTER 8

Natural Texture Patterns and Human Perception

This experiment is similar to the paired comparison experiments conducted previously in that each image stimulus consists of a particular square target pattern centered against a particular background pattern. The X-based Perceptual Experiment Testbed (XPET) was again used to present the observer with every possible pair of stimuli, and each observer responds as to which of the two images has a target which is more distinct from its background. But this experiment differs from those conducted previously in three important ways. First, the texture patterns are real patterns extracted from the new database images, instead of patterns synthesized from GLC texture models. Second, no attempt was made to modify the histograms of the patterns when creating the stimuli, so that both first order and higher order statistics are varying simultaneously. This also means that the stimulus patterns are not quantized to only eight gray levels as before, but are maintained as eight-bit images (256 possible gray levels). Third, the target patterns in the image stimuli are not "blended" into their respective background patterns as before, so that the impact of edge effects at the target/background boundaries can be studied.

This experiment represents a shift in research direction compared to our previous studies. In all of the previous studies, the image stimuli consisted of synthetic, computer generated patterns. This allowed us to carefully control various image

statistics, enabling us to isolate the specific variables we wished to study while keeping other variables constant. In this manner, we were able to study human pattern discrimination in a very fundamental way and to develop and test the ability of many different image metrics to model the psychophysical data.

Although the results of these earlier studies are pleasing, there are two potential problems with this methodology. The first potential problem is that the results obtained using synthetic imagery may not necessarily be directly applicable to real-world situations. For example, we cannot expect to encounter only "histogram matched" battlefields! The other potential problem is that we were not always successful at keeping constant the variables we wished not to study. For example, despite the use of the GLC synthesis "blending" algorithm to minimize edge effects along target/background boundaries, it was noted that some stimuli contained distinctly bright or dark pixel regions very near the target boundaries. From the raw observer decision data, it was evident that some observers may have been using such features as a cue in judging target distinctness.

This experiment utilizes image patterns extracted from real outdoor scenes. No attempt is made to modify the image statistics to force variables to be constant. This makes this experiment more difficult in two ways. The first difficultly is that we will be studying the effects of several independent variables acting simultaneously. The statistical analysis of the resulting data will necessarily utilize more complicated, multivariable techniques. Also, since we are no longer ignoring the effects of target boundaries on judging target distinctness, we will need one or more computational measures of such effects. The second difficulty is that our stimuli can't possibly

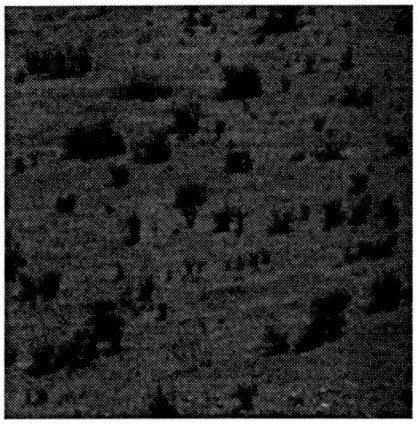
Table 8.1: Descriptions of each of the six real texture samples.

Texture	Content
A	grass and bushes
B	medium length grass
C	stone and grass
D	short grass
E	stones and bushes
F	small bushes

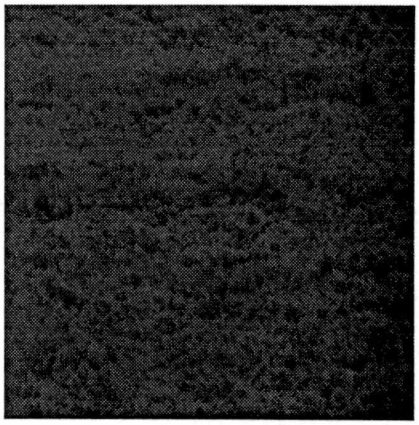
represent a complete, evenly spaced sampling of the "space" of possible patterns. We have attempted to include image patterns possessing a variety of different visual characteristics, but we can't hope to learn everything there is to know from this single experiment. As always, we must remember that our overall success in studying and interpreting the final results of the experiment will depend heavily on scientific intuition.

8.1 Texture Selection and Stimulus Creation

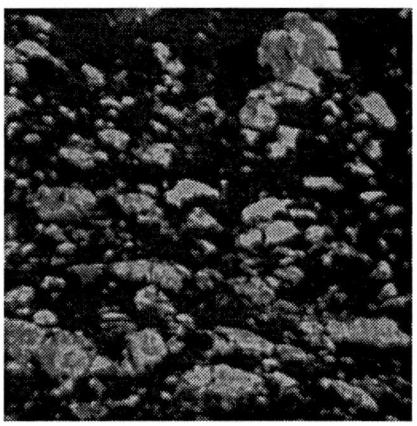
From a database of color images of various outdoor scenes in southern California, six 256×256 sections of relatively homogeneous texture were extracted. Details regarding the database and the procedure of collecting the images are given later in Section 9.1. These six new images were converted to gray scale versions by simply averaging the red, blue, and green channels for each pixel. The resulting images are shown in Figure 8.1. A short description of each texture is given in Table 8.1.



(a) Texture A



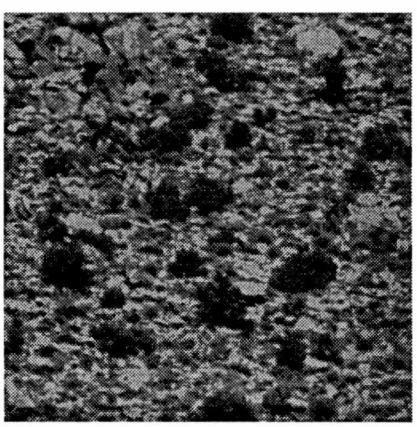
(b) Texture B



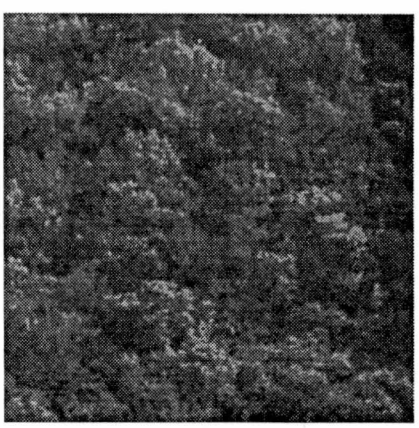
(c) Texture C



(d) Texture D



(e) Texture E



(f) Texture F

Figure 8.1: The six 256×256 real texture samples.

Table 8.2: The scheme for creating the 15 stimuli from the six real texture samples.

	Background					
Target	A	B	C	D	E	F
A		•	•	•	•	•
B			•	•	•	•
C				•	•	•
D					•	•
E						•
F						

It was desired to include one stimulus in the experiment for every possible pair of texture samples. With $T = 6$ texture samples, there are a total of $\binom{T}{2} = \frac{T!}{(T-2)!2!} = \frac{T(T-1)}{2} = \frac{(6)(5)}{2} = 15$ unique pairs of textures. Thus, 15 stimuli were created, according to the scheme indicated in Table 8.2. Each of the 15 bullets in the table represent the target/background combination for one stimulus. A two letter designation will be used to refer to each stimulus, the first letter indicating the target texture and the second indicating the background texture.

The stimuli were created using the binary mask shown in Figure 8.2. The 48×48 white square is considered the target field, while the rest of the 256×256 image is considered to be the background field. Each stimulus was created by simply taking the background texture image and replacing the pixel values in the target field with the corresponding pixel values from the target texture image. The resulting 15 image stimuli are shown in Appendix C of this dissertation. Notice from Table 8.2

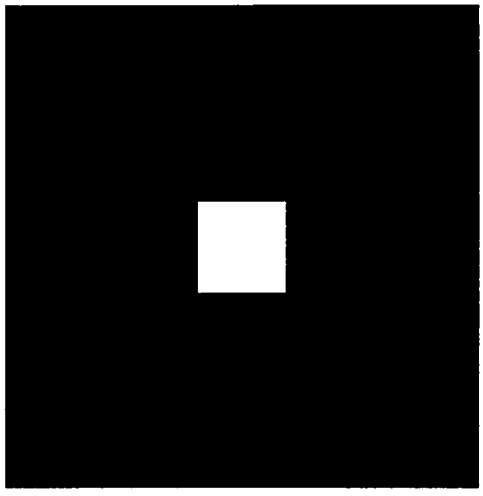
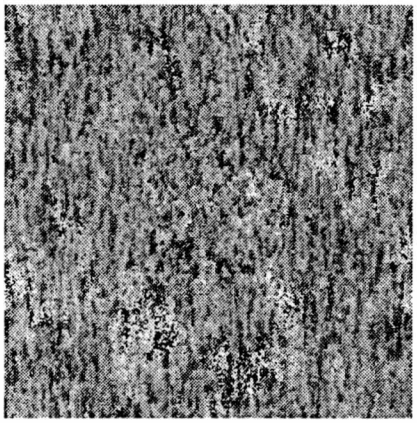


Figure 8.2: The binary mask showing the relative location of the 48×48 target field within each 256×256 background texture.

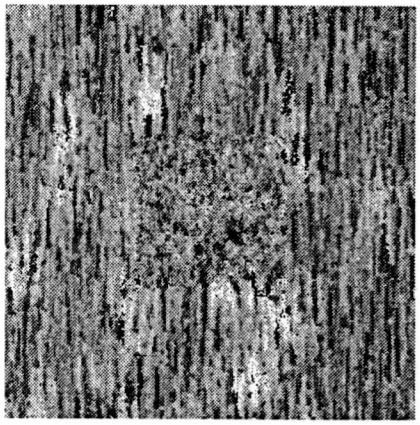
that we could have also included 15 more stimuli with reversed target/background relationships and another six with target and background of the same texture. That would have enabled a more exhaustive study, but a paired comparison experiment with $N = 36$ stimuli would require a total of $\binom{N}{2} = \frac{N!}{(N-2)!2!} = \frac{N(N-1)}{2} = \frac{(36)(35)}{2} = 630$ comparisons! By using $N = 15$ stimuli by the scheme indicated in Table 8.2, we can still study human discrimination of these textures with only $\frac{(15)(14)}{2} = 105$ comparisons.

8.2 Design and Conduct of the Experiment

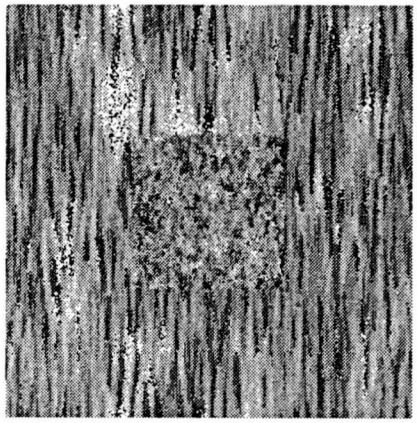
XPET was used to present each observer with all 105 possible pairs of the 15 stimuli, and for each pair the observer responds as to which of the two images has a target pattern that is more distinct from its background. A short training session is first conducted for each observer, to ensure that he understands what is required. This training session utilizes three of the stimuli from the hybrid texture



(a) Hybrid # 3 (0.8, 0.2)



(b) Hybrid # 6 (0.5, 0.5)



(c) Hybrid # 9 (0.2, 0.8)

Figure 8.3: The three 256×256 hybrid texture experiment stimuli that are used in the training session.

paired comparison experiment conducted previously. Briefly, these stimuli consist of target textures synthesized from a grass GLC model and background textures synthesized from GLC models of various grass and wood grain texture hybrids. These three stimuli are shown in Figure 8.3. The background of hybrid stimulus # 3 was synthesized from a GLC model that was weighted 0.8 for grass and only 0.2 for wood grain. The background for stimulus # 6 was an equally weighted (0.5, 0.5) hybrid, and the background for # 9 was weighted less for grass than for wood grain

(0.2, 0.8). Since the target in each stimulus was from a pure grass GLC model, the target in stimulus # 3 is expected to be the least distinct and the target in stimulus # 9 should be the most distinct of the three. In fact, the responses of all ten observers in the previous experiment agreed with this expectation. The training session for the current experiment consists of presenting all three possible pairs of the three images to the observer asking that he choose for each pair the image that has a target pattern that is more distinct from its background. Only if he makes the expected choice for all three pairs is he allowed to continue with the full experiment.

8.3 Results and Analysis of the Data

8.3.1 Estimation of Psychological Scale Values for the Stimuli

Table 8.3 is the raw data for the 15 stimuli used in the LCJ analysis. The element in row j and column k is the number of times (out of 20 observations) that stimulus k was judged greater than stimulus j. The LCJ scale values estimated for each of the 15 stimuli in the experiment are given in Table 8.4. This scale value estimation was performed as explained previously in Section 4.1.3. Figure 8.4 shows the relative locations of the scale values along the perceptual continuum representing target distinctness. These scale values indicate only relative amounts of target distinctness in the stimuli as judged by the observers, and have no absolute meaning. The stimulus containing the target judged least distinct was stimulus DF. This stimulus is assigned a scale value of 0.0, and the scale is constructed from that point. The stimuli containing the most distinct targets as judged by the observers were stimuli CF and CD. These two stimuli fell relatively far above their nearest neighbor on the

Table 8.3: The raw data for the 15 stimuli used in the LCJ analysis. The element in row j and column k is the number of times (out of 20 observations) that stimulus k was judged greater than stimulus j.

	AB	AC	AD	AE	AF	BC	BD	BE	BF	CD	CE	CF	DE	DF	EF
AB	0	14	0	11	6	11	8	14	13	20	3	20	10	0	15
AC	6	0	1	7	1	6	3	12	5	19	1	19	4	0	7
AD	20	19	0	19	17	16	16	18	20	20	12	20	16	6	19
AE	9	13	1	0	1	11	5	16	7	19	6	18	5	0	7
AF	14	19	3	19	0	16	13	18	15	20	9	20	16	0	19
BC	9	14	4	9	4	0	5	16	9	18	3	19	4	0	10
BD	12	17	4	15	7	15	0	18	19	20	6	19	14	0	13
BE	6	8	2	4	2	4	2	0	2	19	4	17	3	0	9
BF	7	15	0	13	5	11	1	18	0	20	3	20	5	0	14
CD	0	1	0	1	0	2	0	1	0	0	0	6	1	0	1
CE	17	19	8	14	11	17	14	16	17	20	0	20	13	2	11
CF	0	1	0	2	0	1	1	3	0	14	0	0	0	0	0
DE	10	16	4	15	4	16	6	17	15	19	7	20	0	0	9
DF	20	20	14	20	20	20	20	20	20	20	18	20	20	0	20
EF	5	13	1	13	1	10	7	11	6	19	9	20	11	0	0

scale, stimulus BE. The remaining stimuli are somewhat evenly spaced along the scale, except for a small cluster near 2.0, formed by stimuli BF, BC, and EF.

8.3.2 Measures of Image Boundary Strength

Since the target patterns in the image stimuli are not blended into their respective backgrounds as in the previous experiments with synthetic textures, it is essential that we include a quantitative measure of target/background boundary strength in our computational models. One way to accomplish this is to compute the average contrast between the pixels lying on either side of the target/background boundary. For a target field that is a rectangular lattice of pixels, such as in our experiment

Table 8.4: The scale values for the 15 stimuli.

Stimulus	LCJ scale
AB	1.72
AC	2.35
AD	0.70
AE	2.06
AF	0.99
BC	1.96
BD	1.29
BE	2.43
BF	1.90
CD	3.60
CE	1.10
CF	3.32
DE	1.58
DF	0.00
EF	1.99

stimuli, the lengths of the boundaries are $n_{top} = n_{bottom} \equiv n_{horiz}$ and $n_{left} = n_{right} \equiv n_{vert}$. For a single point i along a boundary, the contrast is

$$c(i) = |p_t(i) - p_b(i)|, \tag{8.1}$$

where $p_t(i)$ is the gray level of the pixel just on the target side of the boundary and $p_b(i)$ is the gray level of the adjacent pixel just on the background side. The average contrast for one boundary (such as the top boundary) is

$$C_{top} = \frac{1}{n_{horiz}} \sum_{i=1}^{n_{horiz}} c(i), \tag{8.2}$$

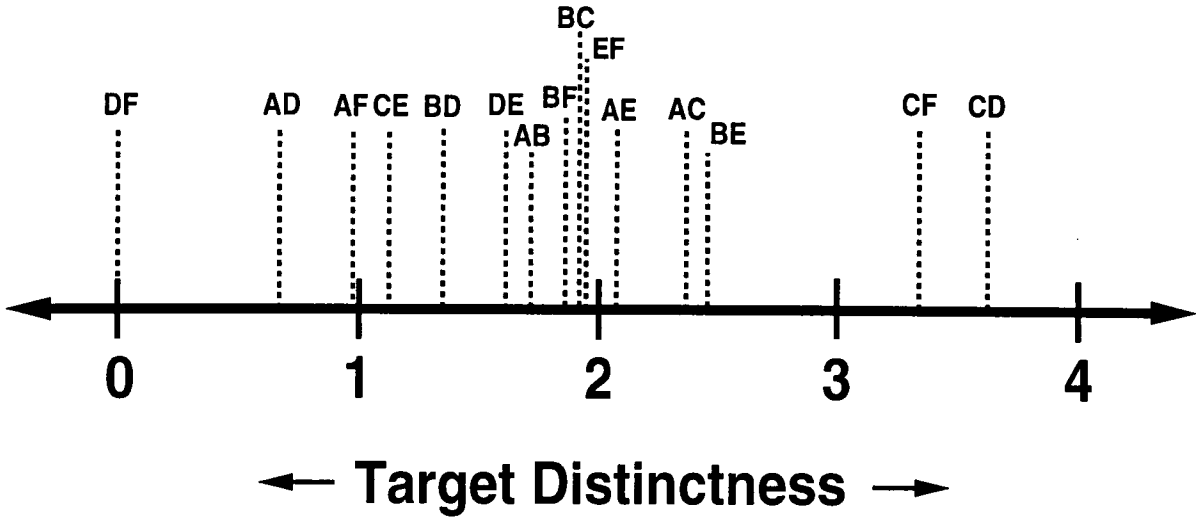


Figure 8.4: The relative locations of the scale values along the perceptual continuum representing target distinctness.

where i is just a summation index for the boundary points. Then the average boundary strength (ABS) for the whole target is:

$$ABS = \frac{1}{2n_{horiz} + 2n_{vert}}(n_{horiz}(C_{top} + C_{bottom}) + n_{vert}(C_{left} + C_{right})). \quad (8.3)$$

For a target field that is a perfect square, such as in our experiment stimuli, we have $n_{horiz} = n_{vert} \equiv n_{bound}$. In this case, the equation for ABS reduces to

$$ABS = \frac{1}{4n_{bound}} n_{bound}(C_{top}+C_{bottom}+C_{left}+C_{right}) = \frac{1}{4}(C_{top}+C_{bottom}+C_{left}+C_{right}). \quad (8.4)$$

The ABS measure does not take into account the values of any pixels that do not lie adjacent to the target/background boundary. However, a target/background boundary that has a high value for ABS may not be very distinct if it is embedded in a region that already is characterized by a large amount of contrast. To take into account the contrast of the entire region, we can use relative average boundary

strength (RABS):

$$RABS = \frac{ABS}{\frac{1}{n_{region}} \sum_{i=1}^{n_{region}} c(i)}, \tag{8.5}$$

where n_{region} is the number of adjacent pixel pairs in the entire region of the target. Essentially, RABS is the ratio of the average contrast along the target/background boundary to the average contrast in the region.

The ABS and RABS measures were computed for the targets in our 15 experiment stimuli. The targets are all 48 × 48 pixels in size, so $n_{bound} = 48$. The results are given in Table 8.5. In the calculation of RABS for each target, the region of the target was considered to be the target pixels plus all background pixels within 48 pixels of the target/background boundary. We can qualitatively judge these two boundary strength metrics by examining the images (Appendix C) and the computed values of the metrics. As an example, consider stimuli CD and CE. These two stimuli yielded relatively large values for ABS, with the value for CE a bit higher than the value for CD. But the computed value of RABS for stimulus CD is the highest over all the stimuli, while the value for CE is medium range. Examining the images, we see that indeed the target/background boundaries for stimulus CD are much more distinct than those for stimulus CE, since the high level of contrast in the background pixels near the target in stimulus CE tends to obscure the high-contrast boundaries.

Table 8.5: The computed values of average boundary strength (ABS) and relative average boundary strength (RABS) for the 15 real texture experiment stimuli.

Stimulus	ABS	RABS
AB	13.688	3.794
AC	32.458	5.165
AD	11.563	2.557
AE	26.750	2.893
AF	17.531	2.829
BC	24.057	4.187
BD	18.589	4.428
BE	28.401	3.210
BF	25.161	4.278
CD	35.182	6.090
CE	40.234	3.883
CF	40.193	5.403
DE	28.620	3.023
DF	16.734	2.623
EF	28.109	3.373

8.3.3 Evaluation of Individual Target Signature Metrics

Sample and Population Correlations

Now that we have perceptual judgments of the level of target distinctness in the stimuli, we can test the various metrics at our disposal to see how they relate to human perception. The correlation coefficient was computed between the vector of stimulus scale values and the vector of each of the target distinctness metrics. This value, represented by r, is the sample correlation coefficient for each metric. In the

previous discrimination experiments, we deterministically generated patterns with specified statistical properties to serve as stimuli. But for the present experiment, we have randomly selected patterns from actual image scenes of the real world, so we can consider our 15 stimuli to be a random sampling from the population of all possible target/background pairings. Thus we consider the sample correlation coefficient r to be an estimate of the population correlation coefficient ρ, which describes the relationship between each metric and the distinctness of all naturally occurring target/background patterns. This allows us to generate a confidence interval for ρ for each metric.

The method for deriving confidence levels for population correlation coefficients ρ is to transform the sampling distribution of r into a variable that is normally distributed for most values of ρ, if the sample size is larger than ten [Har70]. The random variable

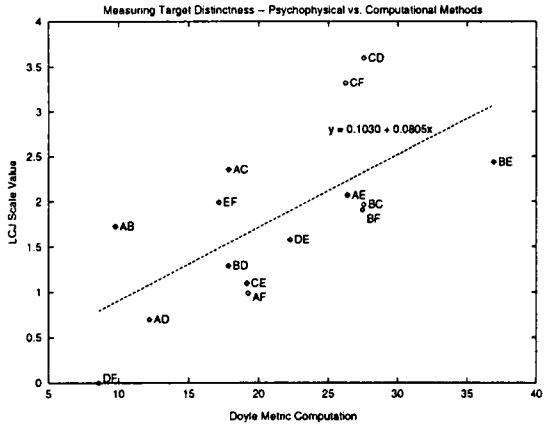
$$z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \tag{8.6}$$

is normally distributed with a mean of $\mu_r = \frac{1}{2} \ln[(1+\rho)/(1-\rho)]$ and standard deviation $\sigma_r = 1/\sqrt{N-3}$. The random variable z_r can be transformed into the standard normal:

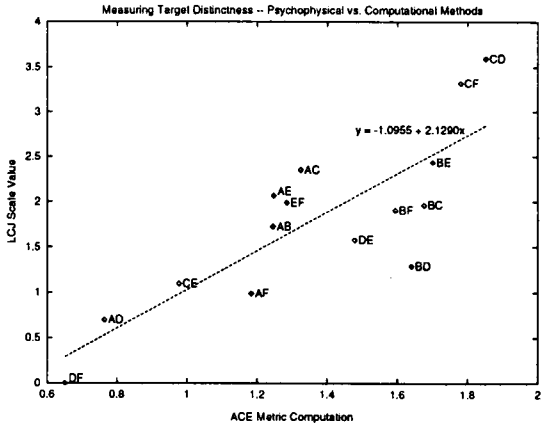
$$z = \frac{z_r - \mu_r}{\sigma_r}. \tag{8.7}$$

To find a confidence interval for ρ, a confidence interval is first calculated for μ_r, and then the interval is transformed into the equivalent interval for ρ by solving for ρ in the relationship $\mu_r = \frac{1}{2} \ln[(1+\rho)/(1-\rho)]$. The interval for μ_r is

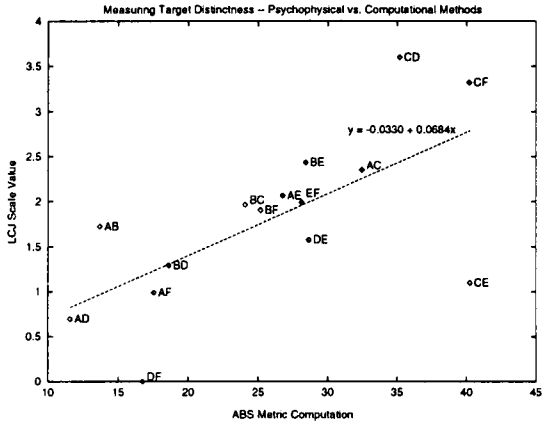
$$z_r - z_{\alpha/2}\sigma_r \leq \mu_r \leq z_r + z_{\alpha/2}\sigma_r. \tag{8.8}$$



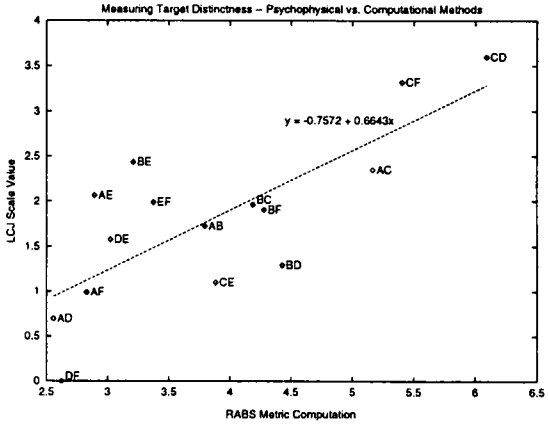
(a) LCJ scale values and the Doyle metric.



(b) LCJ scales and ACE metric.



(c) LCJ scales and ABS metric.



(d) LCJ scales and RABS metric.

Figure 8.5: The stimuli plotted with their LCJ scale values and computed values for various metrics.

The value of α is chosen such that $1 - \alpha$ is the desired level of confidence for the interval.

Table 8.6 gives the sample correlation coefficient (r) for each of the target distinctness metrics, as well as the corresponding 95% confidence interval ($\alpha = 0.05$) for the population correlation coefficient ρ for each metric. Figure 8.5 shows scatter plots of the values of the Doyle, ACE, ABS, and RABS metrics versus the LCJ scale values for the 15 stimuli. Also shown in each case is the best fitting line from

Table 8.6: The sample correlation coefficient (r) computed between the vector of stimulus scale values and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric.

Metric	r	95% Interval
Δ T	0.136	$-0.40 \leq \rho \leq 0.61$
Doyle	0.664	$0.23 \leq \rho \leq 0.88$
Doyle_{Mod}	0.486	$-0.04 \leq \rho \leq 0.80$
Doyle_{Log}	0.188	$-0.36 \leq \rho \leq 0.64$
Doyle_{Hybrid}	0.225	$-0.32 \leq \rho \leq 0.66$
Eff_POT	0.566	$0.08 \leq \rho \leq 0.84$
REDT	0.412	$-0.13 \leq \rho \leq 0.76$
ACE	0.825	$0.54 \leq \rho \leq 0.94$
WACE	0.827	$0.55 \leq \rho \leq 0.94$
AWCE	0.670	$0.24 \leq \rho \leq 0.88$
ASCE	0.325	$-0.22 \leq \rho \leq 0.72$
WASCE	0.362	$-0.18 \leq \rho \leq 0.74$
AWSCE	0.304	$-0.25 \leq \rho \leq 0.71$
VOE	-0.217	$-0.65 \leq \rho \leq 0.33$
ABS	0.654	$0.21 \leq \rho \leq 0.87$
RABS	0.763	$0.41 \leq \rho \leq 0.92$

a linear regression analysis.

In every case, the target statistics were computed from the pixels in the 48×48 square area, and the background statistics were computed from the pixels not in the target square but within 48 pixels of the target. The first seven metrics in the table would be considered first-order metrics. These metrics were all described and evaluated for synthetic images in Chapter 5, with the exception of the Eff_POT metric, an abbreviation for "effective pixels on target." This metric is one that has been used in previous studies [OWD93], and is computed as the number of pixels in the target pattern which have a gray level that differs from the mean gray level of the neighboring background pattern by more than two standard deviations of the background histogram. For each of the three Doyle metric derivatives, $Doyle_{Mod}$, $Doyle_{Log}$, and $Doyle_{Hybrid}$, the value used for the constant k is the same that yielded the highest correlation for the previous data, as described in Section 5.6. The next seven metrics, ACE through VOE, are the second-order metrics based upon GLC matrices. The final two metrics are the new boundary strength metrics.

The Δ T metric, which measures the contrast between the average value of the target pixels and the average value of the background pixels, yielded a relatively low correlation. This indicates that the distinctness of complex patterns cannot be adequately explained by simple contrast. The Doyle metric yielded a correlation of 0.664, which indicates that a great improvement over the Δ T metric can be achieved by taking into account differences in target and background variance. This was also evident from the results of the experiment with first-order statistics described in Chapter 5. The three Doyle metric derivatives, $Doyle_{Mod}$, $Doyle_{Log}$, and

$Doyle_{Hybrid}$, yielded significantly lower correlations than the Doyle metric. All three of these metrics contain a constant that was tuned to yield the highest correlation for the data in the experiment with first-order statistics. The poor performance of these metrics for this new data indicates that perhaps they are not as robust as the original Doyle metric.

The highest correlations among this set of target signature metrics are for ACE and WACE, which both yielded a correlation of about 0.83. The AWCE metric yielded a lower correlation than the ACE and WACE metrics. As explained when this metric was introduced in Chapter 6, the AWCE metric weights the cooccurrence differences inversely by the city-block distance of the matrix element from the diagonal of the GLC matrix. This type of weighting considers the errors in the GLC matrix elements closer to the diagonal to be more important. Since this metric did not perform as well as the ACE and WACE metrics, perhaps all the GLC matrix elements should be considered of equal importance. The three metrics in which GLC differences are squared before averaging, ASCE, WASCE, and AWSCE, all yielded much lower correlations with the psychological scale values. This is the first experiment in which we have seen a large difference between the squared and non-squared GLC metrics. This indicates that perhaps the distinctness between two natural textures is better measured using the absolute differences in GLC elements. The VOE metric, introduced previously in Section 7.2, yielded a very low and slightly negative correlation. This metric, also based on GLC matrices, was developed to measure differences in directionality content in two textures. It is evident that for this data set, directionality was not a strong cue for the observers.

Our two new boundary strength metrics, ABS and RABS, both yielded strong positive correlations with the psychological scale values. The RABS metric correlation was a bit higher, probably because it is beneficial to take into account the contrast of the entire region in order to judge the perceived distinctness of a boundary.

Intercorrelations and Significance of Differences

In addition to confidence intervals for the population correlations (ρ), it would be useful to test the significance of the difference in correlations for two metrics. For example, from Table 8.6 we see that although the sample correlation for the GLC-based metric ACE ($r = 0.825$) is higher than the sample correlation for the first-order Doyle metric ($r = 0.664$), the 95% confidence intervals for the respective population correlations (ρ) overlap. To effectively determine if the new GLC-based metric is indeed an improvement over the purely first-order metric, we need to know with what level of confidence can we say that the population correlation for ACE is indeed higher than the one for Doyle. To make this comparison, it is necessary to also know the level of intercorrelation between the two metrics. This can also give us some indication about how the metrics are related to each other. Table 8.7 gives the sample correlation coefficients computed between pairs of selected metrics.

The problem of whether one of the metrics has a genuinely higher population correlation than another is a hypothesis test. The null hypothesis (H_0) is that the population correlation for the new metric (ρ_2) is either the same as or lower than the population correlation of the existing metric (ρ_3). The alternative hypothesis

Table 8.7: The correlation coefficients computed between pairs of selected metrics.

	Doyle	Eff_POT	ACE	WACE	ABS	RABS
Doyle	1.0000	-0.0536	0.7579	0.7530	0.5421	0.3109
Eff_POT	-0.0536	1.0000	0.3381	0.3346	0.1773	0.5476
ACE	0.7579	0.3381	1.0000	0.9996	0.4580	0.6782
WACE	0.7530	0.3346	0.9996	1.0000	0.4679	0.6874
ABS	0.5421	0.1773	0.4580	0.4679	1.0000	0.5935
RABS	0.3109	0.5476	0.6782	0.6874	0.5935	1.0000

(H_1) is that $\rho_2 > \rho_3$. Hotelling [Hot40, GF78] has developed a t test that takes into account the intercorrelation between the two metrics:

$$t_{d_r} = (r_{12} - r_{13})\sqrt{\frac{(N-3)(1+r_{23})}{2(1-r_{23}^2-r_{12}^2-r_{13}^2+2r_{23}r_{12}r_{13})}}, \tag{8.9}$$

where r_{12} is the sample correlation for the new metric, r_{13} is the sample correlation for the existing metric, r_{23} is the intercorrelation between the two metrics, N is the number of stimulus samples, and $N-3$ is the degrees of freedom. For the ACE and Doyle metrics in this experiment, we have $r_{12} = 0.825$, $r_{13} = 0.664$, $r_{23} = 0.7579$, and $N = 15$. This results in a t score of $t_{d_r} = 1.4261$. From a table of the t-distribution [Har70] with 12 degrees of freedom, we see that the t score needed for $\alpha = 0.10$ is $t = 1.356$ and for $\alpha = 0.05$ is $t = 1.782$. Thus the null hypothesis H_0 is indeed rejected for a level of confidence of $1 - \alpha = 90\%$ but not quite for 95%. We can thus say with 90% confidence that the new ACE metric is more highly correlated with perceived target distinctness for the entire population of natural target/background combinations than the Doyle metric ($\rho_2 > \rho_3$).

8.3.4 Multivariable Linear Regression and Multiple Correlation

Since no attempt was made to equalize the histograms of the patterns when creating the stimuli, we are faced with a situation in which histogram differences, GLC differences, and target/background boundary strengths are all potential cues for the observers. We wish now to compare the psychological scale values to not one, but several variables. So we must extend our linear statistical analysis techniques to the case of multiple independent variables. In this section, we will briefly discuss and then apply the multivariable statistical methods that are commonly used.

The single variable linear regression model is of the form

$$y = \beta_0 + x\beta_1 + \epsilon, \tag{8.10}$$

where y is the response (dependent) variable, x is the independent variable, β_0 and β_1 are regression parameters, and ϵ is the error which is presumed to be normally distributed with mean of $\mu = 0$ and variance of σ^2. For our previous experiments, y represented the stimulus scale value estimated from the psychophysical data and x represented any one of the image metrics that we were studying. With N stimuli in the experiment, we actually have N samples of both y and x, so the entire model is written

$$\mathbf{y} = \beta_0 + \mathbf{x}\beta_1 + \epsilon, \tag{8.11}$$

where $\mathbf{y}' = (y_1, ..., y_N)$ represents the N scale values, $\mathbf{x}' = (x_1, ..., x_N)$ represents the N computed values of the particular image metric, and $\epsilon' = (\epsilon_1, ..., \epsilon_N)$ represents the error for each sample. Thus, when we performed the linear regression for a particular image metric, we solved for a value of β_0 and a value of β_1 that minimized the sum of the squared errors ϵ.

In the current experiment, we have k independent variables (image metrics) interacting simultaneously. Now, the model can be written

$$\mathbf{y} = \mathbf{X}'\beta + \epsilon, \tag{8.12}$$

where the differences are that β is a $k+1$ length vector of regression parameters and $\mathbf{X}$ is a rectangular matrix of computed image metrics with $k+1$ rows and N columns. (Actually, the first row of $\mathbf{X}$ consists of all 1's which are dummy variables so that the additive constant parameter β_0 is included.) The least squares solution for β is given by $\hat{\beta} = (\mathbf{X}\mathbf{X}')^{-1}$ [SC83].

Our other major statistical tool, correlation, can also be extended to multiple independent variables. We previously performed a simple correlation to measure the degree of linear association between two random variables. We can now utilize multiple correlation to measure the maximum correlation between the dependent variable and a linear combination of a set of independent variables. This will enable us to test the ability of various linear models for the human texture discrimination process to explain the empirical data. The multiple correlation was computed for selected pairs of metrics considered. This value is defined as the highest value of the correlation coefficient computed between the scale values and a linear combination of the two metrics. The results are given in Table 8.8. From the table, we see that of the pairs of metrics considered, Doyle and Eff_POT yielded the highest multiple correlation, at 0.897. This is despite the fact that the simple correlations between each these two metrics and the scale values were among the lowest of the six.

We can get some indication as to why this is so by considering the correlations between each of the metrics and each of the others, as given previously in Table 8.7.

Table 8.8: The multiple correlation coefficients for selected pairs of metrics.

	Doyle	Eff_POT	ACE	WACE	ABS	RABS
Doyle	-	0.897	0.827	0.829	0.751	0.886
Eff_POT	-	-	0.879	0.882	0.798	0.783
ACE	-	-	-	0.828	0.881	0.870
WACE	-	-	-	-	0.880	0.869
ABS	-	-	-	-	-	0.803
RABS	-	-	-	-	-	-

We see that the Doyle and Eff_POT are almost uncorrelated ($r = -0.0536$), while most other pairs are more strongly correlated. If two metrics are highly correlated with each other, then a linear combination of the two will not greatly improve the predictive capability of either considered singly. Although the ACE and WACE metrics yielded the highest single correlations with the scale values, they are strongly correlated with all of the metrics so that combining them with another metric does not greatly improve the results.

Additionally, we can use multiple correlation to test the effectiveness of various models consisting of linear combinations of more than two metrics to predict the psychological data. For this analysis, four metrics were selected as the most promising out of the six which were tested with pairwise correlations. These four metrics are assigned numerals 1-4 as follows:

1. Doyle

2. Eff_POT

Table 8.9: The multiple correlation coefficients and corresponding regression parameters for selected linear combinations of metrics.

Metrics	Max Correlation	Regression Parameters				
		β_0	β_1	β_2	β_3	β_4
1,2,3,4	0.941	-1.18e+00	6.58e-02	5.48e-04	1.11e-01	3.08e-01
1,2,3	0.914	-8.05e-01	5.34e-02	6.77e-04	8.50e-01	-
1,2,4	0.940	-1.16e+00	6.91e-02	5.57e-04	-	3.22e-01
1,3,4	0.892	-1.55e+00	4.03e-02	-	6.00e-01	4.37e-01
2,3,4	0.893	-1.20e+00	-	3.38e-04	1.51e+00	2.06e-01

3. ACE

4. RABS

The models tested are a linear combination of all four and every possible combination of three. The results of this are given in Table 8.9. In the table, the second column lists the value of the maximum correlation coefficient computed between the scale values and the linear combination of metrics in the first column. The remaining columns list the values of the regression parameters for which the model yields the maximum correlation value, corresponding to the $k + 1$ β parameters in Equation 8.12. In each case, the value listed in the β_0 column is the value of the additive constant parameter in the linear model for the optimum case. The values of these regression parameters do not absolutely indicate the relative importance of each metric in the model, since they provide both weighting and normalizing of the metrics. They are included simply to illustrate that although the maximum correlations for the models are rather high, their eventual utility depends on the

proper selection of values for several parameters.

8.4 Conclusions

A striking result of this latest experiment is that the single GLC-based error metrics ACE and WACE yield about 0.83 correlation coefficient compared to the psychological scale values for the stimuli. This is surprising because of the fact that the stimuli in the latest experiment consist of target and background patterns that differ in both first-order and second-order probabilities, and also that there was no attempt made to obscure abrupt edges along target/background boundaries. It seems that the GLC models for the patterns were able to capture at least some of all of these variables. But this is not so surprising if one considers the fact that second-order probabilities inherently contain first-order probabilities. Also, if two patterns have GLC models that are significantly different, it is apparent that a distinctly abrupt boundary is more likely if the two patterns are placed adjacent to each other.

When the metrics were considered two at a time, the highest correlation (0.897) was obtained for a linear combination of the Doyle metric with the Eff_POT metric. These two metrics were also found, in a previous experiment performed at the U.S. Army Night Vision and Electronic Sensors Directorate, to be the best predictors of the probability of finding low observable military targets in simulated infrared imagery [OWD93]. High correlations (0.87-0.89) were also obtained for a combination of one of the boundary measures with one of the GLC-based metrics and for the Doyle metric combined with RABS.

When combinations of three or four metrics were considered, a correlation of 0.940 resulted for the combination of Doyle, Eff_POT, and RABS. The inclusion of the GLC-based ACE metric does not significantly improve this result. Thus, it seems that for the stimuli and resulting psychological scale values in this experiment, it is best to use a GLC-based error metric if a single metric is desired as a measure of target distinctness. However, if we allow the inclusion of multiple metrics in the model, it is best to discard the GLC-based metric and instead use the Doyle, Eff_POT, and RABS metrics. The reason for this is that the GLC-based metrics are so strongly correlated with the other metrics that combining them with other metrics does not greatly improve their predictive capability. At any rate, before such a combination model can be used in practice, it will certainly be necessary to conduct further experimentation to either confirm the robustness of the regression parameters that were best for this experiment or to determine values that are the better for the particular imagery being used.

CHAPTER 9

Integrating Visual Search with Discrimination

9.1 Creation of the Image Scene Stimuli

The images used in the visual search experiment were extracted from a set of natural scenes of various locations in southern California. All of the images were obtained using a Nikon N8008s AF auto-focus 35mm camera fitted with a 70-210mm f/4-5.6 zoom lens and developed as 8 × 10in. color enlargements. The photographs were taken in midday in bright sunlight, using 100 and 200 speed Kodak film. The enlargements were digitized at 120 pixels per inch using a Hewlett Packard HP ScanJet 3C digital scanner. The scenes include a wide variety of both terrain and vegetation conditions such as forests, mountains, fields, and deserts. Great care was taken to ensure that no man-made objects or animals appear in the scenes. Also, the viewing perspective of each scene is such that the viewer is looking down from above, which is analogous to a military observer on high ground or in a low altitude aircraft. The viewing distance varies from as close as 100m to as far as several kilometers.

Ten 800 × 1200 images were selected from the database as representative of the wide variety of possible terrain and vegetation conditions. The color images were converted to gray scale by averaging the red, green, and blue channels. These ten raw images were used to create ten stimulus images according to the scheme

Table 9.1: The number of targets and the raw images used to contribute the target and background pixels for the ten stimuli in the search experiment.

Stimulus Image	Number of Targets	Raw Target Image	Raw Bkgnd. Image
A	5	b14	b18
B	5	b18	c15
C	4	c15	c19
D	6	c19	c20
E	6	c20	d9
F	5	d9	d11
G	4	d11	d12
H	6	d12	d13
I	5	d13	d22
J	6	d22	b14

in Table 9.1. For each stimulus, a random number of either four, five, or six was chosen for the number of targets. Every target was a square region 48 pixels on each side. A random location was chosen for each target square, with the restrictions that no target pixels could lie within 96 pixels, or two target dimensions, of the boundaries of the image, and no target pixels could lie within 144 pixels, or three target dimensions, of another target's pixels. If a target location was chosen that did not meet these two restrictions, it was discarded and another random location was chosen.

Once the number of targets and target locations for a particular stimulus were randomly chosen, the stimulus image was created by using the pixel values of the raw background image for all pixels except target pixels. The values for the target pixels were taken from the pixels in the raw target image at the corresponding locations. The resulting coordinate locations for all 52 targets in the ten stimuli are given in Appendix D. For each target, the row and column of the top left pixel is given, where the row coordinate in each image varies from 0 to 799 top to bottom and the column coordinate varies from 0 to 1199 left to right.

In Appendix E, the ten stimulus images as well as transparent images representing the target ground truth locations are exhibited. Also, some sample fixation point data is given for each stimulus, which will be explained later. From inspection of the stimulus images, one can see that the 52 targets represent a wide variety of possible patterns that are superimposed on a wide variety of background patterns.

9.2 Design and Conduct of the Experiment

The search experiment was conducted using ITEMS, which was described previously in Section 4.2. Data was collected from a total of 12 different observers, who were volunteers with various university backgrounds including students, faculty, and staff. The data collection procedure for every observer consisted of an explanation of what was required, a practice session, a review of the data from the practice session, and finally the experiment session.

The instructions given to every observer included an explanation of the importance that the observer's head movements be kept to a minimum during the data

collection. The proper procedure for controlling the display of image stimuli was then described, as well as what is required during the calibration process. The observer was then told that each of the image scenes contained between four and six targets each, and that every target is a square region of a specified size that contains a pattern which looks as if it doesn't belong in its location, in that it looks "unnatural" or "out of the ordinary." He was asked to identify each target as soon as he sees it, and to find as many of the targets in each image before proceeding to the next.

The practice session, conducted before the actual experiment session for every observer, was designed to ensure that the observer understands what is required and is comfortable with the procedure, that ITEMS is capable of segregating the observer's pupil during eye movements, that the observer has seen example stimulus images to be familiar with the nature of the targets, and that the observer is able to avoid excessive head movements that can create large errors. The practice session was identical to the actual experiment session, except that there were only three image scene stimuli presented instead of ten. The three practice image scenes were created in the exact same manner as the ten in the actual experiment, from different images in the same database.

After completion of the practice session, the raw data was processed to produce the fixation point estimates. Then the images were played back, with the fixation point estimates superimposed. This process allows the observer to see the data he provided, and the nature of any error caused by head movements. In addition, any targets he may have missed in the image scenes were pointed out by the experi-

menter.

The experiment session was conducted immediately after reviewing the practice data. The ten stimuli were presented to each observer in a different randomly chosen order. Together with five calibration images and four zero point images, each observer was presented a total of 19 images in the experiment. This typically required about 10-15 minutes, during which time the observer was required to hold his head still. The entire time required for data collection from a single observer was typically about one hour, which includes explanation of instructions, review of practice data, and both the practice and actual experiment sessions.

9.3 Results and Analysis of the Data

9.3.1 Target Fixation and Identification Statistics

The data provided by ITEMS for every observer consists of the fixation point coordinates in the display image and the corresponding timestamp for each sample, along with the timestamp and button identifier for every press of a mouse button during the session. Since the mouse button presses are the means by which the observer both controls the image display process and indicates he is fixating targets, and the locations of all targets in the image stimuli are known (Appendix D), this data is sufficient for computing various statistics describing the observer's process of search and discrimination of the targets. Since there is a certain amount of error in the fixation locations provided by ITEMS, it is necessary to establish a threshold for the maximum distance from a target center for a fixation point sample to be considered a fixation of the target. From extensive trial runs conducted during the

development of ITEMS, it was observed that the fixation point error, defined as the x- and y-distance in pixels from a fixation point estimate to the true location of the fixation, varies over spatial locations in the display image, over different observers, and over time for a single observer. But this error for any particular sample almost never exceeds 48 pixels in either x- or y-direction, which corresponds to one target dimension. Since it is important for our analysis to be able to determine the moment that an observer identifies a target, we don't want an unusually large amount of error at the exact moment of a mouse press to identify a target to cause us to consider it a false detection rather than a true one, we must consider the area of a target to extend a bit beyond its actual boundaries. For this reason, the area of a target in the display image was considered to be the pixels within the target square as well as all pixels within one and one half target dimensions outside the target square.

Table 9.2 gives the number of targets identified (out of 52) and the average distances in pixels of the fixation point estimates at the instant of identification for the 12 observers in the experiment. The number of false alarms given is the total over all ten stimuli. The raw fixation point data for observer 'a' is given in Appendix E. For each of the ten stimulus scenes, a transparent overlay is included to show the locations of the square target regions and another is included to show the observer's fixation points during the display of that image. Each sample point appears as a cross, with consecutive sample points connected by a straight line to indicate the eye movement. The fixation point at the instant the observer identified a target is shown as a solid block. These images are all oriented sideways, and

Table 9.2: The number of targets identified (out of 52) and the average distances in pixels of the fixation point estimates at the instant of identification for the 12 observers in the search experiment.

Observer	Targets Identified	Avg. Error (Pixels X)	Avg. Error (Pixels Y)	False Alarms
a	43	9.95	11.47	3
b	38	13.13	18.26	4
c	38	10.87	10.16	5
d	43	9.28	18.58	3
e	41	9.10	31.63	4
f	35	17.94	24.37	3
g	37	21.49	29.57	3
h	38	13.82	17.24	7
i	37	24.05	31.76	4
j	23	26.87	33.30	3
k	40	18.53	17.78	8
l	46	10.13	13.43	6

should be viewed from the right side.

Appendix F gives the computed statistics of identification probability (P_{ID}), average time to identification ($\overline{T_{<ID}}$), fixation probability (P_{fix}), average time to first fixation ($\overline{T_{<fix}}$), and average total fixation time ($\overline{T_{fix}}$) for the 52 targets in the search experiment. The computations of P_{ID} and P_{fix} are more properly the likelihood of identification and fixation for each target, as they are simply calculated as the proportion of the 12 observers that identified and fixated the target. The statistics $\overline{T_{<ID}}$ and $\overline{T_{<fix}}$ are computed as the time elapsed from the moment the image is

Table 9.3: The correlation coefficients between pairs of computed statistics from the search experiment.

	P_{ID}	$\overline{T_{<ID}}$	P_{fix}	$\overline{T_{<fix}}$	$\overline{T_{fix}}$
P_{ID}	1.0000	-0.6126	0.3741	-0.6256	0.1356
$\overline{T_{<ID}}$	-0.6126	1.0000	-0.3936	0.7883	0.3009
P_{fix}	0.3741	-0.3936	1.0000	-0.4921	0.3208
$\overline{T_{<fix}}$	-0.6256	0.7883	-0.4921	1.0000	-0.0089
$\overline{T_{fix}}$	0.1356	0.3009	0.3208	-0.0089	1.0000

first displayed until the observer first identifies or fixates the target, averaged over only those observers that did indeed identify or fixate the target. The statistic $\overline{T_{fix}}$ is computed as the total time the observer spent fixating the target area, averaged over all 12 observers.

Table 9.3 the correlation coefficients between pairs of computed statistics. From these calculations, we can see that there is a relatively strong negative correlation between P_{ID} and $\overline{T_{<ID}}$ of -0.6126, which seems reasonable since the more distinct a target is, the higher the probability it will be detected and the less time it will take to do so. There is a large positive correlation of 0.7883 between $\overline{T_{<ID}}$ and $\overline{T_{<fix}}$, which indicates that these two variables are closely related.

9.3.2 Evaluation of Target Signature Metrics

The various target signature metrics which were compared to the data from the previous discrimination experiments were computed for all 52 targets in the search experiment. The resulting values are given in Appendix G. For each calculation,

Table 9.4: The sample correlation coefficient (r) computed between the vector of computed identification statistics and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric.

	P_{ID}		$\overline{T_{<ID}}$	
	r	95% Interval	r	95% Interval
Δ T	0.2970	$0.03 \leq \rho \leq 0.53$	-0.5520	$-0.72 \leq \rho \leq -0.33$
Doyle	0.2983	$0.03 \leq \rho \leq 0.53$	-0.5434	$-0.71 \leq \rho \leq -0.32$
Eff_POT	0.3535	$0.09 \leq \rho \leq 0.57$	-0.4295	$-0.63 \leq \rho \leq -0.18$
ACE	0.4224	$0.17 \leq \rho \leq 0.62$	-0.6243	$-0.77 \leq \rho \leq -0.42$
WACE	0.4300	$0.18 \leq \rho \leq 0.63$	-0.6291	$-0.77 \leq \rho \leq -0.43$
AWCE	0.4404	$0.19 \leq \rho \leq 0.64$	-0.5986	$-0.75 \leq \rho \leq -0.39$
ASCE	0.4109	$0.16 \leq \rho \leq 0.61$	-0.4790	$-0.67 \leq \rho \leq -0.24$
WASCE	0.4111	$0.16 \leq \rho \leq 0.61$	-0.4786	$-0.67 \leq \rho \leq -0.24$
AWSCE	0.3963	$0.14 \leq \rho \leq 0.60$	-0.4425	$-0.64 \leq \rho \leq -0.19$
VOE	0.0855	$-0.19 \leq \rho \leq 0.35$	-0.3452	$-0.56 \leq \rho \leq -0.08$
ABS	0.2264	$-0.05 \leq \rho \leq 0.47$	-0.2524	$-0.49 \leq \rho \leq 0.02$
RABS	0.4331	$0.18 \leq \rho \leq 0.63$	-0.4340	$-0.63 \leq \rho \leq -0.18$

the background was considered to consist of all pixels not in the target square but within one target dimension. Table 9.4 gives the sample correlation coefficient (r) computed between the two vectors of computed identification statistics and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric.

Table 9.5: The sample correlation coefficient (r) computed between the vector of computed fixation statistics and the vector of each of the target distinctness metrics, as well as the corresponding 95% confidence interval for the population correlation coefficient ρ for each metric.

	P_{fix}		$\overline{T_{<fix}}$		$\overline{T_{fix}}$	
	r	95% Interval	r	95% Interval	r	95% Interval
Δ T	0.2793	$0.01 \leq \rho \leq 0.51$	-0.4680	$-0.66 \leq \rho \leq -0.22$	-0.3245	$-0.55 \leq \rho \leq -0.06$
Doyle	0.3387	$0.07 \leq \rho \leq 0.56$	-0.4901	$-0.67 \leq \rho \leq -0.25$	-0.2870	$-0.52 \leq \rho \leq -0.02$
Eff_POT	0.2439	$-0.03 \leq \rho \leq 0.48$	-0.3804	$-0.59 \leq \rho \leq -0.12$	-0.2507	$-0.49 \leq \rho \leq 0.02$
ACE	0.3128	$0.04 \leq \rho \leq 0.54$	-0.5551	$-0.72 \leq \rho \leq -0.33$	-0.3269	$-0.55 \leq \rho \leq -0.06$
WACE	0.3140	$0.05 \leq \rho \leq 0.54$	-0.5588	$-0.72 \leq \rho \leq -0.34$	-0.3263	$-0.55 \leq \rho \leq -0.06$
AWCE	0.3264	$0.06 \leq \rho \leq 0.55$	-0.5368	$-0.71 \leq \rho \leq -0.31$	-0.2838	$-0.52 \leq \rho \leq -0.01$
ASCE	0.3395	$0.07 \leq \rho \leq 0.56$	-0.4274	$-0.63 \leq \rho \leq -0.17$	-0.2271	$-0.47 \leq \rho \leq 0.05$
WASCE	0.3414	$0.08 \leq \rho \leq 0.56$	-0.4283	$-0.63 \leq \rho \leq -0.18$	-0.2250	$-0.47 \leq \rho \leq 0.05$
AWSCE	0.3336	$0.07 \leq \rho \leq 0.56$	-0.3996	$-0.61 \leq \rho \leq -0.14$	-0.2054	$-0.45 \leq \rho \leq 0.07$
VOE	0.0883	$-0.19 \leq \rho \leq 0.35$	-0.2137	$-0.46 \leq \rho \leq 0.06$	-0.2807	$-0.51 \leq \rho \leq -0.01$
ABS	0.1913	$-0.09 \leq \rho \leq 0.44$	-0.3080	$-0.54 \leq \rho \leq -0.04$	0.0908	$-0.19 \leq \rho \leq 0.36$
RABS	0.3535	$0.09 \leq \rho \leq 0.57$	-0.4195	$-0.62 \leq \rho \leq -0.17$	-0.0505	$-0.32 \leq \rho \leq 0.23$

The confidence intervals were calculated as discussed previously in Section 8.3.3. Table 9.5 gives the same information for the computed fixation statistics. From Table 9.4, we see that the strongest correlations for P_{ID} were obtained for the set of GLC-based texture metrics and for the boundary strength metric RABS. For $\overline{T_{<ID}}$, the three GLC-based metrics in which the differences are not squared performed better than for the other three. In addition, the first-order metrics Δ T and Doyle performed better than RABS. These trends are similar for the corresponding fixation statistics P_{fix} and $\overline{T_{<fix}}$ in Table 9.5. For $\overline{T_{fix}}$, the correlations are lower, but still highest for the GLC-based metrics and also for Δ T.

As discussed previously in Section 8.3.3, it can be useful to test the significance

Table 9.6: The correlation coefficients computed between pairs of selected metrics.

	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
Δ T	1.0000	0.9555	0.5188	0.8070	0.8036	0.7286	0.6377	0.6399	0.5989	0.2024	0.4534	0.5163
Doyle	0.9555	1.0000	0.4324	0.8787	0.8752	0.7957	0.6826	0.6825	0.6407	0.2278	0.5043	0.5635
Eff_POT	0.5188	0.4324	1.0000	0.5613	0.5641	0.5788	0.5359	0.5402	0.5019	0.0922	-0.0537	0.3734
ACE	0.8070	0.8787	0.5613	1.0000	0.9995	0.9638	0.8383	0.8351	0.7892	0.2967	0.3079	0.5888
WACE	0.8036	0.8752	0.5641	0.9995	1.0000	0.9664	0.8430	0.8401	0.7941	0.3087	0.3041	0.5948
AWCE	0.7286	0.7957	0.5788	0.9638	0.9664	1.0000	0.9360	0.9329	0.9014	0.4153	0.2233	0.6920
ASCE	0.6377	0.6826	0.5359	0.8383	0.8430	0.9360	1.0000	0.9996	0.9949	0.4586	0.1767	0.7348
WASCE	0.6399	0.6825	0.5402	0.8351	0.8401	0.9329	0.9996	1.0000	0.9952	0.4563	0.1771	0.7354
AWSCE	0.5989	0.6407	0.5019	0.7892	0.7941	0.9014	0.9949	0.9952	1.0000	0.4468	0.1758	0.7367
VOE	0.2024	0.2278	0.0922	0.2967	0.3087	0.4153	0.4586	0.4563	0.4468	1.0000	-0.1567	0.4065
ABS	0.4534	0.5043	-0.0537	0.3079	0.3041	0.2233	0.1767	0.1771	0.1758	-0.1567	1.0000	0.5535
RABS	0.5163	0.5635	0.3734	0.5888	0.5948	0.6920	0.7348	0.7354	0.7367	0.4065	0.5535	1.0000

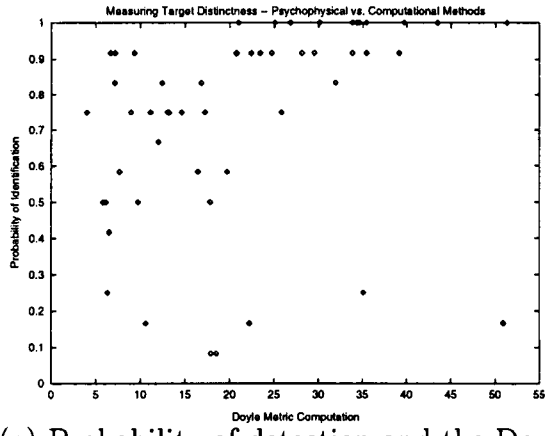
of the difference in correlations for two metrics. To make such a comparison, it is necessary to know the level of intercorrelation between the two metrics. This can also give us some indication about how the metrics are related to each other. Table 9.6 shows the results of this. The Hotelling t test [Hot40, GF78], given previously in Equation 8.9, allows us to test the hypothesis that one of the metrics has a genuinely higher population correlation (ρ) than another. This type of test was applied to three different pairs of metrics for the P_{ID} statistic and one pair for the $\overline{T_{<ID}}$ statistic. For P_{ID} and the ACE and Doyle metrics in this experiment, we have $r_{12} = 0.4224$, $r_{13} = 0.2983$, $r_{23} = 0.8787$, and $N = 52$. This results in a t score of $t_{d_r} = 1.9743$. From a table of the t-distribution [Har70] with $N-3 = 49$ degrees of freedom, we see that this t score is sufficient to pass a $1-\alpha$ test for $\alpha = 0.05$ but not quite sufficient for $\alpha = 0.025$. In fact, with as many as $N = 52$ samples, the t-distribution closely resembles the standard normal distribution [Har70]. Using the standard normal table with $z = 1.9743$ tells us that the null hypothesis should be rejected with 97% probability, meaning that we can say with about 97% confidence that the new

ACE metric is more highly correlated with P_{ID} for the entire population of natural target/background combinations than the Doyle metric ($\rho_2 > \rho_3$). For the ABS and RABS metrics and P_{ID}, the result is $t_{d_r} = 1.6990$, meaning that we can also say with 95% confidence that RABS is more highly correlated with P_{ID} than is ABS for the entire population. For RABS and Doyle and P_{ID}, the result is $t_{d_r} = 1.1234$, meaning that the value of ρ for RABS is higher than for Doyle with about 86% confidence. For $\overline{T_{<ID}}$, ACE and Doyle were tested. Since these correlations are negative, the resulting value for t_{d_r} is also negative, at $t_{d_r} = -1.4722$. This means that we can say with 93% confidence that the ACE is more strongly negatively correlated with $\overline{T_{<ID}}$ for the population. Figure 9.1 through Figure 9.4 show plots of the 52 targets in the search experiment, with the horizontal axis representing the computed value of a target signature metric and the vertical axis representing the measured value of one of the search statistics.

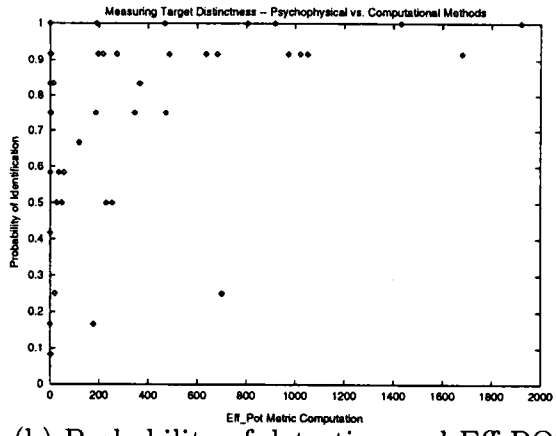
Next we consider pairs of metrics, computing the multiple correlation between each possible pair of metrics and each of the five computed search statistics. The multiple correlation is defined as the maximum correlation coefficient computed between the search statistic and a linear combination of the metrics. The results of these computations for each of the five search statistics are given in Table 9.7 through Table 9.11.

9.4 Conclusions

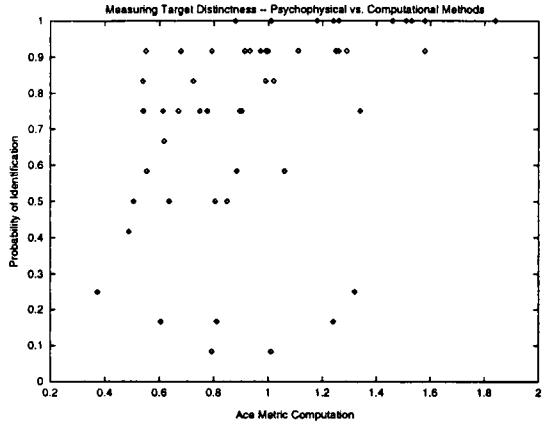
We have found (Table 9.4) that the magnitudes of the correlations between the individual target signature metrics and the probability of identification (P_{ID}) were



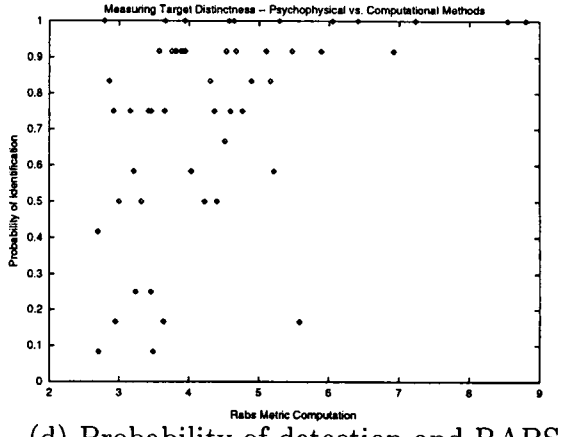
(a) Probability of detection and the Doyle metric.



(b) Probability of detection and Eff_POT metric.



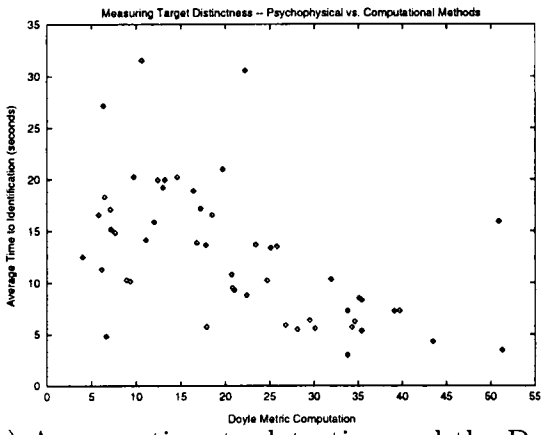
(c) Probability of detection and ACE metric.



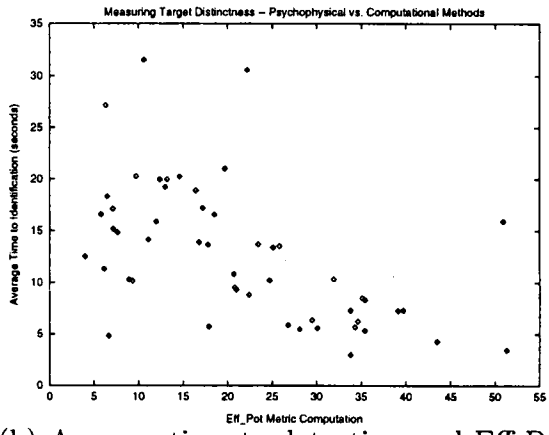
(d) Probability of detection and RABS metric.

Figure 9.1: The 52 targets in the search experiment plotted with their probability of detection and the computed values for various metrics.

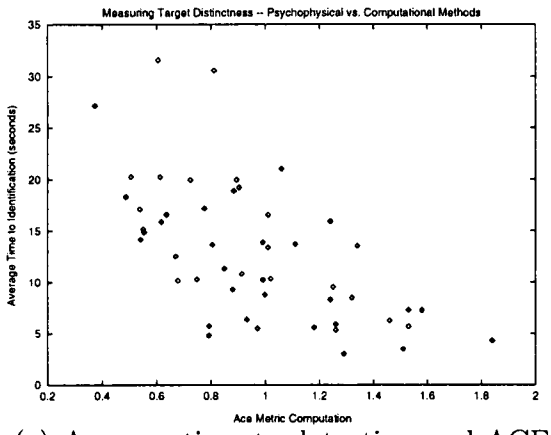
as high as 0.44 and for average time to identification ($\overline{T_{<ID}}$) as high as 0.63. Although these values are not quite as high as the correlation magnitudes between the metrics and the psychological scalings of stimuli in "target distinctness" in the paired comparison discrimination experiments previously conducted, we do see strong relationships. We must remember that in all of the previous experiments, at least



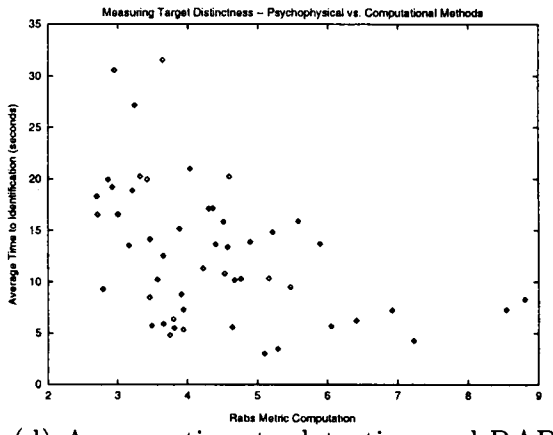
(a) Average time to detection and the Doyle metric.



(b) Average time to detection and Eff_POT metric.



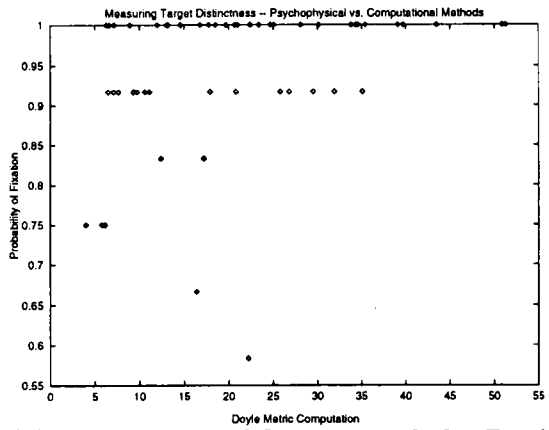
(c) Average time to detection and ACE metric.



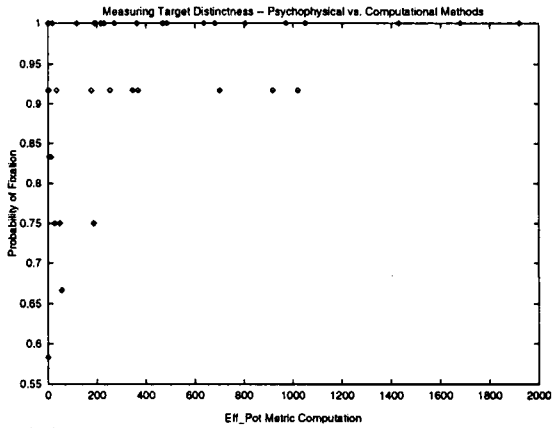
(d) Average time to detection and RABS metric.

Figure 9.2: The 52 targets in the search experiment plotted with their average time to detection and the computed values for various metrics.

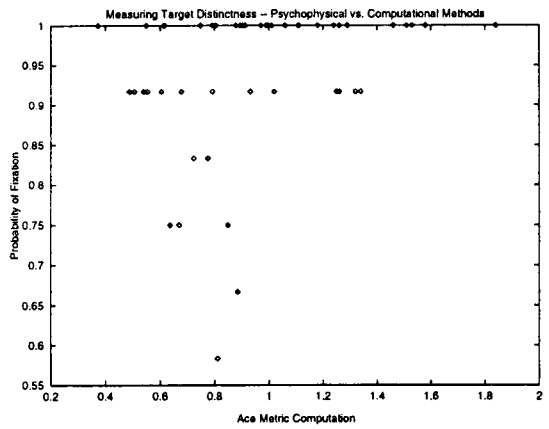
some attempt was made to have the target and background fields each consist of a homogeneous pattern, which would correspond to concept of stationary signals that have the same statistics over time or space. The search experiment is the only one in which the locations of targets was chosen randomly, so occasionally the target pattern or its local background pattern will consist of a homogeneous texture pat-



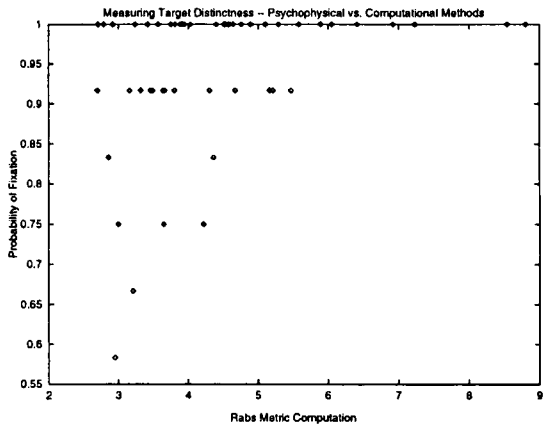
(a) Probability of fixation and the Doyle metric.



(b) Probability of fixation and Eff_POT metric.



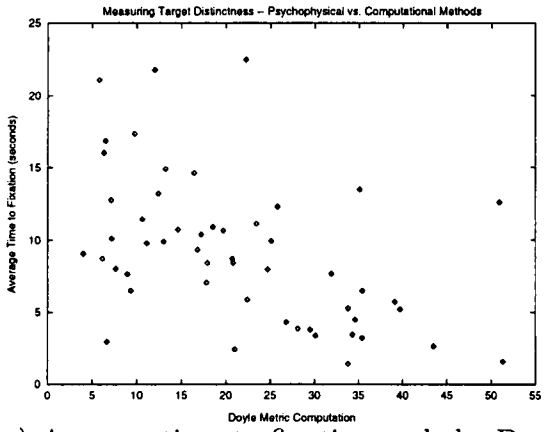
(c) Probability of fixation and ACE metric.



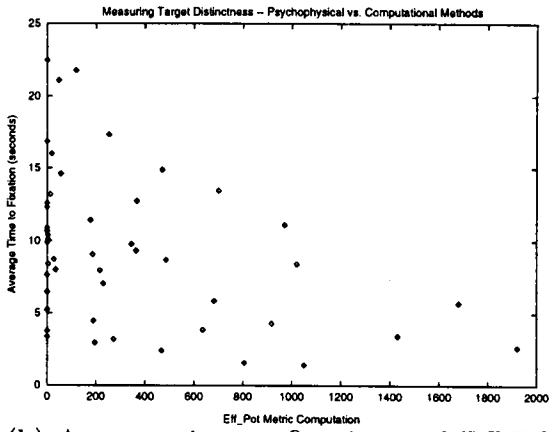
(d) Probability of fixation and RABS metric.

Figure 9.3: The 52 targets in the search experiment plotted with their probability of fixation and the computed values for various metrics.

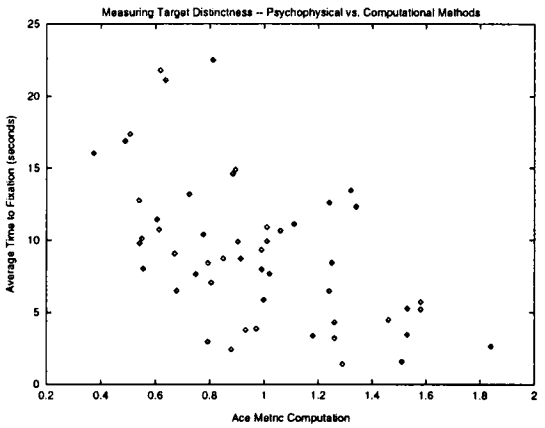
tern. Any of the metrics computed from a field should characterize the field best if it is homogeneous. We must also realize that there are many more variables contributing to whether an observer identifies a target and the time required to locate a target than just the distinctness of the target. In this respect, it seems that search combined with discrimination is a much more complicated task than pure discrim-



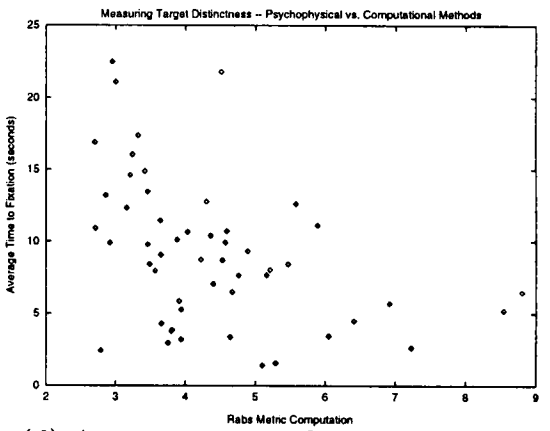
(a) Average time to fixation and the Doyle metric.



(b) Average time to fixation and Eff_POT metric.



(c) Average time to fixation and ACE metric.



(d) Average time to fixation and RABS metric.

Figure 9.4: The 52 targets in the search experiment plotted with their average time to fixation and the computed values for various metrics.

ination. It is also important to realize that even if there is a direct relationship between two variables, the computed value of a correlation coefficient between them will not be as high if the relationship is not linear.

It also is evident that despite the complicated nature of the processes, the GLC-based texture metrics, particularly the three metrics in which cooccurrence error is

Table 9.7: The multiple correlation coefficients computed between each possible pair of metrics and P_{ID} computed from the search experiment data.

	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
Δ T		0.3010	0.3776	0.4289	0.4377	0.4418	0.4134	0.4135	0.4032	0.2981	0.3143	0.4415
Doyle			0.3885	0.4492	0.4593	0.4488	0.4116	0.4118	0.4005	0.2988	0.3110	0.4381
Eff_POT				0.4452	0.4505	0.4567	0.4402	0.4398	0.4347	0.3574	0.4305	0.4799
ACE					0.4938	0.4405	0.4351	0.4355	0.4347	0.4245	0.4344	0.4801
WACE						0.4408	0.4393	0.4397	0.4394	0.4329	0.4416	0.4833
AWCE							0.4404	0.4404	0.4404	0.4533	0.4596	0.4749
ASCE								0.4111	0.4290	0.4269	0.4396	0.4541
WASCE									0.4313	0.4268	0.4397	0.4541
AWSCE										0.4093	0.4271	0.4479
VOE											0.2574	0.4443
ABS												0.4334
RABS												

Table 9.8: The multiple correlation coefficients computed between each possible pair of metrics and $\overline{T_{<ID}}$ computed from the search experiment data.

	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
Δ T		0.5547	0.5769	0.6296	0.6339	0.6221	0.5761	0.5756	0.5695	0.6013	0.5520	0.5788
Doyle			0.5847	0.6244	0.6292	0.6088	0.5632	0.5630	0.5571	0.5890	0.5440	0.5650
Eff_POT				0.6316	0.6355	0.6072	0.5209	0.5199	0.5033	0.5279	0.5105	0.5211
ACE					0.6460	0.6244	0.6296	0.6291	0.6296	0.6464	0.6275	0.6297
WACE						0.6301	0.6362	0.6357	0.6360	0.6488	0.6323	0.6335
AWCE							0.6416	0.6384	0.6392	0.6080	0.6109	0.5993
ASCE								0.4791	0.5852	0.4994	0.5085	0.4941
WASCE									0.5891	0.4994	0.5080	0.4937
AWSCE										0.4722	0.4768	0.4705
VOE											0.4642	0.4717
ABS												0.4343
RABS												

not squared (ACE, WACE, and AWCE), are the best overall metrics at predicting the five measured search statistics. The RABS metric, which measures boundary strength, shows some promise also, but not as much as the texture metrics. The various first-order metrics that are commonly used for target signature evaluation did not yield correlations as high as for the second-order metrics. It is also clear

Table 9.9: The multiple correlation coefficients computed between each possible pair of metrics and P_{fix} computed from the search experiment data.

	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
Δ T		0.3707	0.3023	0.3161	0.3173	0.3320	0.3491	0.3505	0.3480	0.2812	0.2886	0.3711
Doyle			0.3555	0.3402	0.3407	0.3515	0.3697	0.3708	0.3712	0.3389	0.3396	0.3918
Eff_POT				0.3235	0.3243	0.3333	0.3473	0.3486	0.3451	0.2526	0.3184	0.3735
ACE					0.3161	0.3265	0.3434	0.3451	0.3432	0.3128	0.3283	0.3765
WACE						0.3264	0.3434	0.3451	0.3432	0.3142	0.3298	0.3763
AWCE							0.3404	0.3421	0.3388	0.3305	0.3483	0.3712
ASCE								0.3482	0.3419	0.3478	0.3647	0.3725
WASCE									0.3472	0.3497	0.3664	0.3734
AWSCE										0.3404	0.3598	0.3697
VOE											0.2257	0.3587
ABS												0.3536
RABS												

Table 9.10: The multiple correlation coefficients computed between each possible pair of metrics and $\overline{T_{<fix}}$ computed from the search experiment data.

	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
Δ T		0.4901	0.4949	0.5562	0.5597	0.5484	0.4970	0.4971	0.4911	0.4835	0.4802	0.5120
Doyle			0.5245	0.5552	0.5588	0.5468	0.5063	0.5066	0.5026	0.5012	0.4951	0.5199
Eff_POT				0.5613	0.5644	0.5435	0.4635	0.4635	0.4505	0.4206	0.5029	0.4839
ACE					0.5699	0.5552	0.5595	0.5588	0.5587	0.5575	0.5735	0.5669
WACE						0.5590	0.5647	0.5640	0.5635	0.5605	0.5773	0.5692
AWCE							0.5775	0.5733	0.5709	0.5369	0.5704	0.5409
ASCE								0.4291	0.4963	0.4278	0.4883	0.4548
WASCE									0.5064	0.4287	0.4889	0.4552
AWSCE										0.4016	0.4669	0.4404
VOE											0.4065	0.4222
ABS												0.4293
RABS												

that combining pairs of metrics did not improve the correlations by any significant amount for any of the five measured search statistics, as was found for the previous discrimination experiment (Chapter 8).

As for the experimental methodology of this last experiment, it seems that the study of human search and discrimination is greatly improved by measuring the

Table 9.11: The multiple correlation coefficients computed between each possible pair of metrics and $\overline{T_{fix}}$ computed from the search experiment data.

	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
Δ T		0.3338	0.3385	0.3427	0.3427	0.3318	0.3256	0.3253	0.3248	0.3918	0.4202	0.3521
Doyle			0.3195	0.3269	0.3263	0.3013	0.2902	0.2898	0.2884	0.3623	0.3959	0.3170
Eff_POT				0.3369	0.3361	0.3030	0.2737	0.2724	0.2671	0.3602	0.2624	0.2550
ACE					0.3273	0.3473	0.3381	0.3384	0.3380	0.3793	0.3839	0.3711
WACE						0.3486	0.3383	0.3387	0.3381	0.3772	0.3824	0.3720
AWCE							0.3042	0.3046	0.3068	0.3356	0.3249	0.3485
ASCE								0.2391	0.3042	0.3017	0.2632	0.2846
WASCE									0.2934	0.3011	0.2612	0.2818
AWSCE										0.2946	0.2425	0.2538
VOE											0.2847	0.2892
ABS												0.1512
RABS												

eye fixations of observers. Although we can always have an observer report his judgments of a scene, knowledge of the eye fixation points provides us with invaluable insights into the process through which the observer reached his decisions. The search experiment that was conducted in this manner has provided us with data that can be used to develop or test models describing any portion of the search and discrimination processes. Not only do we have fixation data that include two-dimensional image coordinates, but also a third dimension of time, which allows us to include this dimension in the model.

CHAPTER 10

Concluding Remarks

10.1 Review of Accomplishments

This dissertation is a study of the use of quantitative image metrics to measure the signature strength of a camouflaged target in a sensed image. Various metrics were evaluated by comparison to the results of experiments in which human observers perform the tasks of search and discrimination with digital imagery. The real-world analogy of these experiments is that of a military observer looking upon the sensed image of a static scene to find camouflaged enemy targets that are reported to be in the area. The perceptual experiments were conducted so as to learn how to measure the fundamental characteristics of target and background patterns that serve as important discrimination cues for humans and to increasingly resemble the conditions of the real-world analogy.

To generalize the wide variety of potential target and background patterns in a sensed image, a computational model for image texture was utilized. Experiments were performed to study the perceptual qualities of a texture in relationship to the shape of the second-order probability distributions within its gray level cooccurrence (GLC) matrices. A texture synthesis technique was used to generate realizations of image texture possessing specified second-order probabilities. The GLC matrices of these synthesized textures were displayed so as to study the information content of

the matrices and to compare this to the perceptual qualities of the texture. Based on the results of these experiments, several GLC-based texture metrics were created as an attempt to measure the distinctness of two texture patterns.

Several unique computer testbeds for investigating human visual perception were developed. These testbeds allow for the conduct of paired comparison, triadic comparison, and eye tracking experiments. In the comparison experiments, image stimuli were synthesized with target and background patterns possessing specific first-order and/or second-order statistics, with minimized target/background boundaries. These experiments resulted in psychological scale values describing the relative amount of "target distinctness" in the stimuli. Various existing first-order image metrics and the newly developed GLC-based second-order metrics which attempt to describe target signature strength were evaluated by computing the correlations between computed values of the metrics and the psychological scale values.

The metrics shown to be so effective for stimuli that are synthesized with specified statistics were then evaluated with a paired comparison experiment using stimuli that are composed of natural target and background patterns that were extracted from actual images of outdoor scenes. In these stimuli, there were simultaneous differences in first-order and higher-order probabilities, and also no minimization of the target/background boundaries. The data from this experiment allowed comparison of the relative effectiveness of first-order and second-order metrics as well as two metrics developed to measure the strength of the target/background boundary.

Finally, the experimental methodology was expanded to integrate the task of search with discrimination. In the search experiment, observers were asked to view

images of real-world outdoor scenes that contain targets in random spatial locations. The observer's fixation point in the image is measured by tracking his eye movements during the experiment. From the data, various statistics were computed for each target to describe how easily the observers located it, including the likelihood the target was fixated or identified and the time required to do so, and the target signature metrics were compared to these statistics. This final experiment is the one most closely resembling the real-world analogy.

10.2 Summary of Findings

From the perceptual experiments in this dissertation, we have discovered several cues that humans use when performing search and discrimination of a particular target pattern in a background. Certainly differences in target and background first-order pixel probabilities are important, since they represent pattern contrast and variation. But second-order probabilities are important too, since they better represent the general concept of texture by taking into account the spatial relationships between pixels. It appears that the directionality of textures is an important cue in the discrimination of some textures. And the strength of the boundary between target and background is important also. This was shown previously by some simple demonstrations by Muller [Mul86], and this dissertation has found that the strength of this cue is not absolute but depends on the boundary contrast relative to the strength of intensity changes in the region. It is also evident that the cues that are utilized vary with the imagery, and even with different targets in the same imagery.

Of the first-order metrics considered, it is evident that the Doyle metric (Equation 5.7) generally outperforms the standard contrast, or Δ T metric (Equation 5.2), which is commonly used in sensor performance modeling. The Doyle metric accounts for differences in target and background variance, which is evidently important as a cue in addition to simple contrast. The various Doyle derivative metrics showed some promise, but did not prove to be robust in their performance over the series of experiments. For measuring the boundary strength cue, it was shown that the new RABS metric (Equation 8.5) is more effective than simple average boundary strength (ABS). But overall, no single metrics were as successful as the new GLC-based second-order metrics such as ACE (Equation 6.1) in predicting the psychophysical data. These metrics appear to be able to represent the distinctness of two image textures. This dissertation demonstrates that the GLC-based metrics hold great promise for application to the fields of camouflage development and evaluation as well as automatic target recognition.

10.3 Relevance to Other Research

The work most closely related to this dissertation is that of Rotman *et al.* [RTK94, RKG94, TRA+94] and that of O'Kane *et al.* [OWD93]. Rotman et al. [RTK94] consider measures of target contrast as compared to qualitative judgments of images of outdoor scenes of rural New York state. Later, Tidhar *et al.* [TRA+94] presented 12 of the images, each with a synthetically placed target, to observers and compared the values of contrast measures to the average detection time. Only the contrast of the target pixels was controlled. Rotman *et al.* [RKG94] conducted

an experiment with images of simulated scenes, comparing the popular fixation points in the scene to the value of metrics computed within square blocks of the image. O'Kane *et al.* [OWD93] used the same database of simulated scenes to compare first-order metrics computed for targets to the probability of observers finding them. These previous studies are all attempts at finding image metrics that correspond to human target acquisition. In every study in which targets were used, the case of possible camouflage was never considered. In this dissertation, we have allowed our targets to appear as any randomly chosen naturally occurring pattern, which generalizes the problem and includes the possibility of camouflage. And although there have been previous attempts to measure boundary discrimination separately from field discrimination [OWD93], this is the first work in which a model for texture was used in modeling the target and background fields.

Although there are several different models for texture being utilized in the research community, we have found GLC probabilities to serve well for modeling target and background patterns. We have also seen that GLC models generalize some other texture analysis techniques, such as Gibbs/Markov random fields [EP94]. This can be seen by comparing Figure 2.2, which shows the spatial displacement definitions for GLC probabilities, to Figure 2.1, which shows a typical spatial neighborhood for defining Gibbs/Markov probabilities. We also know that in several studies to compare the relative power of various texture analysis techniques to perform automated texture discrimination, GLC matrices generally outperformed other methods [WDR76, CH80, OD92]. And although most researchers who want to mimic the physiological processes in human preattentive vision utilize spatial frequency chan-

nels, it is possible that the human vision system generates something analogous to GLC matrices when performing preattentive discrimination [CN89].

10.4 Future Research Directions

The most direct potential application of the image metrics that quantify target signature strength would be in the synthesis of new camouflage patterns and the assessment of existing camouflage effectiveness. The metrics could be particularly useful for "active" camouflage systems of the future, where a camouflage pattern is created in the field. This "site specific" camouflage would be designed for the particular location of the vehicle or structure. From the results of our experiments, we have seen that such a system would definitely need the ability to perform some type of blending at its edges, so that abrupt changes don't serve as strong cues. Also, to be most useful to the commanders in the field, the metrics that have shown such promise at predicting the probability of detecting the target and the time required to do so must eventually be related to the variable of detection range. This information would be necessary for tactical planning.

Computer assisted targeting would be another potential use for the metrics. In such a system, the computer would assist the soldier in the search process, by automatically identifying candidates for attention. This would be especially useful since the visual displays used by tank gunners have a rather narrow field of view. In this regard, sensor bandwidth isn't wasted on areas of the scene that likely do not contain potential targets. At any rate, it is likely that any targeting system actually deployed in the future will include a human in the loop at some point. The human

will still most likely perform the discrimination process, as it will be his ultimate decision to pull the trigger.

BIBLIOGRAPHY

BIBLIOGRAPHY

[AK89] M. Amadasun and R. King. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-19(5):1264–1274, 1989.

[Ant84] H. Anton. *Calculus with Analytic Geometry.* John Wiley and Sons, New York, 1984.

[BA89] D. Blostein and N. Ahuja. Shape from texture: Integrating texture-element extraction and surface estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1233–1251, 1989.

[BCG90] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12:55–73, 1990.

[BL91] J. R. Bergen and M. S. Landy. Computational Modeling of Visual Texture Segregation. In M. S. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing.* MIT Press, Cambridge, MA, 1991.

[Bro66] P. Brodatz. *Textures - A Photographic Album for Artists and Designers.* Dover Publications, New York, 1966.

[Cal89] T. C. Callaghan. Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception and Psychophysics*, 46(4):299–311, 1989.

[Car92] J. M. Carstensen. *Description and Simulation of Visual Texture.* PhD thesis, Technical University of Denmark, 1992.

[CCB85] R. Chellappa, S. Chatterjee, and R. Bagdazian. Texture synthesis and compression using gaussian-markov random field models. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:298–303, 1985.

[CH80] R. W. Conners and C. A. Harlow. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(3):204–222, May 1980.

[CH95] J. Cartier and D. Hsu. Human visual search – a two state process. In G. C. Holst, editor, *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing VI*, 1995. Proc. SPIE 2470.

[Cha86] W. G. Chase. Visual Information Processing. In K. R. Boff, editor, *Handbook of Perception and Human Performance, Volume II*, chapter 28. John Wiley and Sons, New York, 1986.

[CJ78] T. Caelli and B. Julesz. On perceptual analyzers underlying visual texture discrimination: Part 1. *Biol. Cybernetics*, 28:167–175, 1978.

[CJ83] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1):25–39, 1983.

[CJG78] T. Caelli, B. Julesz, and E. Gilbert. On perceptual analyzers underlying visual texture discrimination: Part 2. *Biol. Cybernetics*, 29:201–214, 1978.

[CN89] R. W. Conners and C. T. Ng. Developing a quantitative model of human preattentive vision. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1384–1407, 1989.

[CTM96] A. C. Copeland, M. M. Trivedi, and J. R. McManamey. Evaluation of image metrics for target discrimination using psychophysical experiments. *Optical Engineering*, 35(6):1714–1722, June 1996.

[DD94] M. Dubuisson and R. C. Dubes. Efficacy of fractal features in segmenting images of natural textures. *Pattern Recognition Letters*, 15:419–431, April 1994.

[DH73] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1973.

[Dod87] N. Dodd. Multispectral texture synthesis using fractal concepts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9:703–707, 1987.

[EP94] I. M. Elfadel and R. W. Picard. Gibbs random fields, cooccurrences, and texture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):24–37, Jan. 1994.

[Fal86] J. C. Falmagne. Psychophysical Measurement and Theory. In K. R. Boff, editor, *Handbook of Perception and Human Performance, Volume I.* John Wiley and Sons, New York, 1986.

[FS89] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological Cybernetics*, 61:103–113, 1989.

[FVPBGG87] A. R. Figueiras-Vidal, J. M. Paez-Borrallo, and R. Garcia-Gomez. On using cooccurrence matrices to detect periodicities. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(1):114–116, Jan. 1987.

[Gag81] A. Gagalowicz. A new method for texture fields synthesis: Some applications to the study of human vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(5):520–533, Sept. 1981.

[GDM85] A. Gagalowicz and Song De Ma. Sequential synthesis of natural textures. *Computer Vision, Graphics, and Image Processing*, 30:289–315, 1985.

[Ges84] G. A. Gescheider. *Psychophysics : Method, Theory, and Applications.* L. Erlbaum, Bergenfield, N.J., 1984.

[GF78] J. P. Guilford and B. Fruchter. *Fundamental Statistics in Psychology and Education.* McGraw-Hill, New York, 1978.

[GK90] C. C. Gotlieb and H. E. Kreyszig. Texture descriptors based on co-occurrence matrices. *Computer Vision, Graphics, and Image Processing*, 51:70–86, 1990.

[GW92] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Addison-Wesley, Reading, MA, 1992.

[Har70] D. L. Harnett. *Introduction to Statistical Methods.* Addison-Wesley, Reading, MA, 1970.

[Har79] R. M. Haralick. Statistical and structural approaches to texture. *Proc. of the IEEE*, 67(5):786–804, May 1979.

[Hot40] H. Hotelling. The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Mathematical Statistics*, 11:271–283, 1940.

[HSA92] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen. Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):474–481, 1992.

[HSD73] R. M. Haralick, K. Shanmugan, and I. H. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610–621, Nov. 1973.

[HTCP86] C. A. Harlow, M. M. Trivedi, R. W. Conners, and D. Phillips. Scene analysis of high resolution aerial scenes. *Optical Engineering*, 25(3):347–355, March 1986.

[HW90] D. H. Hubel and T. N. Wiesel. Brain Mechanisms of Vision. In I. Rock, editor, *The Perceptual World: Readings from Scientific American.* W. H. Freeman and Co., New York, 1990.

[JB83] B. Julesz and J. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal*, 62(6):1619–1645, 1983.

[JGV78] B. Julesz, E. N. Gilbert, and J. D. Victor. Visual discrimination of textures with identical third-order statistics. *Biol. Cybernetics*, 31:137–140, 1978.

[Juh86] M. Juhola. A syntactic method for analysis of saccadic eye movements. *Pattern Recognition*, 19(5):353–359, 1986.

[Jul62] B. Julesz. Visual pattern discrimination. *IRE Trans. Information Theory*, 8(2):84–92, Feb. 1962.

[Jul81] B. Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biol. Cybernetics*, 41:131–138, 1981.

[Loh94] G. Lohmann. Co-occurrence-based analysis and synthesis of textures. In *12th IAPR International Conference on Pattern Recognition (ICPR)*, volume 1, pages 449–453, Jerusalem, Israel, Oct. 1994.

[Mar76] D. Marr. Early processing of visual information. *Phil. Trans. of Royal Soc. of London.*, 275(942):483–524, Oct. 1976.

[MJ92] J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2):173–188, 1992.

[MRR+53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Chem. Phys.*, 21:1087–1092, 1953.

[MSM81] J. Monne, F. Schmitt, and D. Massaloux. Bidimensional texture synthesis by markov chains. *Computer Graphics and Image Processing*, 17:1–23, 1981.

[Mul86] M. J. Muller. Texture boundaries: Important cues for human texture discrimination. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 464–468, Miami Beach, FL, June 1986.

[MW94] J. R. McManamey and G. W. Walker. Predicting target detectability using a mu/sigma response surface. In *Proc. of the 5th Annual Ground Target Modeling and Validation Conf.*, 1994.

[MWD93] J. R. McManamey, G. W. Walker, and B. Dolling. A mu/sigma response surface for the human visual system. In *Proc. of the 4th Annual Ground Target Modeling and Validation Conf.*, pages 15–25, 1993.

[Nic94] J. Nicoll. A mathematical framework for an improved search model. Technical Report P-2901, Institute for Defense Analyses, Oct. 1994.

[OD92] P. P. Ohanian and R. C. Dubes. Performance evaluation for four classes of textural features. *Pattern Recognition*, 25(8):819–833, 1992.

[OWD93] B. L. O'Kane, C. P. Walters, and J. D'Agostino. Report on perception experiments in support of low observables thermal performance models. Technical report, U.S. Army, Night Vision and Electronic Sensors Directorate, Fort Belvoir, VA, 1993.

[PE92] R. W. Picard and I. M. Elfadel. Structure of aura and co-occurrence matrices for the gibbs texture model. *Journal of Mathematical Imaging and Vision*, 2(1):5–25, 1992.

[PEP91] R. W. Picard, I. M. Elfadel, and A. P. Pentland. Markov/gibbs texture modeling: Aura matrices and temperature effects. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 371–377, Lahaina, Maui, Hawaii, June 1991.

[PM95] R. W. Picard and T. P. Minka. Vision texture for annotation. Technical Report 302, M.I.T. Media Laboratory, 1995. Perceptual Computing Section.

[PSO90] J. Parkinnen, K. Selkainaho, and E. Oja. Detecting texture periodicity from the cooccurrence matrix. *Pattern Recognition Letters*, 11:43–50, Jan. 1990.

[Rao90] A. Ravishankar Rao. *A Taxonomy for Texture Description and Identification.* Springer-Verlag, New York, 1990.

[RKG94] S. R. Rotman, M. L. Kowalczyk, and V. George. Modeling human search and target acquisition performance: Fixation-point analysis. *Optical Engineering*, 33(11):3803–3809, November 1994.

[RKT94] G. Ravichandran, E. J. King, and M. M. Trivedi. Texture synthesis: A multiresolution approach. In *Proc. of the Ground Target Modeling and Validation Conf.*, Houghton, MI, 1994. Keweenaw Research Center.

[RL93] A. R. Rao and G. L. Lohse. Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing*, 55(3):218–233, May 1993.

[RTK94] S. R. Rotman, G. Tidhar, and M. L. Kowalczyk. Clutter metrics for target detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, 30(1):81–91, Jan. 1994.

[RW90] T. R. Reed and H. Wechsler. Segmentation of textured images and gestalt organization using spatial and spatial-frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12:1–12, 1990.

[SB92] H. C. Shen and C. Y. C. Bie. Feature frequency matrices as texture image representation. *Pattern Recognition Letters*, 13:195–205, March 1992.

[SBG89] A. Sutter, J. Beck, and N. Graham. Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception and Psychophysics*, 46(4):312–332, 1989.

[SC83] M. S. Srivastava and E. M. Carter. *An Introduction to Applied Multivariate Statistics.* North-Holland, New York, 1983.

[SS93] H. Sari-Sarraf. *Multiscale Wavelet Representation and its Application to Signal Classification.* PhD thesis, Univ. of Tennessee, Knoxville, 1993.

[ST92] M. V. Shirvaikar and M. M. Trivedi. Developing texture-based image clutter measures for object detection. *Optical Engineering*, 31:2628–2639, Dec. 1992.

[SW83] D. E. Schmieder and M. R. Weathersby. Detection performance in clutter with variable resolution. *IEEE Transactions on Aerospace and Electronic Systems*, 19(4):622–630, July 1983.

[SYTN93] L. Stark, I. Yamashita, G. Tharp, and H. X. Ngo. Keynote Lecture: Search Patterns and Search Paths in Human Visual Search. In D. Brogan, A. Gale, and K. Carr, editors, *Visual Search 2*, chapter 2. Taylor and Francis, London, 1993.

[THCG84] M. M. Trivedi, C. A. Harlow, R. W. Conners, and S. Goh. Object detection based on gray level cooccurrence. *Computer Vision, Graphics, and Image Processing*, 28:199–219, 1984.

[Thu27] L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34:273–286, 1927.

[TJ90] M. Tuceryan and A. K. Jain. Texture segmentation using voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:211–216, 1990.

[TJ93] M. Tuceryan and A. K. Jain. Texture analysis. In C. H. Chen, L. F. Pau, and P. S. P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*, pages 235–276. World Scientific, Singapore, 1993.

[TMY78] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-8(6):460–473, June 1978.

[Tor58] W. S. Torgerson. *Theory and Methods of Scaling.* John Wiley and Sons, New York, 1958.

[TRA+94] G. Tidhar, G. Reiter, Z. Avital, Y. Hadar, S. R. Rotman, V. George, and M. L. Kowalczyk. Modeling human search and target acquisition performance iv.: Detection probability in the cluttered environment. *Optical Engineering*, 33(3):801–808, March 1994.

[TT90] F. Tomita and S. Tsuji. *Computer Analysis of Visual Textures.* Kluwer Academic Publishers, Boston, 1990.

[VE84] R. E. Vasquez-Espinosa. *A Study of the Texture Measurement Definition Problem.* PhD thesis, Louisiana State Univ., 1984.

[WC92] C. Wu and Y. Chen. Statistical feature matrix for texture analysis. *CVGIP: Graphical Models and Image Processing*, 54(5):407–419, Sept. 1992.

[WDR76] J. S. Weszka, C. R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 6:269–285, Apr. 1976.

[WWHL88] G. Waldman, J. Wooten, G. Hobson, and K. Luetemeyer. A normalized clutter measure for images. *Computer Vision, Graphics, and Image Processing*, 42(2):137–156, May 1988.

[Yar67] A. L. Yarbus. *Eye Movements and Vision.* Plenum Press, New York, 1967.

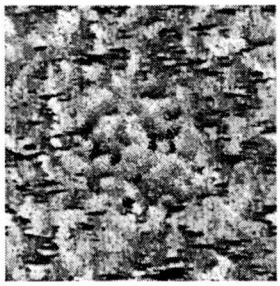
[YH79] R. Yokoyama and R. M. Haralick. Texture pattern image generation by regular markov chain. *Pattern Recognition*, 11:225–233, 1979.

[Zuc76] S. W. Zucker. Toward a model of texture. *Computer Graphics and Image Processing*, 5:190–202, 1976.

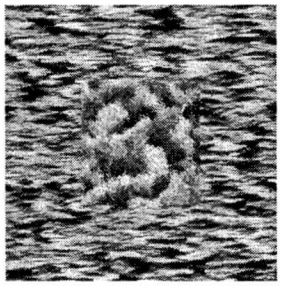
APPENDICES

APPENDIX A

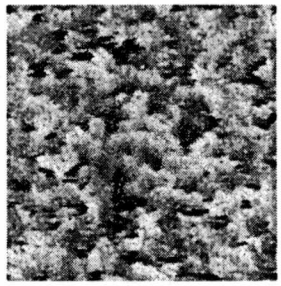
The 42 Triadic Comparisons Experiment Stimuli



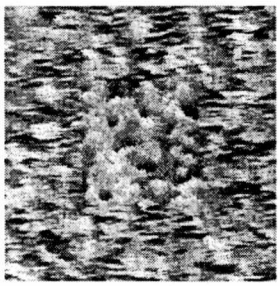
Stimulus 1_on_2



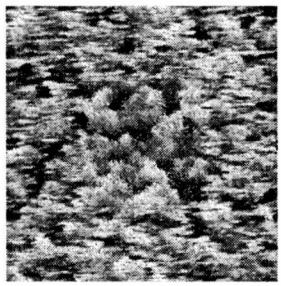
Stimulus 1_on_3



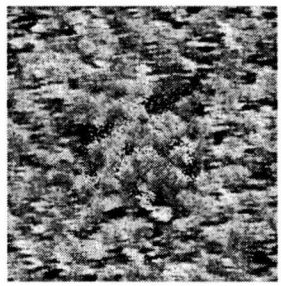
Stimulus 1_on_4



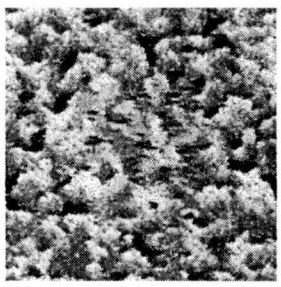
Stimulus 1_on_5



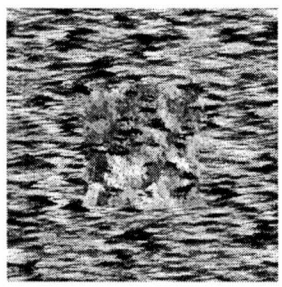
Stimulus 1_on_6



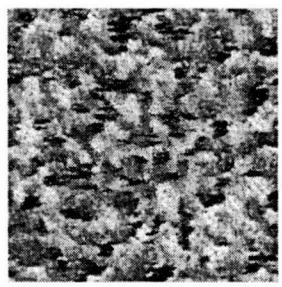
Stimulus 1_on_7



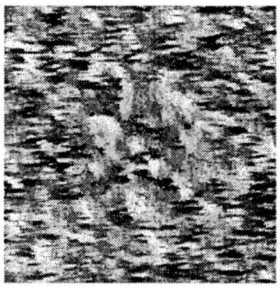
Stimulus 2_on_1



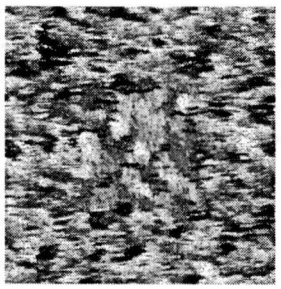
Stimulus 2_on_3



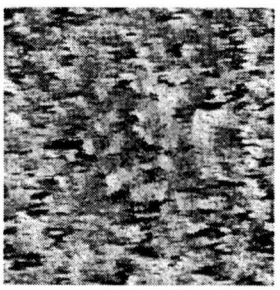
Stimulus 2_on_4



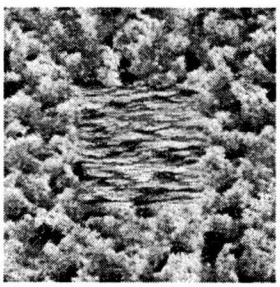
Stimulus 2_on_5



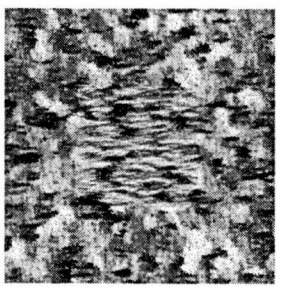
Stimulus 2_on_6



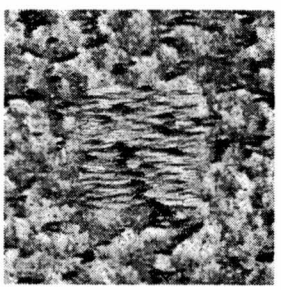
Stimulus 2_on_7



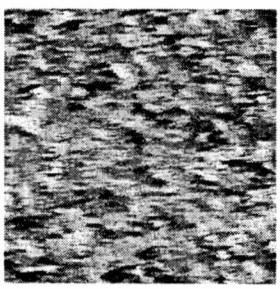
Stimulus 3_on_1



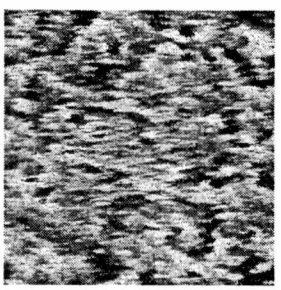
Stimulus 3_on_2



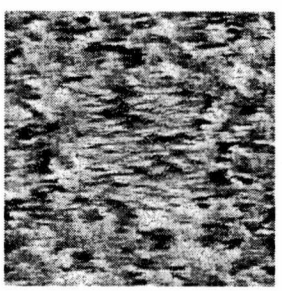
Stimulus 3_on_4



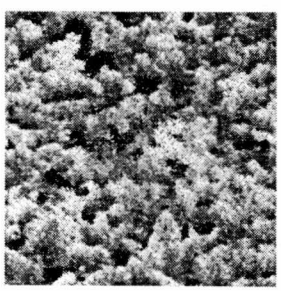
Stimulus 3_on_5



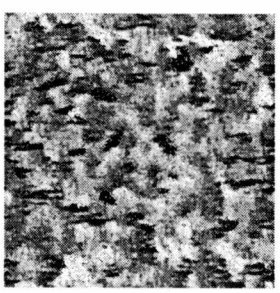
Stimulus 3_on_6



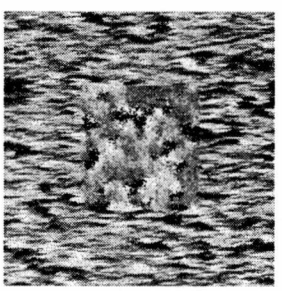
Stimulus 3_on_7



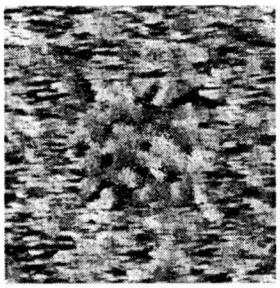
Stimulus 4_on_1



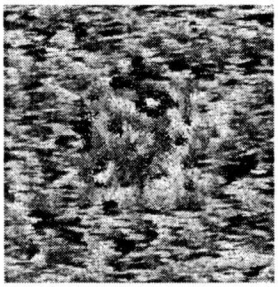
Stimulus 4_on_2



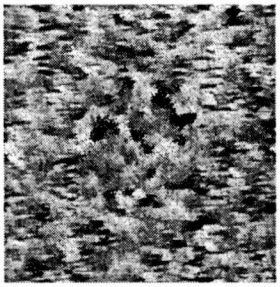
Stimulus 4_on_3



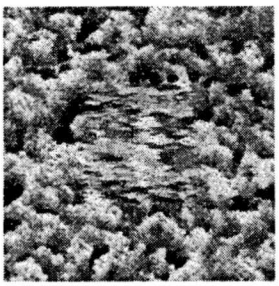
Stimulus 4_on_5



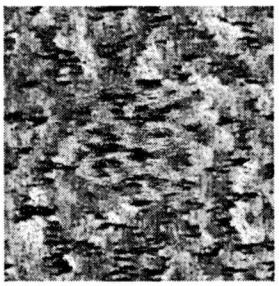
Stimulus 4_on_6



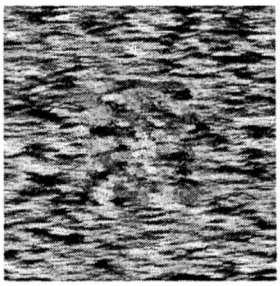
Stimulus 4_on_7



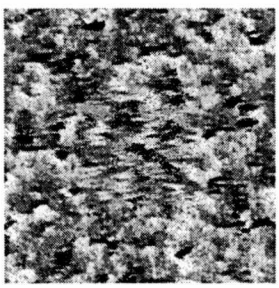
Stimulus 5_on_1



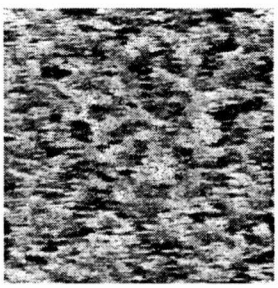
Stimulus 5_on_2



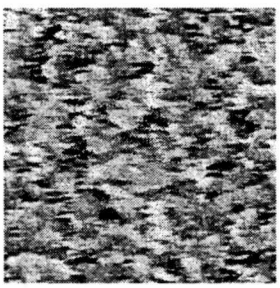
Stimulus 5_on_3



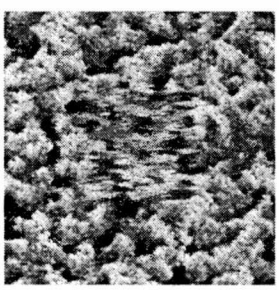
Stimulus 5_on_4



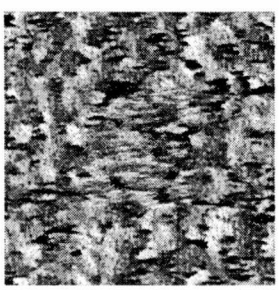
Stimulus 5_on_6



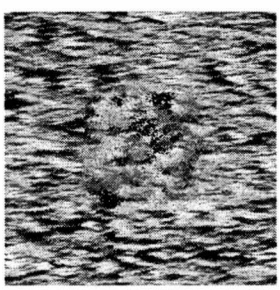
Stimulus 5_on_7



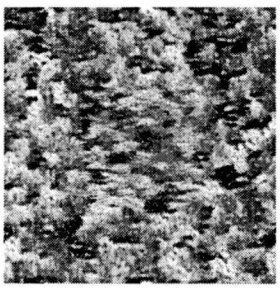
Stimulus 6_on_1



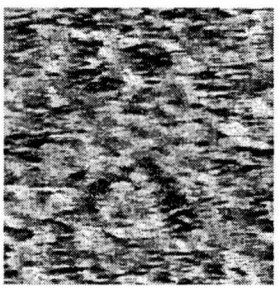
Stimulus 6_on_2



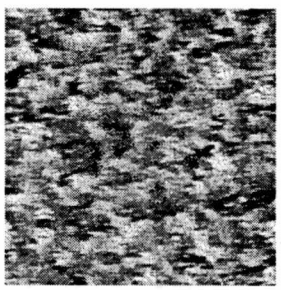
Stimulus 6_on_3



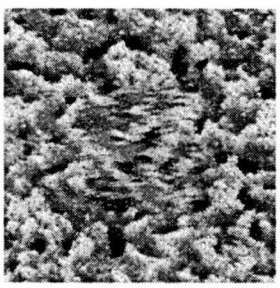
Stimulus 6_on_4



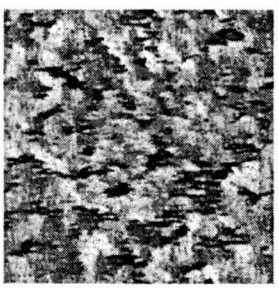
Stimulus 6_on_5



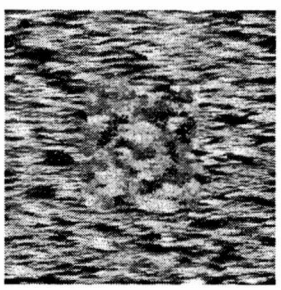
Stimulus 6_on_7



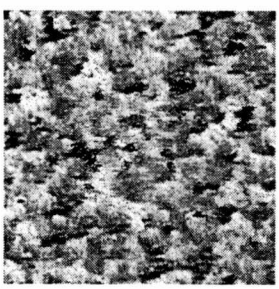
Stimulus 7_on_1



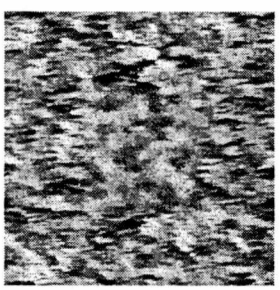
Stimulus 7_on_2



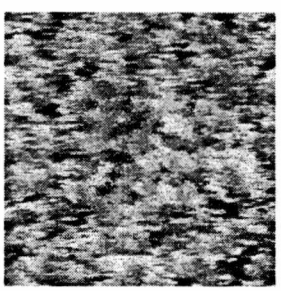
Stimulus 7_on_3



Stimulus 7_on_4



Stimulus 7_on_5



Stimulus 7_on_6

APPENDIX B

The Raw Decision Data from the Triadic Comparisons Experiment

	Target	Left Bkgnd.	Right Bkgnd.	Left Picks	Right Picks	ACE (L.-R.)	ACE R/W	VOE (L.-R.)	VOE R/W
1	7	3	4	15	0	9.7799e-02	1	4.5729e-04	1
2	6	4	1	2	13	-1.7630e-02	1	-1.4190e-05	1
3	1	2	5	1	14	-3.3030e-02	1	-2.6129e-04	1
4	3	1	6	15	0	1.6464e-01	1	2.2222e-04	1
5	4	6	1	15	0	5.1737e-02	1	1.9384e-04	1
6	5	3	1	6	9	-7.6630e-02	1	1.4298e-04	-1
7	1	7	4	15	0	5.3687e-02	1	1.2992e-04	1
8	4	7	5	0	15	-6.6879e-02	1	-1.5006e-04	1
9	5	1	6	13	2	1.2825e-01	1	2.2222e-04	1
10	7	2	4	8	7	3.2669e-02	1	-1.8690e-05	-1
11	7	5	2	15	0	-5.5863e-02	-1	3.8830e-05	1
12	2	3	4	15	0	1.9872e-01	1	6.7975e-04	1
13	3	2	4	1	14	6.4000e-04	-1	-1.8690e-05	1
14	7	3	6	15	0	1.3025e-01	1	5.0910e-04	1
15	4	6	3	0	15	-1.4629e-01	1	-5.0910e-04	1
16	6	3	1	12	3	-5.0000e-05	-1	2.8688e-04	1
17	4	5	1	15	0	6.9357e-02	1	2.6579e-04	1
18	3	2	5	15	0	1.4705e-01	1	2.6129e-04	1
19	6	4	5	13	2	5.1744e-02	1	1.3608e-04	1
20	5	6	4	1	14	-6.9364e-02	1	-2.0803e-04	1
21	6	7	4	0	15	-8.1707e-02	1	-1.2992e-04	1
22	2	6	4	14	1	1.0050e-01	1	1.7065e-04	1
23	2	3	5	15	0	1.4689e-01	1	4.3715e-04	1
24	4	3	5	15	0	1.2867e-01	1	4.3715e-04	1
25	4	2	7	1	14	-2.5300e-03	1	-1.1123e-04	1

	Target	Left Bkgnd.	Right Bkgnd.	Left Picks	Right Picks	ACE (L.-R.)	ACE R/W	VOE (L.-R.)	VOE R/W
26	6	1	5	15	0	6.9374e-02	1	1.5027e-04	1
27	2	6	3	0	15	-9.8220e-02	1	-5.0910e-04	1
28	5	4	7	14	1	9.0073e-02	1	1.2992e-04	1
29	1	2	3	0	15	-1.3874e-01	1	-6.9844e-04	1
30	7	4	6	5	10	3.2448e-02	-1	5.1812e-05	-1
31	5	3	6	14	1	5.1624e-02	1	3.6520e-04	1
32	4	2	6	0	15	-5.1789e-02	1	-1.8934e-04	1
33	6	3	2	15	0	-3.1130e-02	-1	3.1976e-04	1
34	5	2	6	13	2	5.1784e-02	1	1.8934e-04	1
35	1	2	4	13	2	9.5217e-02	1	1.8691e-05	1
36	3	7	1	0	15	-1.3368e-01	1	-1.4411e-04	1
37	7	3	1	15	0	4.6590e-02	1	4.4310e-04	1
38	4	3	2	15	0	1.9808e-01	1	6.9844e-04	1
39	7	3	5	15	0	1.2099e-01	1	4.3715e-04	1
40	2	7	1	11	4	-6.0070e-02	-1	7.8348e-05	1
41	1	6	5	3	12	-5.8880e-02	1	-7.1950e-05	1
42	2	1	5	0	15	4.3440e-02	-1	-2.2841e-04	1
43	6	7	2	0	15	-1.3042e-01	1	-1.1123e-04	1
44	5	2	4	11	4	-1.7580e-02	-1	-1.8690e-05	-1
45	2	5	4	15	0	5.1829e-02	1	2.4260e-04	1
46	3	2	1	0	15	-3.5290e-02	1	-3.2880e-05	1
47	1	4	5	0	15	-1.2825e-01	1	-2.7998e-04	1
48	7	1	4	15	0	5.1209e-02	1	1.4190e-05	1
49	2	7	4	12	3	3.5199e-02	1	9.2538e-05	1
50	1	6	3	0	15	-1.6459e-01	1	-5.0910e-04	1
51	2	7	6	1	14	-6.5300e-02	1	-7.8110e-05	1
52	6	4	2	2	13	-4.8710e-02	1	1.8690e-05	-1
53	5	1	2	12	3	7.6470e-02	1	3.2880e-05	1
54	6	5	3	0	15	-6.9324e-02	1	-4.3715e-04	1
55	1	5	3	0	15	-1.0571e-01	1	-4.3715e-04	1

	Target	Left Bkgnd.	Right Bkgnd.	Left Picks	Right Picks	ACE (L.-R.)	ACE R/W	VOE (L.-R.)	VOE R/W
56	2	3	7	15	0	1.6352e-01	1	5.8721e-04	1
57	1	7	3	0	15	-1.8027e-01	1	-5.8721e-04	1
58	5	7	3	0	15	-7.2333e-02	1	-2.8709e-04	1
59	2	5	7	14	1	1.6630e-02	1	1.5006e-04	1
60	3	7	5	11	4	4.8660e-02	1	1.5006e-04	1
61	4	2	1	11	4	-5.2000e-05	-1	4.5010e-06	1
62	2	6	1	15	0	5.2300e-03	1	1.5646e-04	1
63	7	6	5	2	13	-9.2540e-03	1	-7.1952e-05	1
64	1	3	4	15	0	2.3396e-01	1	7.1713e-04	1
65	7	3	2	15	0	6.5130e-02	1	4.7598e-04	1
66	5	4	3	2	13	1.7740e-02	-1	-1.5717e-04	1
67	3	2	7	15	0	9.8390e-02	1	1.1123e-04	1
68	3	7	6	15	0	3.0960e-02	1	7.8110e-05	1
69	1	6	7	8	7	1.5680e-02	1	7.8110e-05	1
70	3	1	4	9	6	3.5930e-02	1	1.4190e-05	1
71	6	2	5	15	0	1.0045e-01	1	1.1739e-04	1
72	4	1	7	0	15	-2.4780e-03	1	-1.1573e-04	1
73	5	3	2	13	2	-1.6000e-04	-1	1.7586e-04	1
74	4	6	5	10	5	-1.7620e-02	-1	-7.1950e-05	-1
75	5	2	7	14	1	7.2493e-02	1	1.1123e-04	1
76	3	7	4	0	15	-9.7750e-02	1	-1.2992e-04	1
77	4	1	3	0	15	-1.9803e-01	1	-7.0294e-04	1
78	7	1	5	6	9	7.4403e-02	-1	-5.9500e-06	1
79	6	3	7	15	0	9.9287e-02	1	4.3099e-04	1
80	1	2	7	4	11	4.1530e-02	-1	-1.1123e-04	1
81	3	5	1	0	15	-1.8234e-01	1	-2.9417e-04	1
82	3	2	6	15	0	1.2935e-01	1	1.8934e-04	1
83	7	6	1	0	15	-8.3657e-02	1	-6.6002e-05	1
84	5	7	1	0	15	-1.4896e-01	1	-1.4411e-04	1
85	4	2	5	0	15	-6.9409e-02	1	-2.6129e-04	1

	Target	Left Bkgnd.	Right Bkgnd.	Left Picks	Right Picks	ACE (L.-R.)	ACE R/W	VOE (L.-R.)	VOE R/W
86	7	6	2	11	4	-6.5117e-02	-1	-3.3122e-05	-1
87	7	5	4	15	0	-2.3194e-02	-1	2.0140e-05	1
88	7	1	2	15	0	1.8540e-02	1	3.2880e-05	1
89	6	2	1	1	14	3.1080e-02	-1	-3.2880e-05	1
90	1	6	2	14	1	-2.5850e-02	-1	1.8934e-04	1
91	1	6	4	14	1	6.9367e-02	1	2.0803e-04	1
92	6	3	4	15	0	1.7580e-02	1	3.0107e-04	1
93	4	3	7	15	0	1.9555e-01	1	5.8721e-04	1
94	1	7	5	4	11	-7.4560e-02	1	-1.5006e-04	1
95	6	5	7	13	2	2.9963e-02	1	-6.1580e-06	-1
96	3	5	4	0	15	-1.4641e-01	1	-2.7998e-04	1
97	5	7	6	3	12	-2.0709e-02	1	7.8110e-05	-1
98	3	6	4	0	15	-1.2871e-01	1	-2.0803e-04	1
99	2	3	1	15	0	1.0345e-01	1	6.6556e-04	1
100	4	6	7	15	0	4.9259e-02	1	7.8110e-05	1
101	5	1	4	12	3	5.8890e-02	1	1.4190e-05	1
102	3	5	6	12	3	-1.7700e-02	-1	-7.1950e-05	-1
103	6	7	1	0	15	-9.9337e-02	1	-1.4411e-04	1
104	2	5	6	8	7	-4.8670e-02	-1	7.1950e-05	1
105	2	4	1	9	6	-9.5269e-02	-1	-1.4190e-05	-1

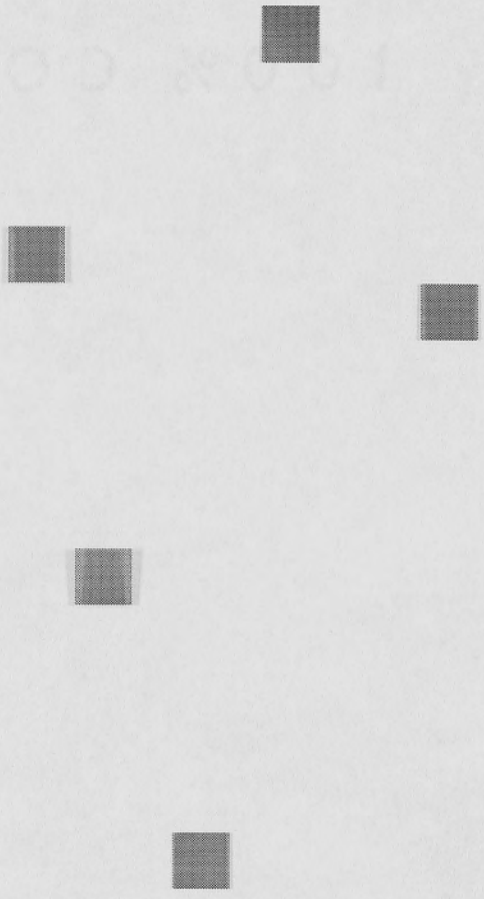


Figure E.1: Ground truth for image stimulus A.

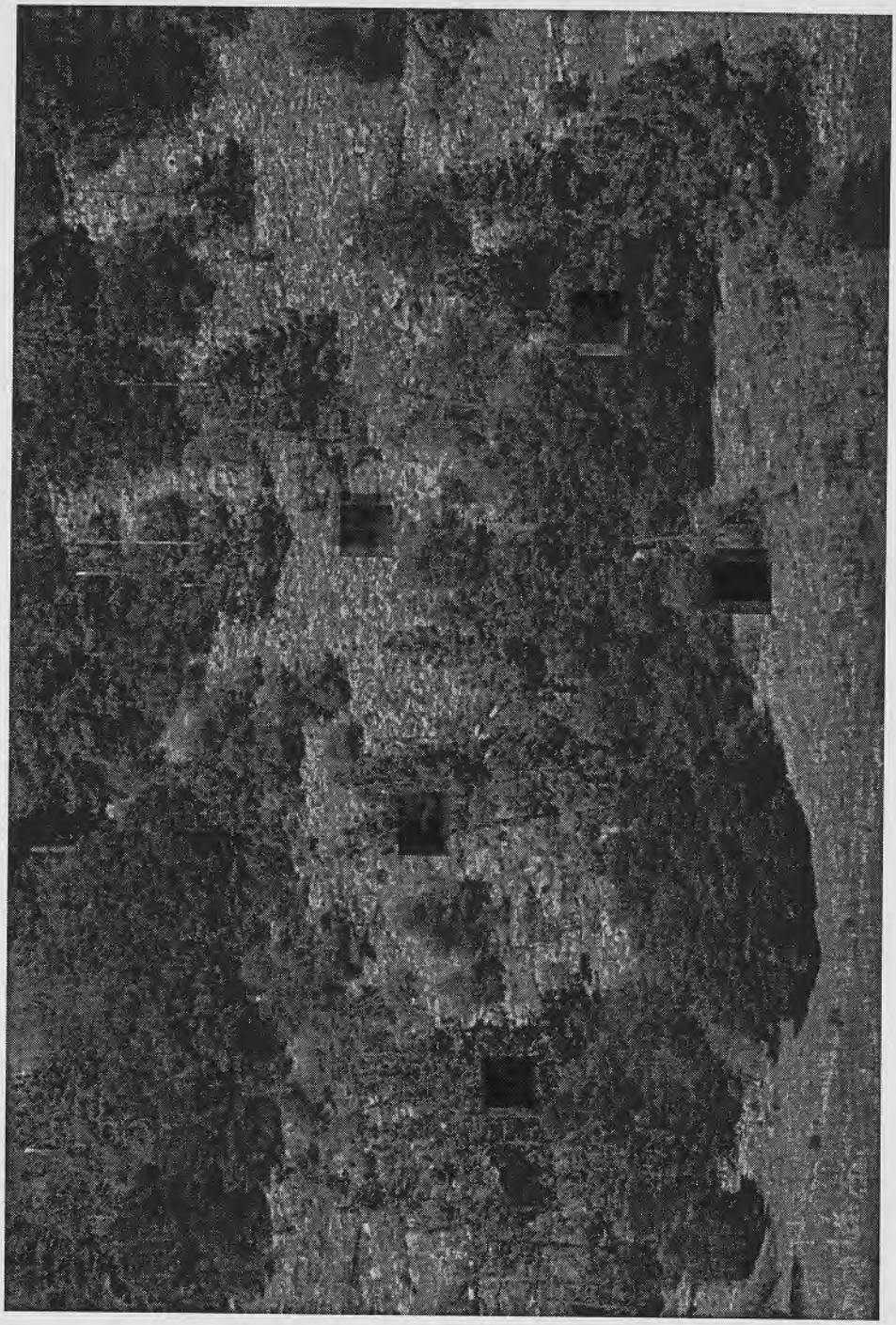


Figure E.1: Ground truth for image stimulus A.

Figure E.3: Image stimulus A.

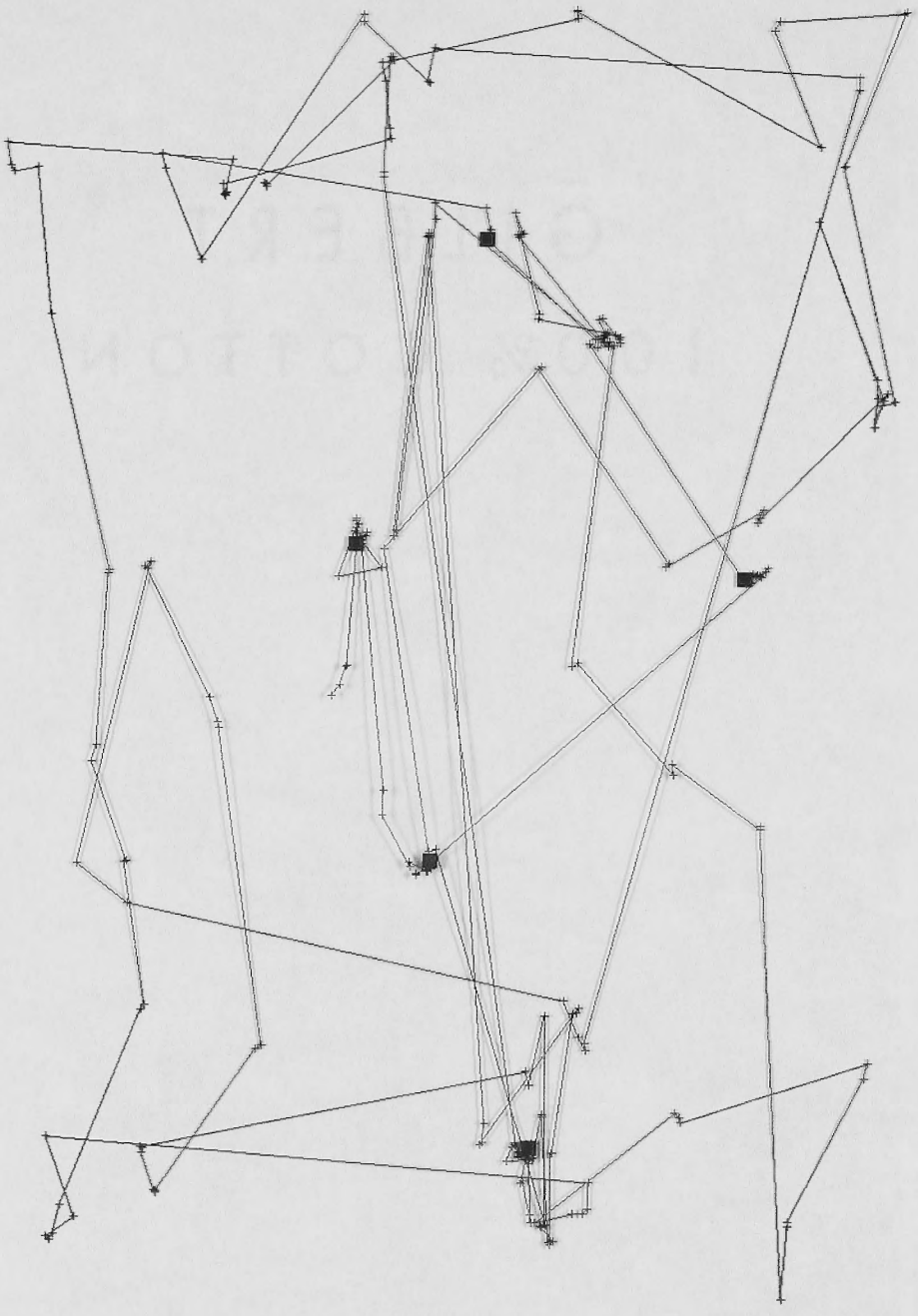


Figure E.2: Observer a's fixations for image stimulus A.

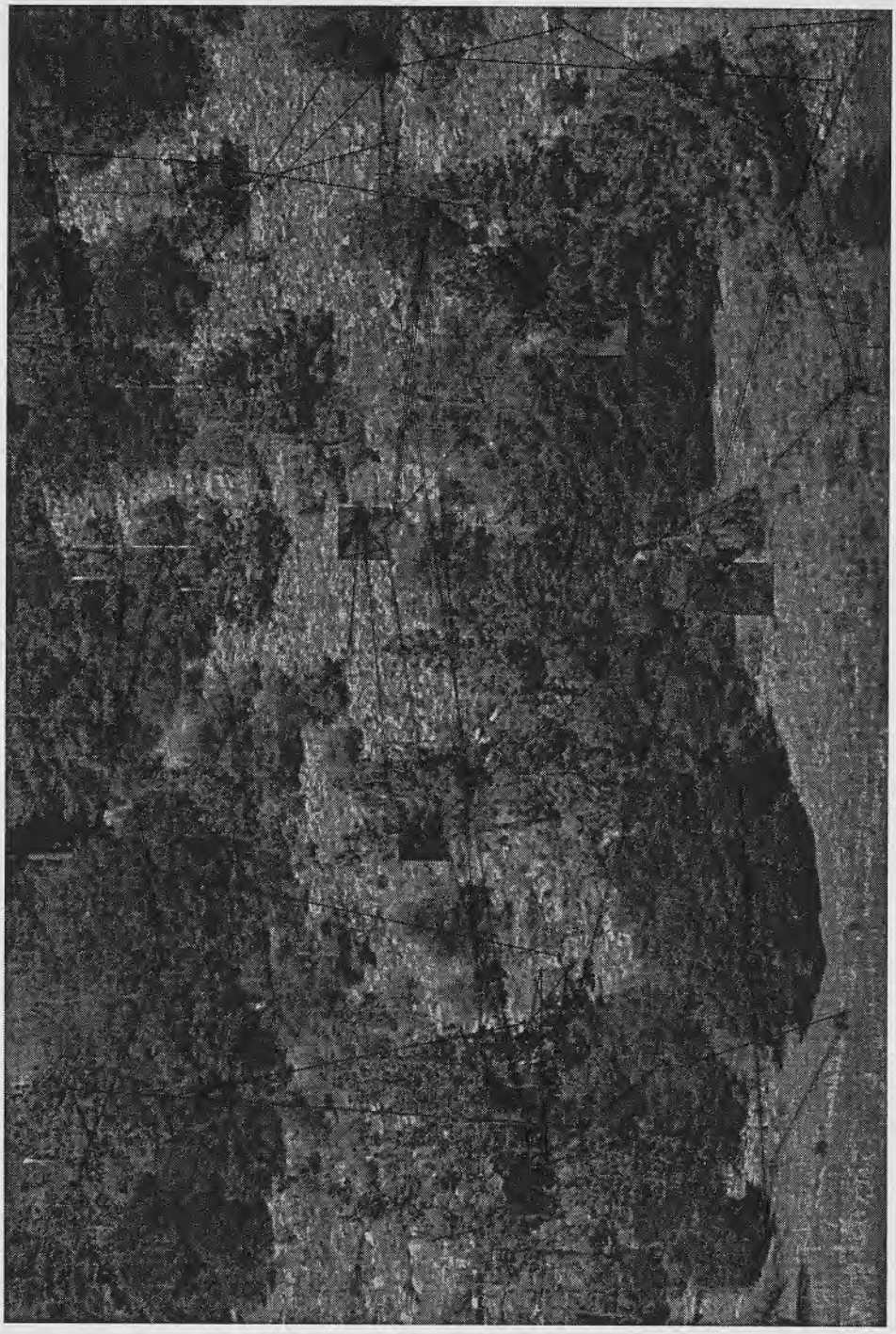


Figure E.2: Observer a's fixations for image stimulus A.

Figure E.3: Image stimulus A.

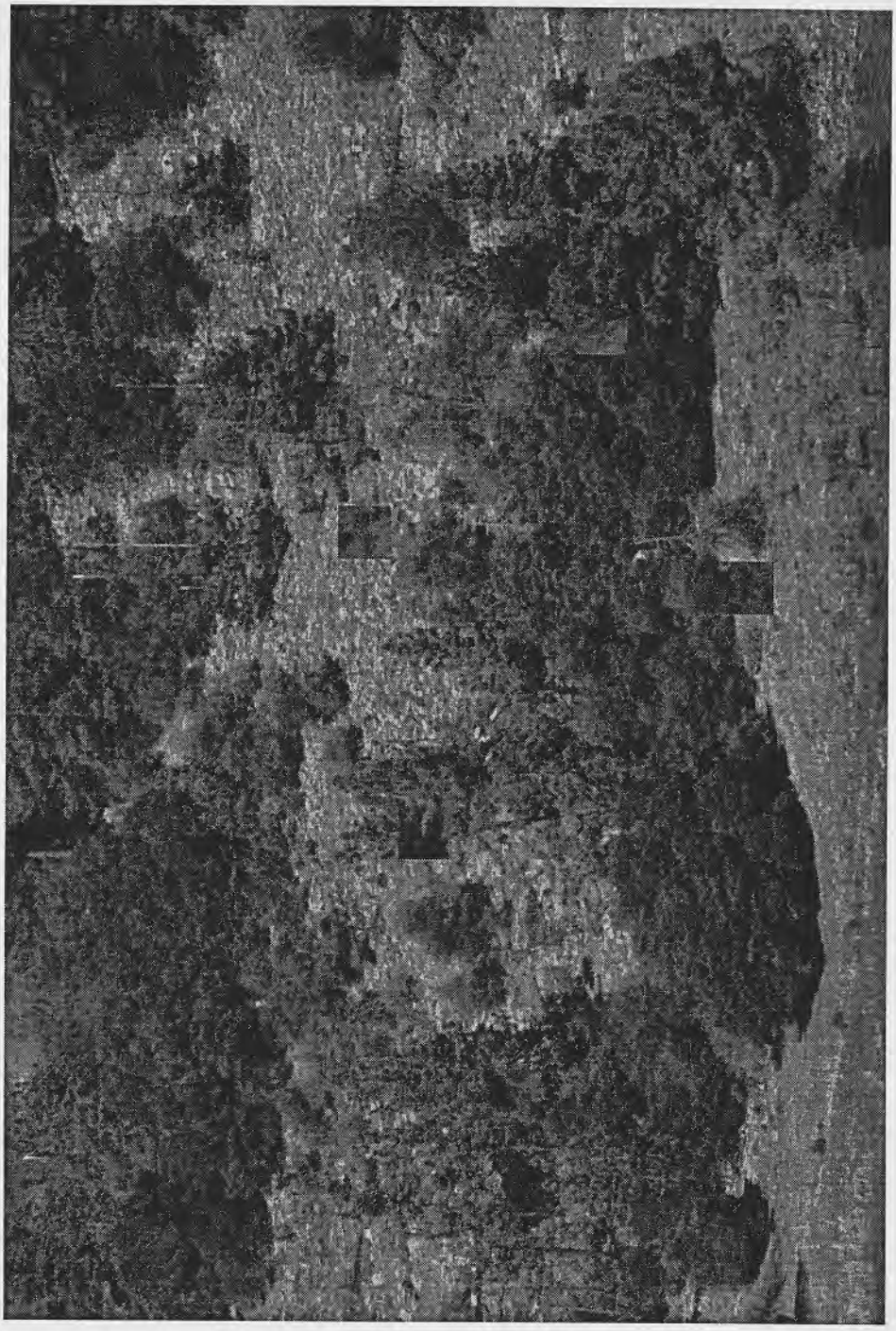


Figure E.3: Image stimulus A.

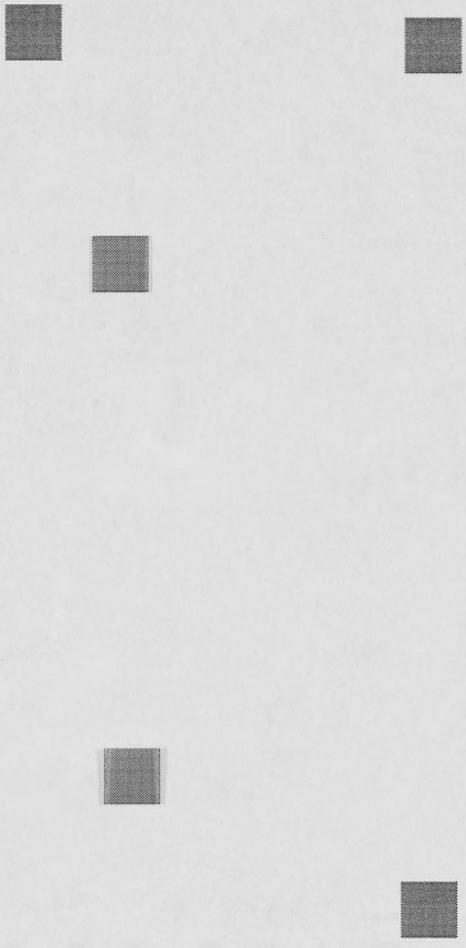


Figure E.4: Ground truth for image stimulus B.

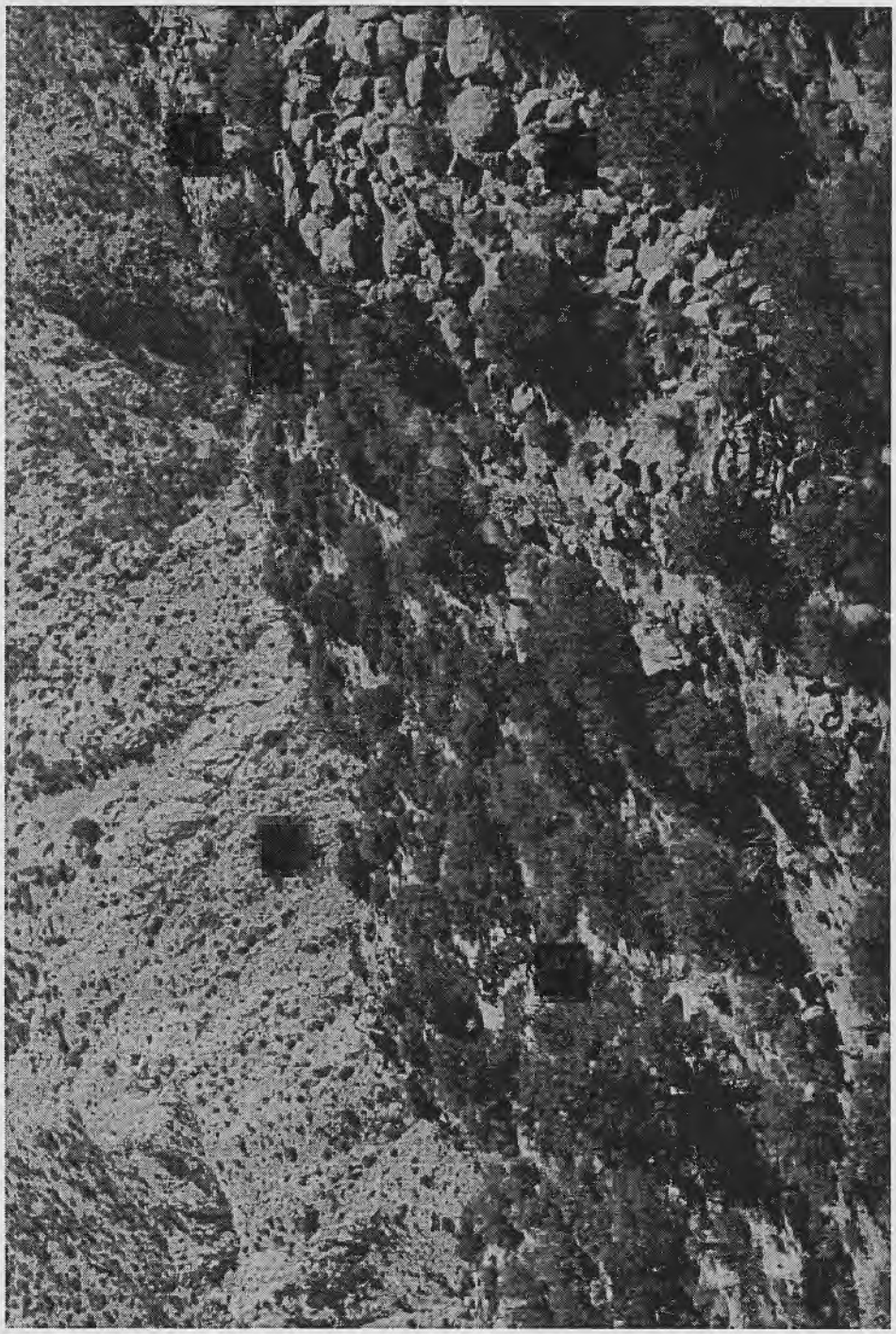


Figure E.4: Ground truth for image stimulus B.
Figure E.6: Image stimulus B.

Figure E.5: Observer a's fixations for image stimulus B.

Figure E.6: Image stimulus B.



Figure E.6: Image stimulus B.



Figure E.7: Ground truth for image stimulus C.

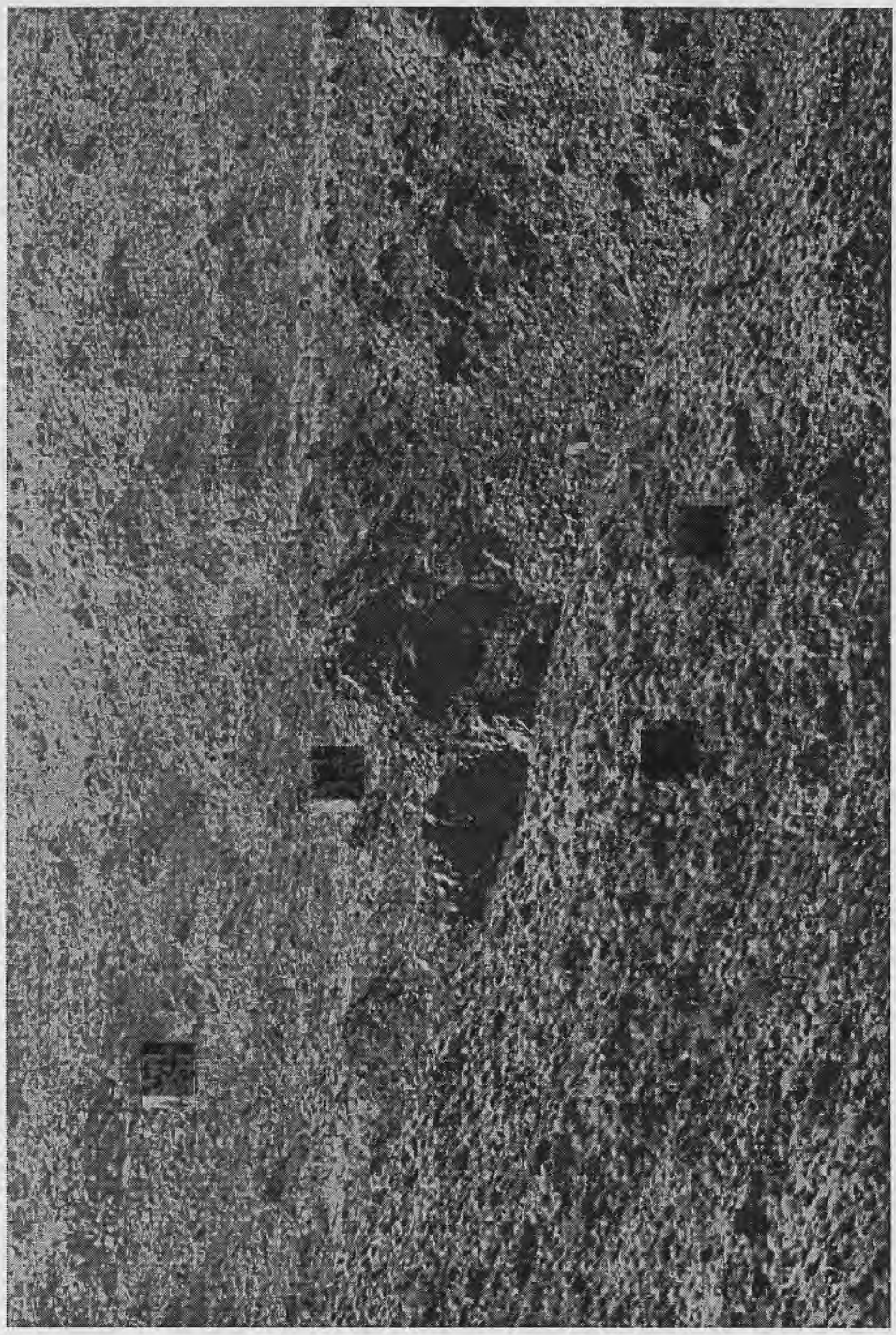


Figure E.7: Ground truth for image stimulus C.

Figure E.9: Image stimulus C.

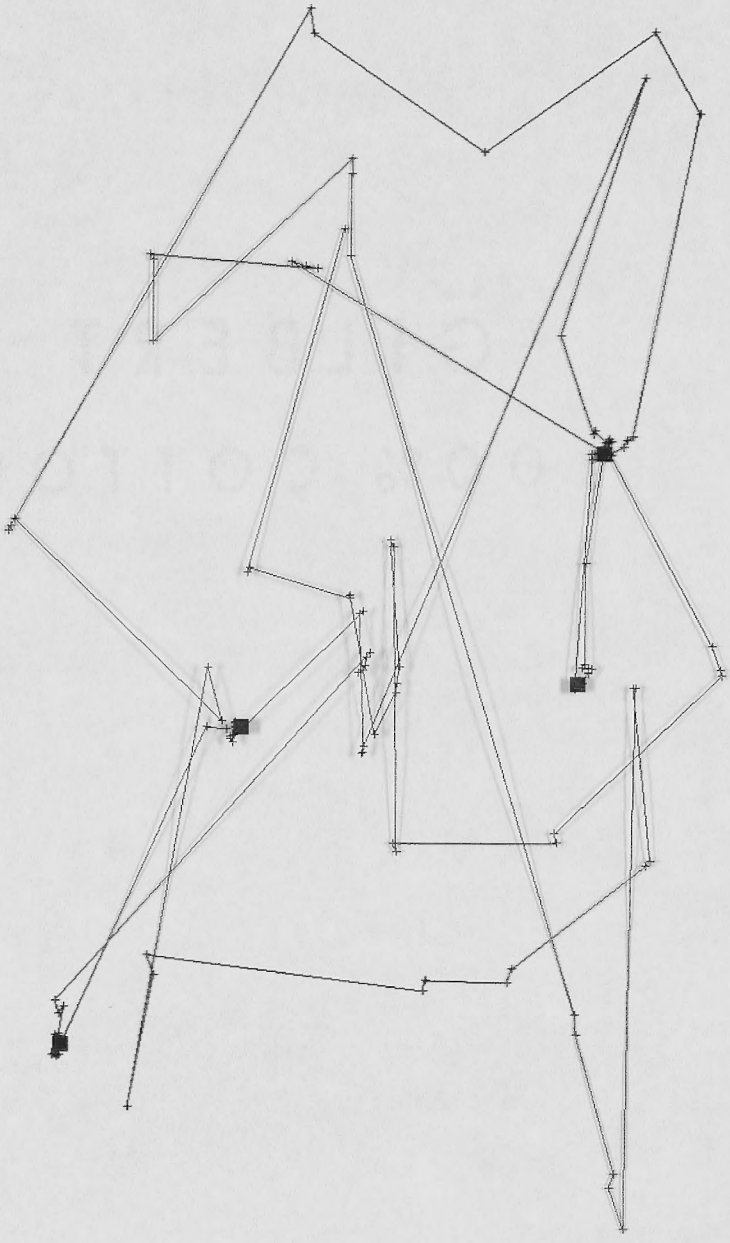


Figure E.8: Observer a's fixations for image stimulus C.



Figure E.8: Observer a's fixations for image stimulus C.

Figure E.9: Image stimulus C.

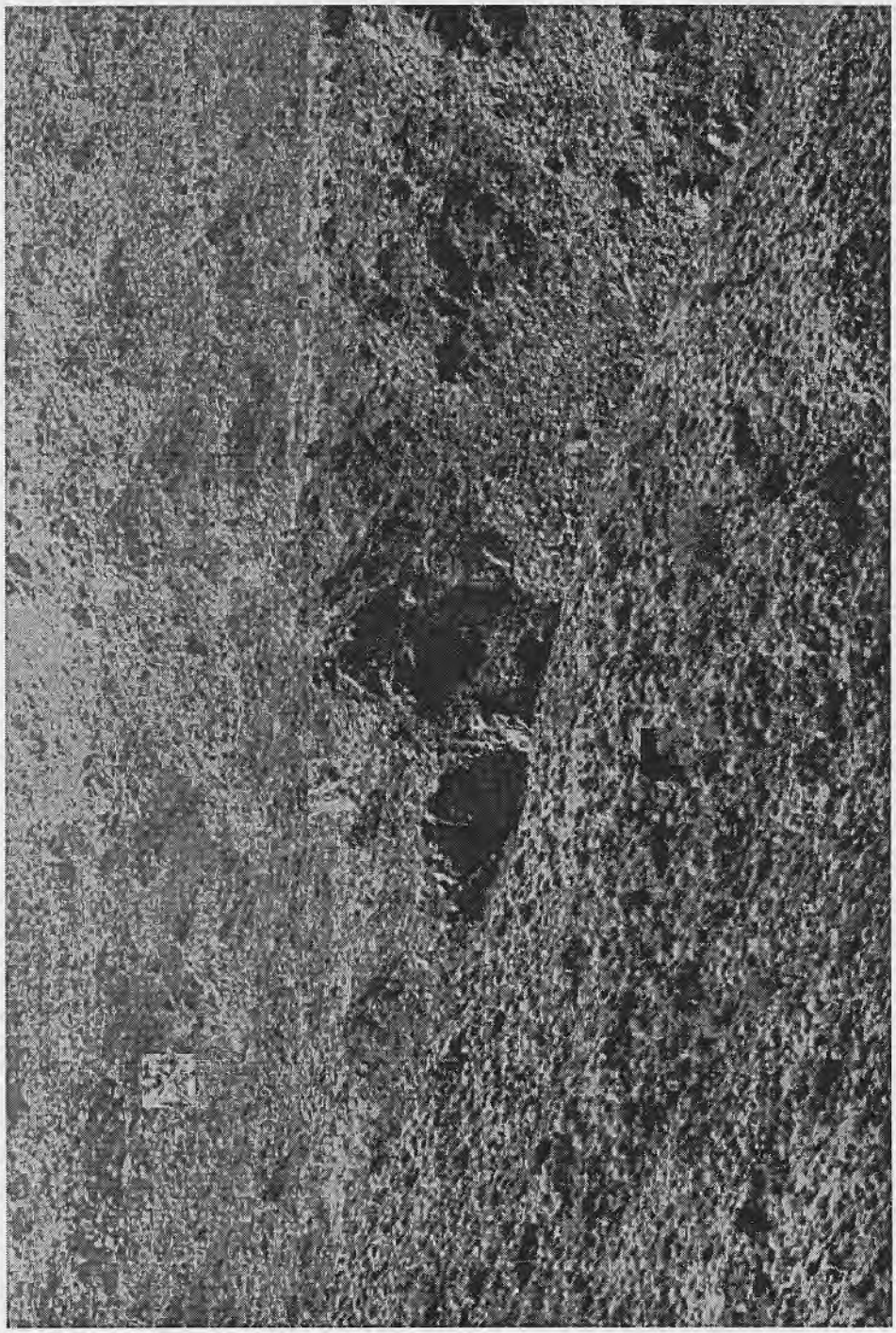


Figure E.9: Image stimulus C.

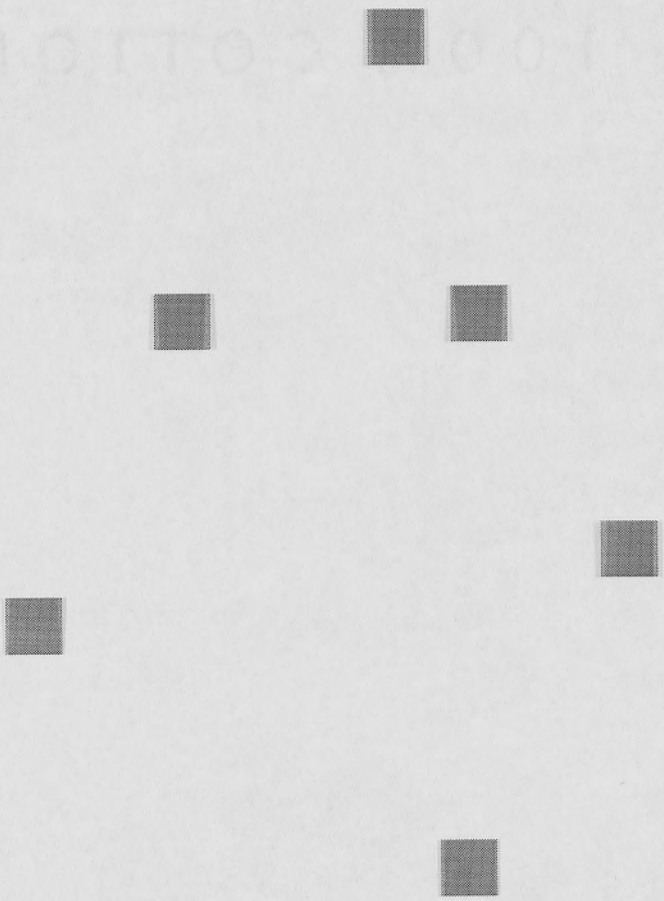


Figure E.10: Ground truth for image stimulus D.

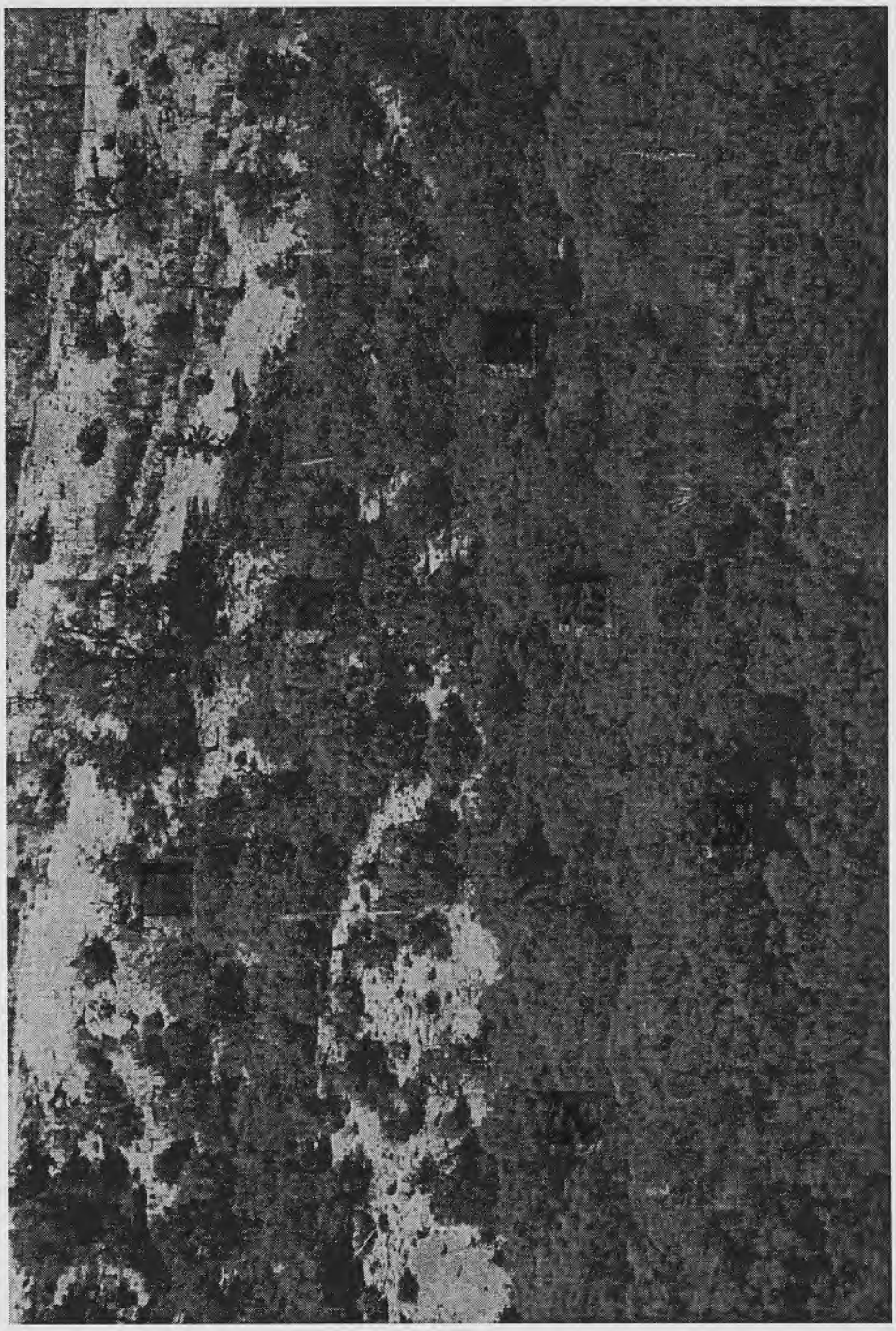


Figure E.10: Ground truth for image stimulus D.
Figure E.12: Image stimulus D.

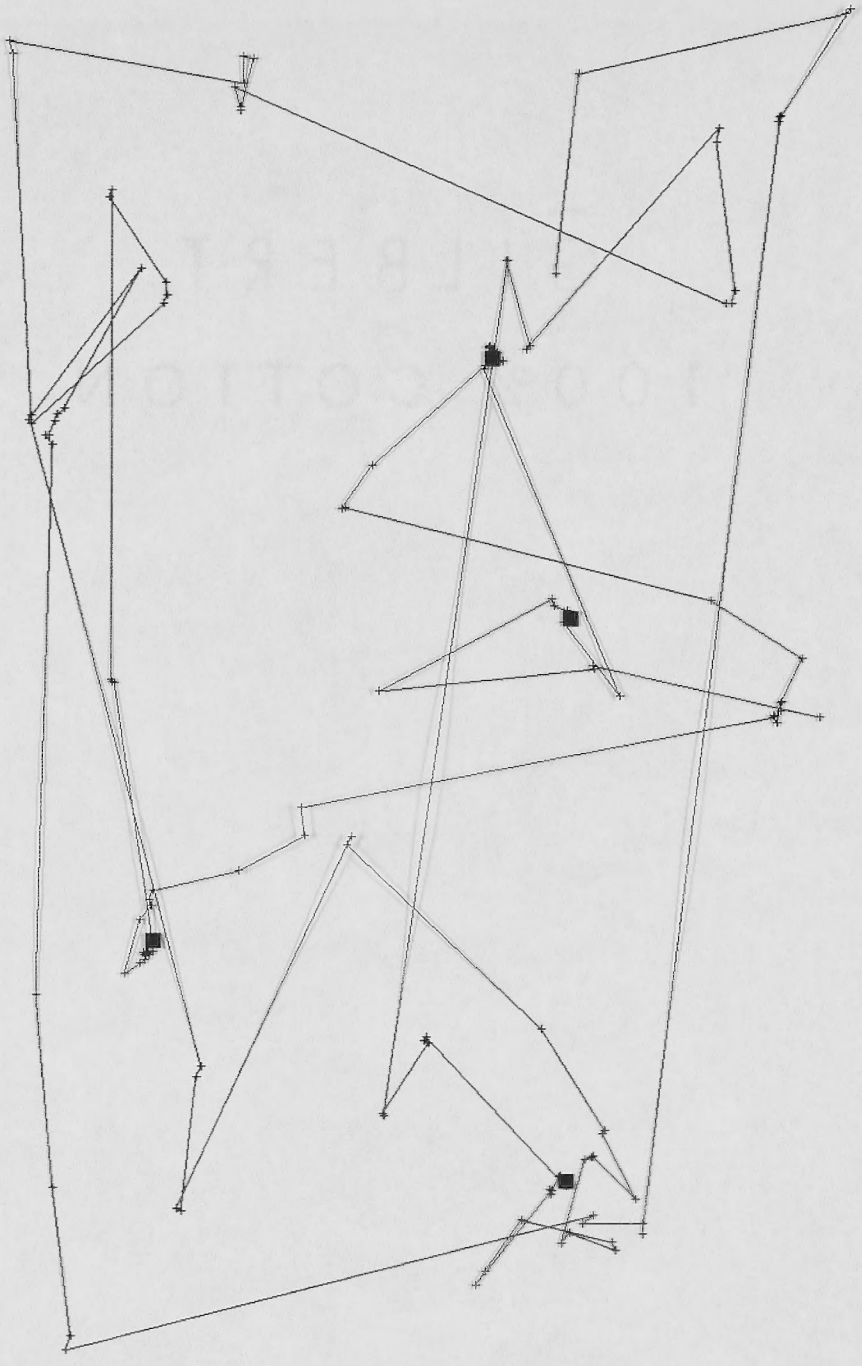


Figure E.11: Observer a's fixations for image stimulus D.

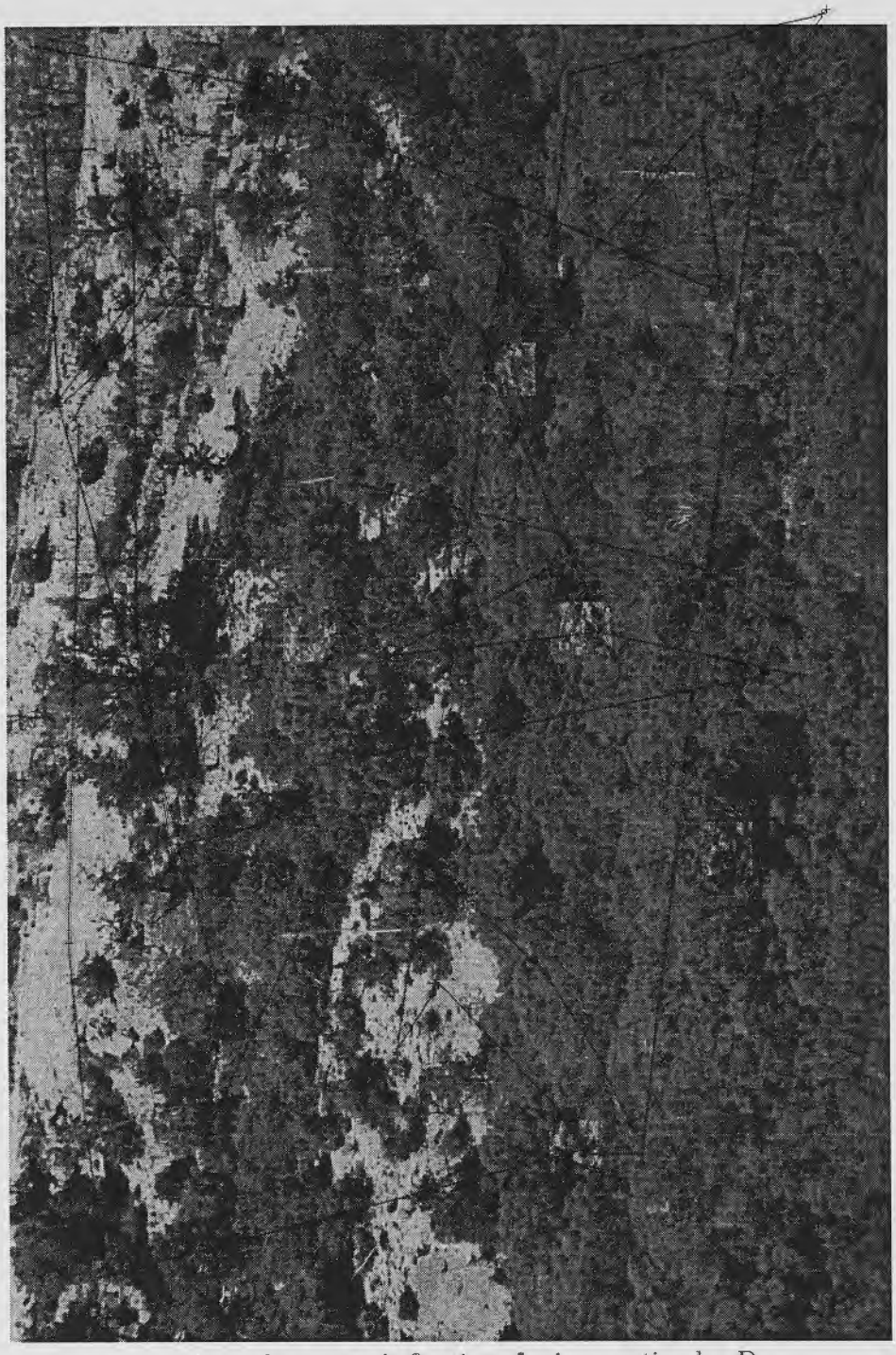


Figure E.11: Observer a's fixations for image stimulus D.
Figure E.12: Image stimulus D.

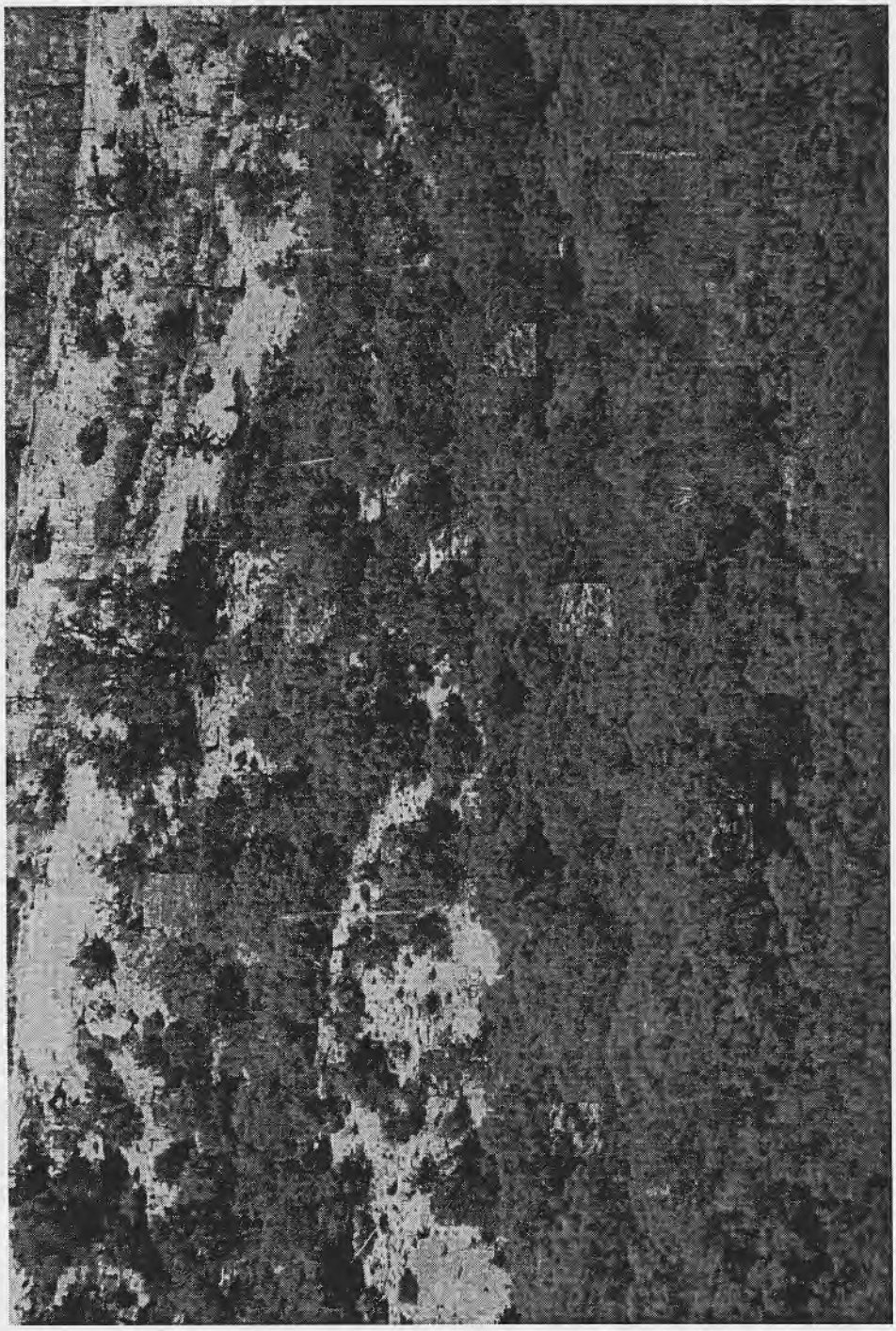


Figure E.12: Image stimulus D.

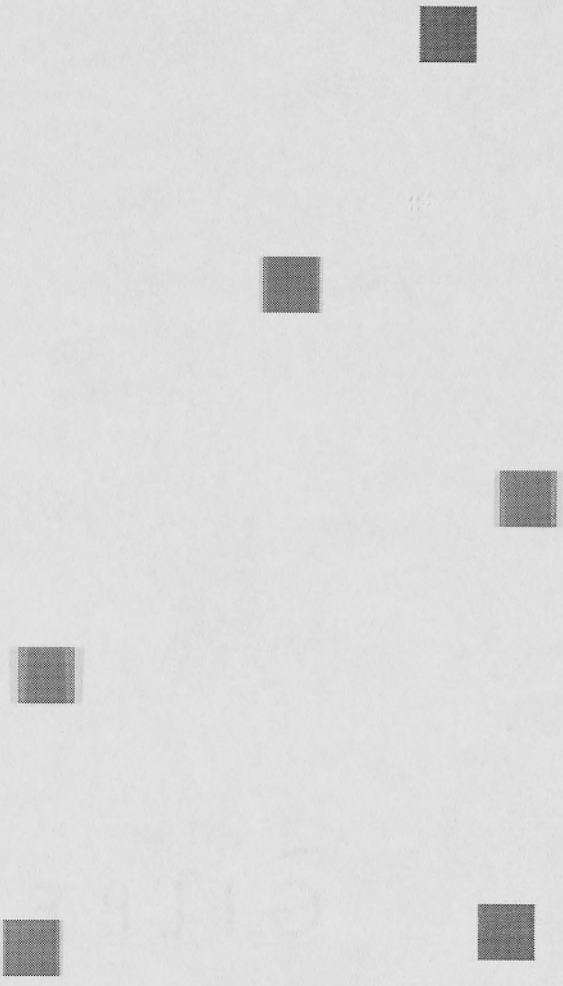


Figure E.13: Ground truth for image stimulus E.

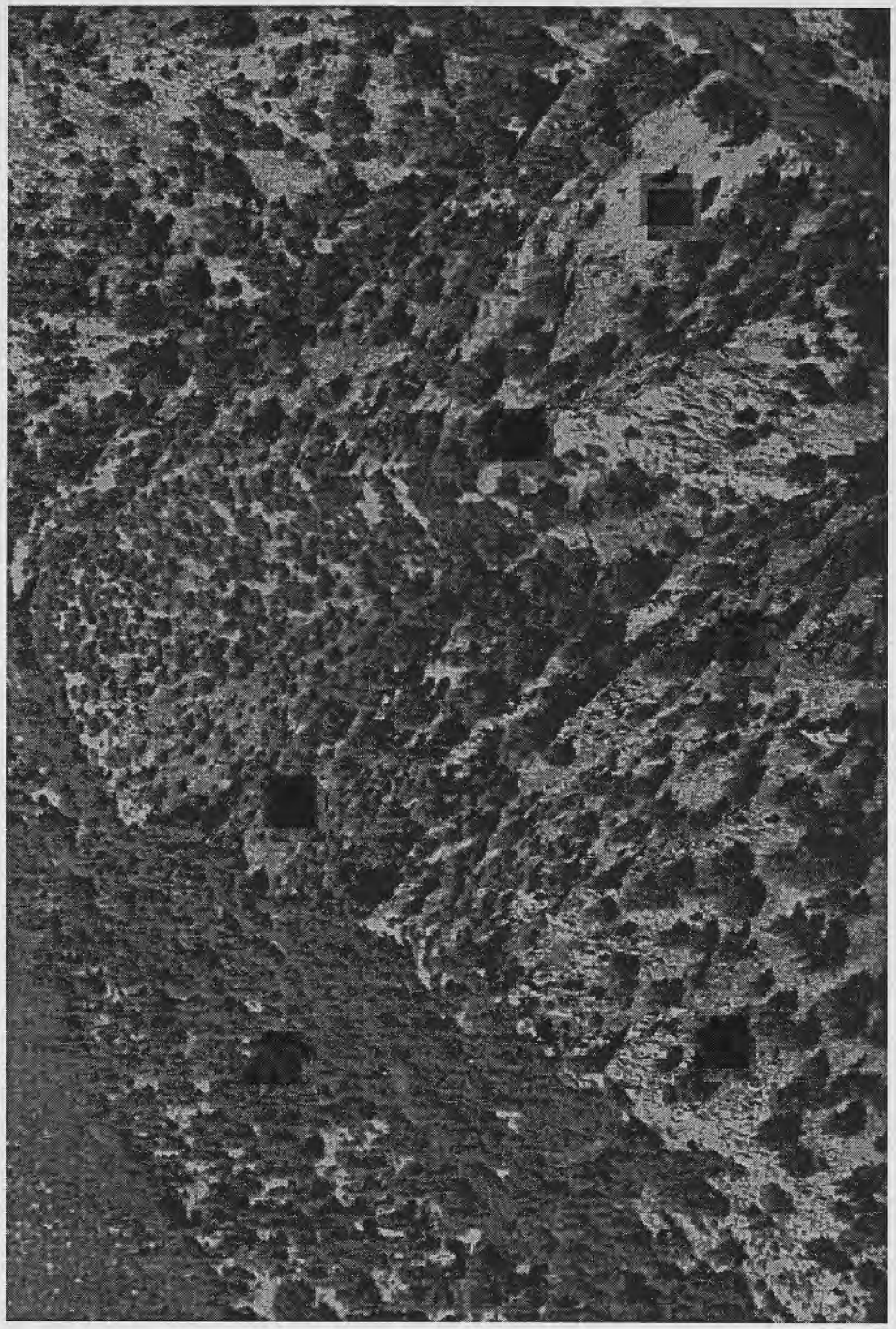


Figure E.13: Ground truth for image stimulus E.

Figure E.15: Image stimulus E.

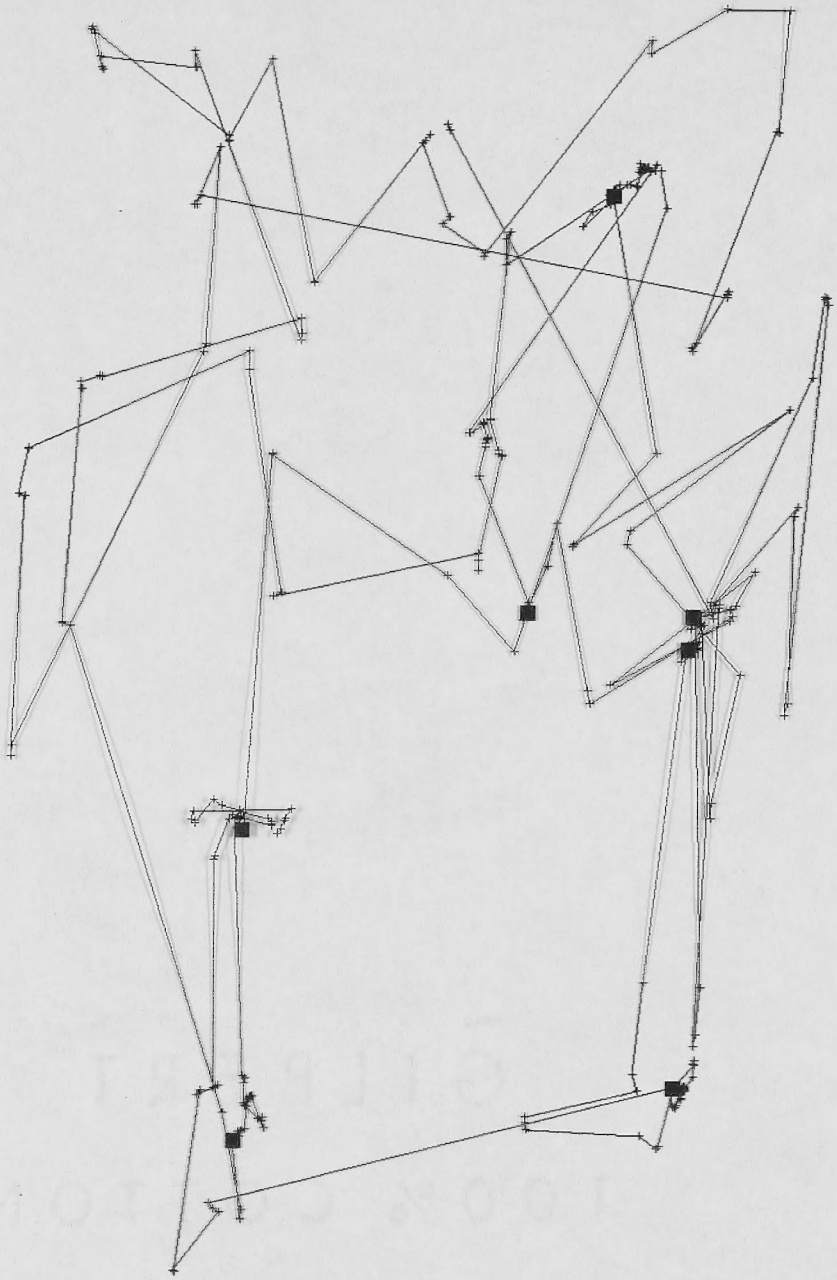


Figure E.14: Observer a's fixations for image stimulus E.

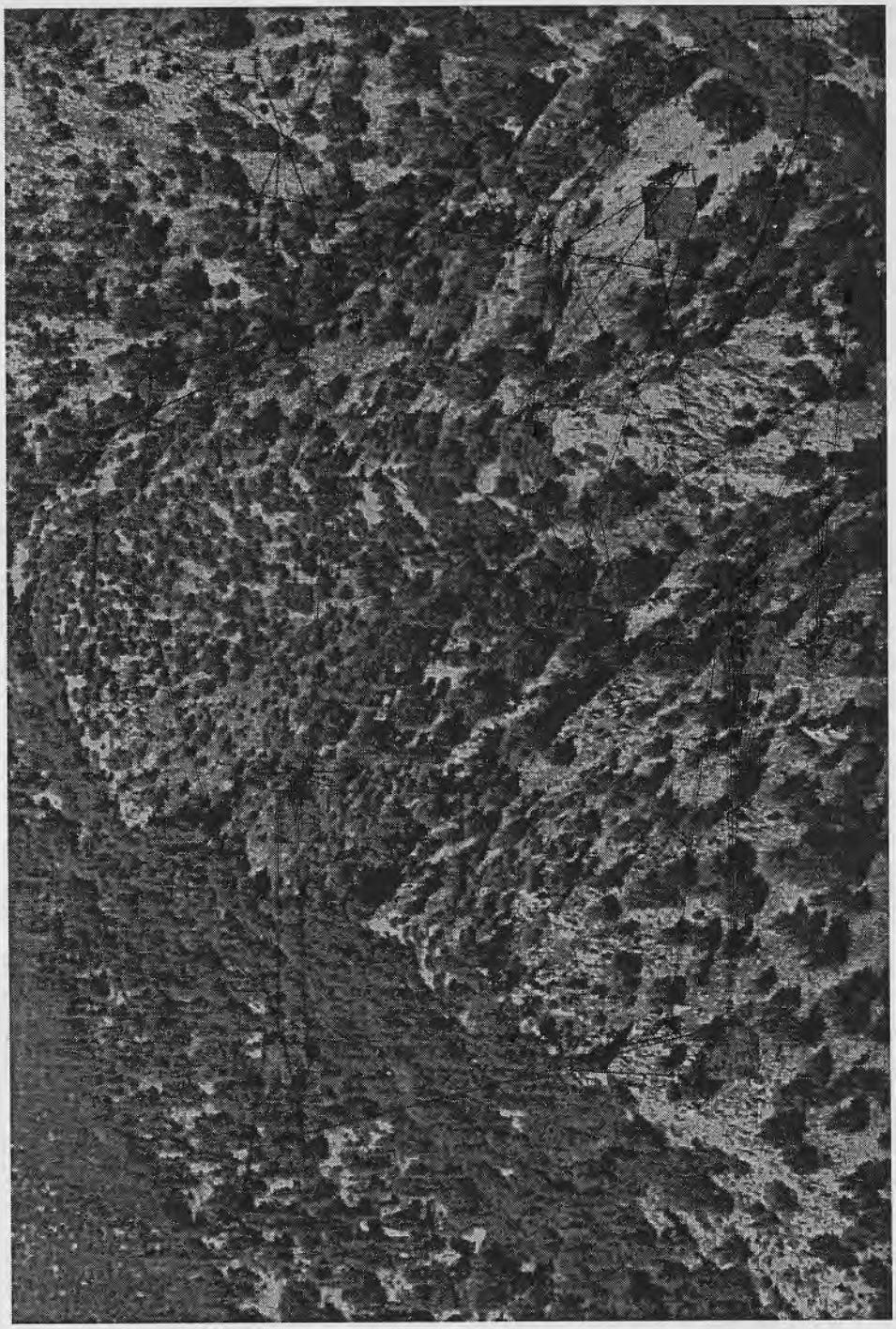


Figure E.14: Observer a's fixations for image stimulus E.

Figure E.15: Image stimulus E.

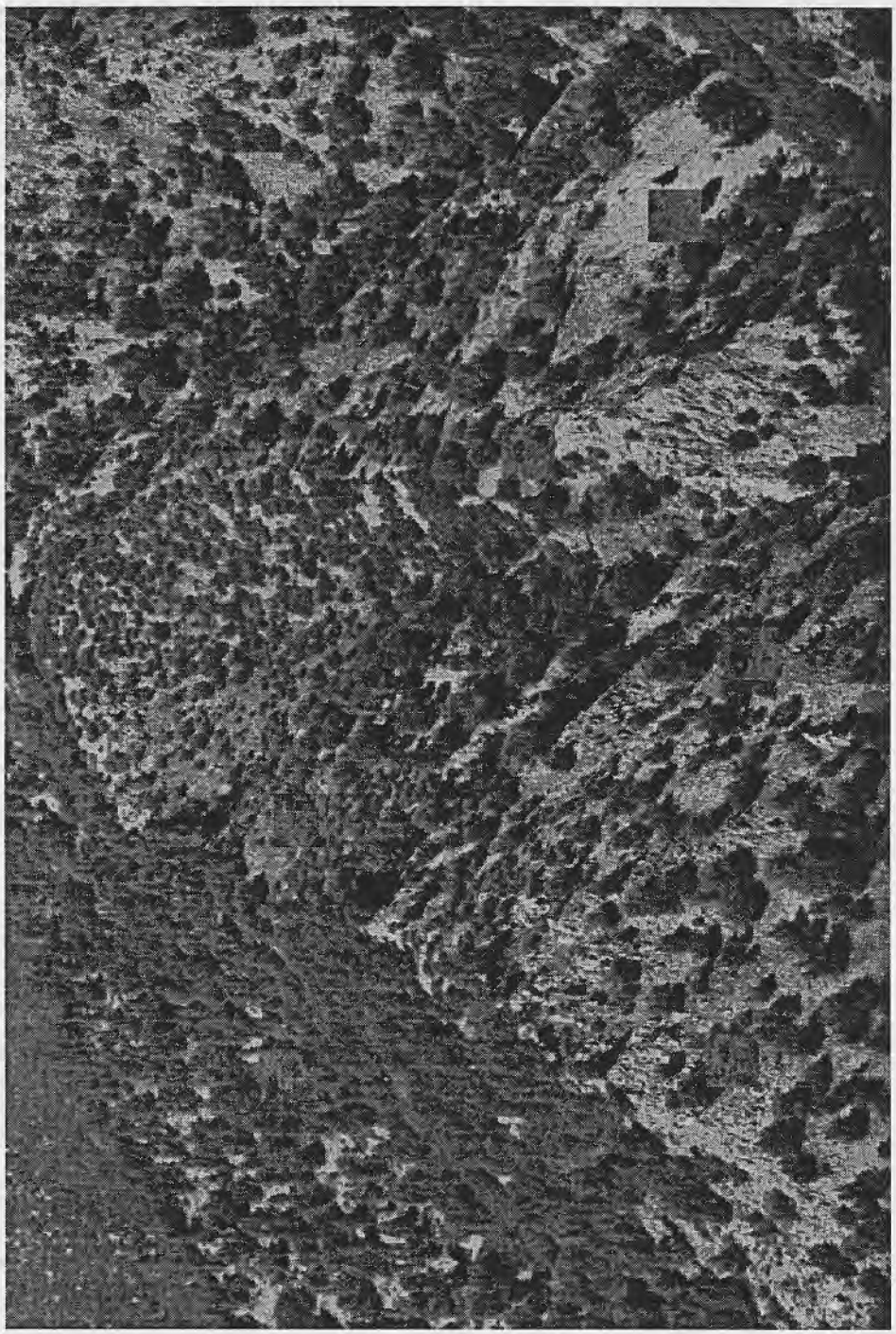


Figure E.15: Image stimulus E.

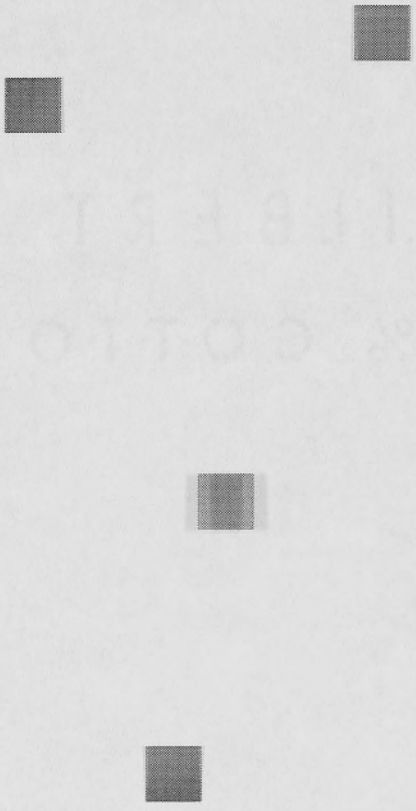


Figure E.16: Ground truth for image stimulus F.

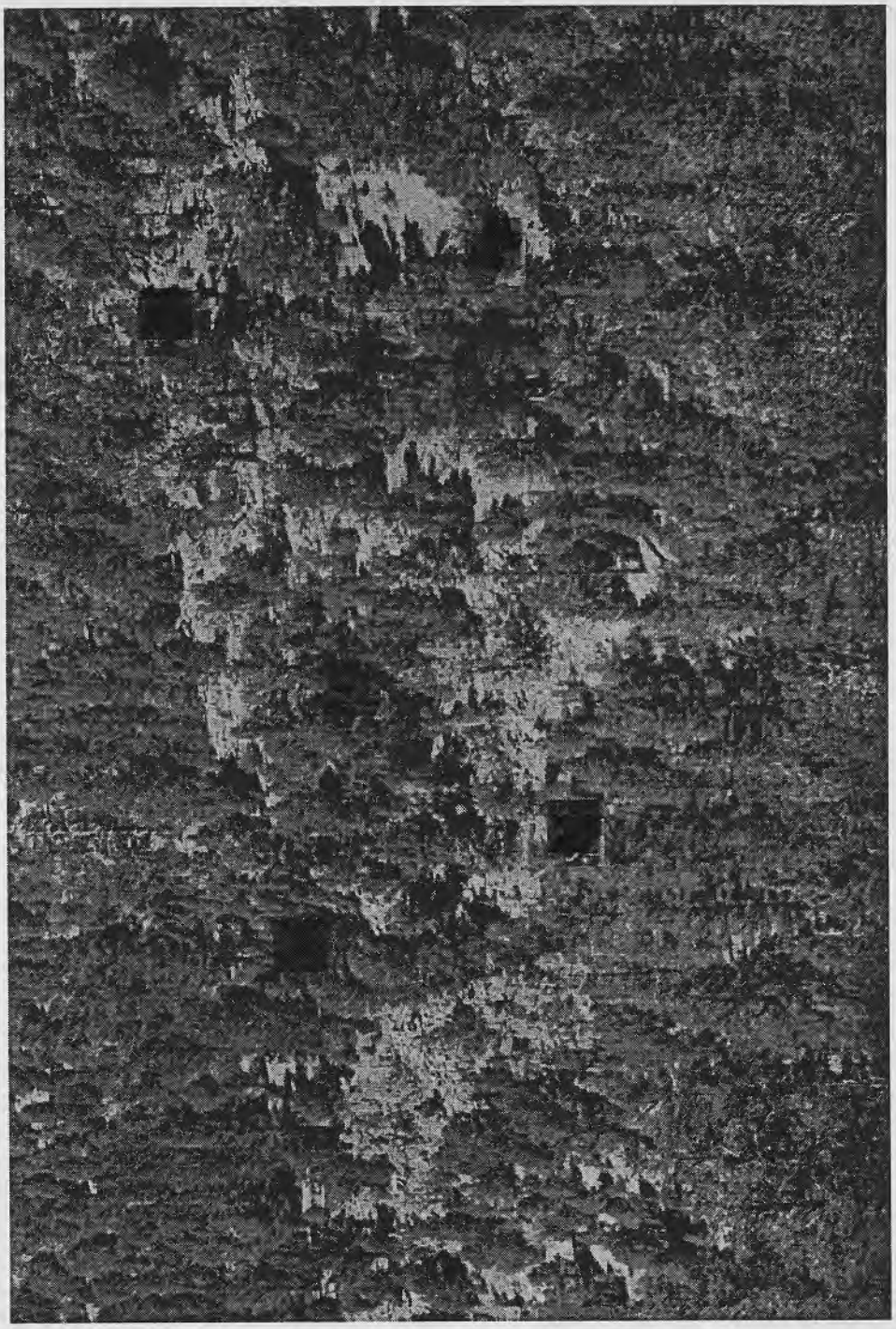


Figure E.16: Ground truth for image stimulus F.
Figure E.18: Image stimulus F.

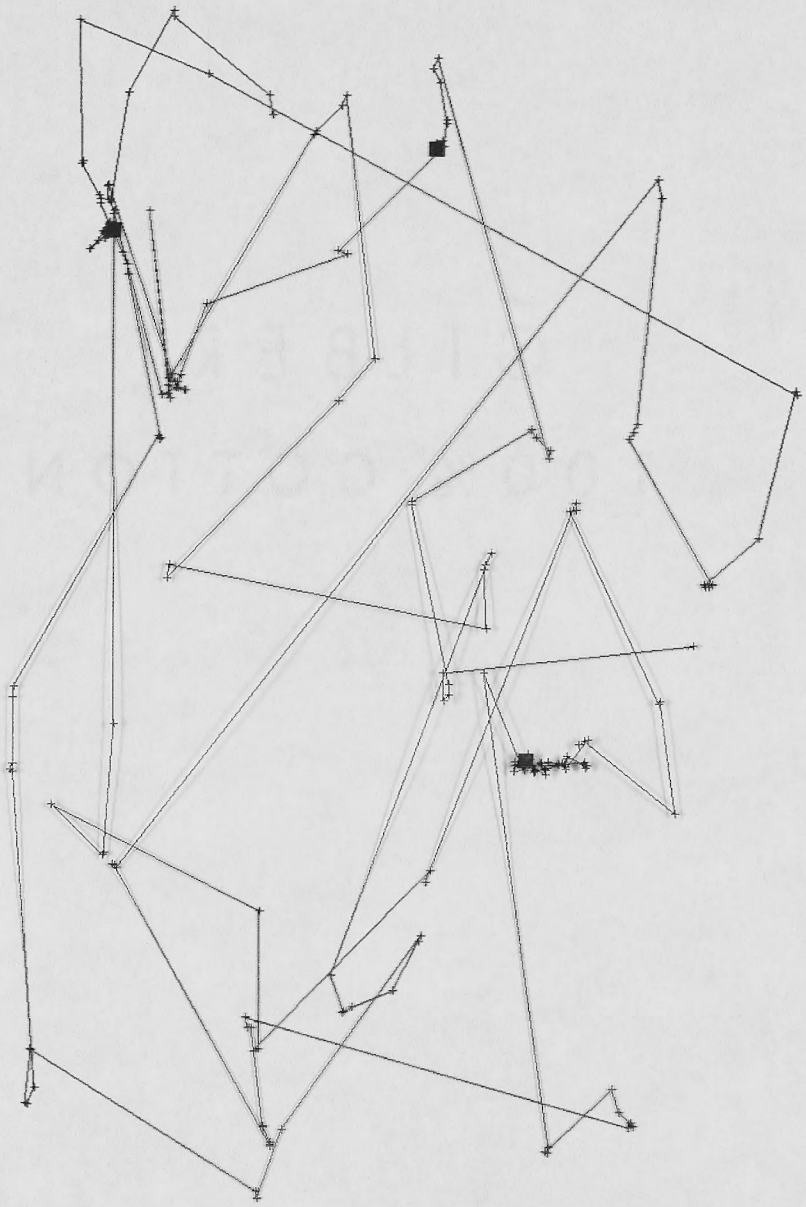


Figure E.17: Observer a's fixations for image stimulus F.



Figure E.17: Observer a's fixations for image stimulus F.

Figure E.18: Image stimulus F.

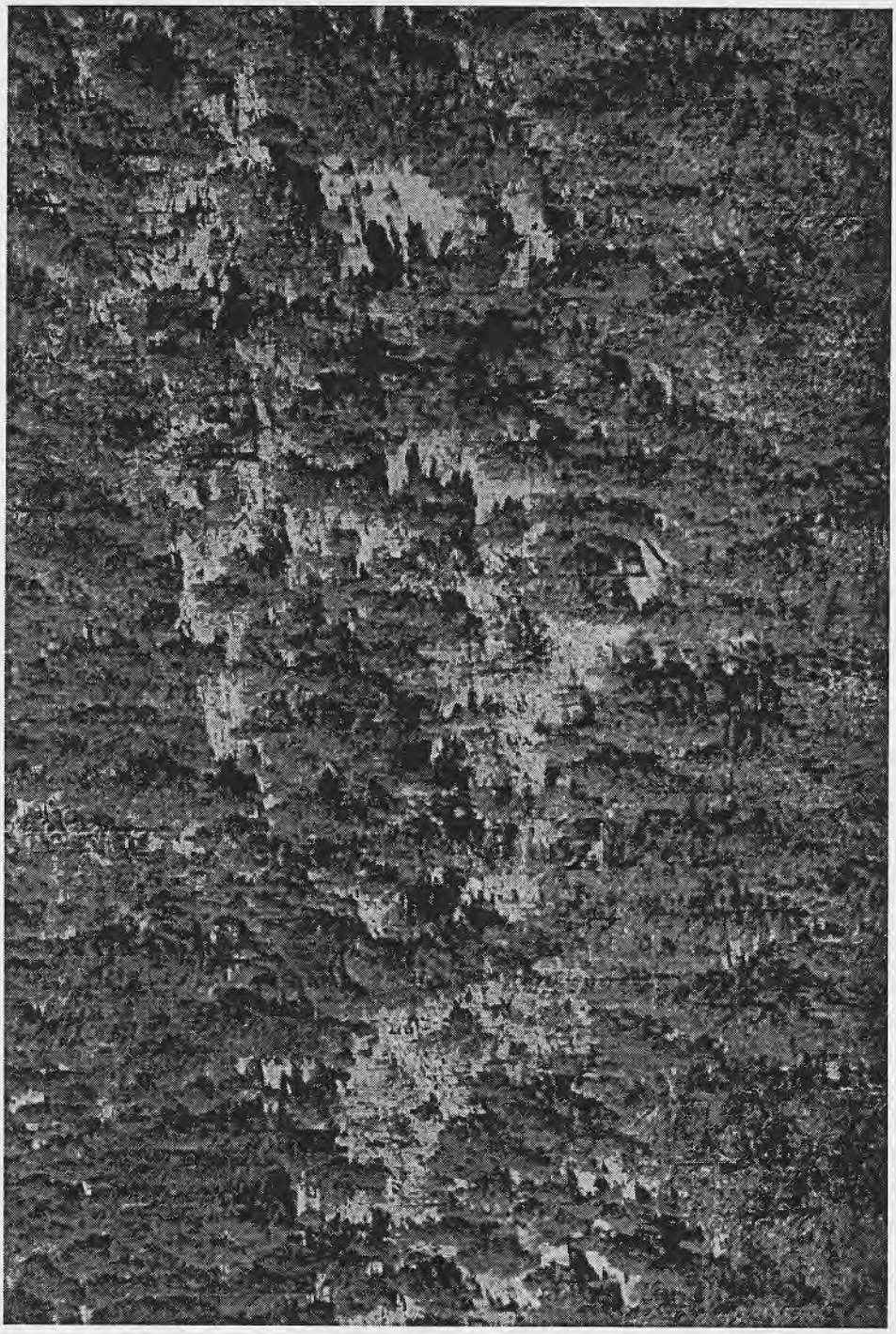


Figure E.18: Image stimulus F.

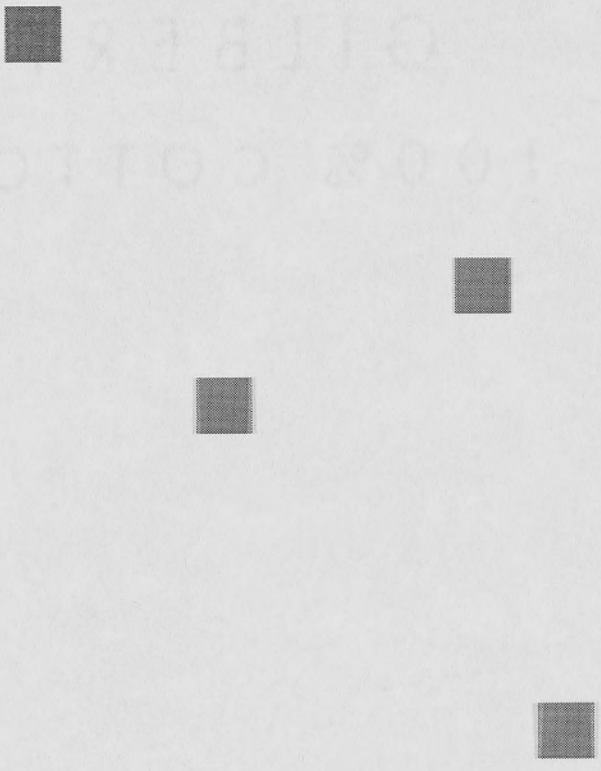


Figure E.19: Ground truth for image stimulus G.



Figure E.19: Ground truth for image stimulus G.
Figure E.21: Image stimulus G.

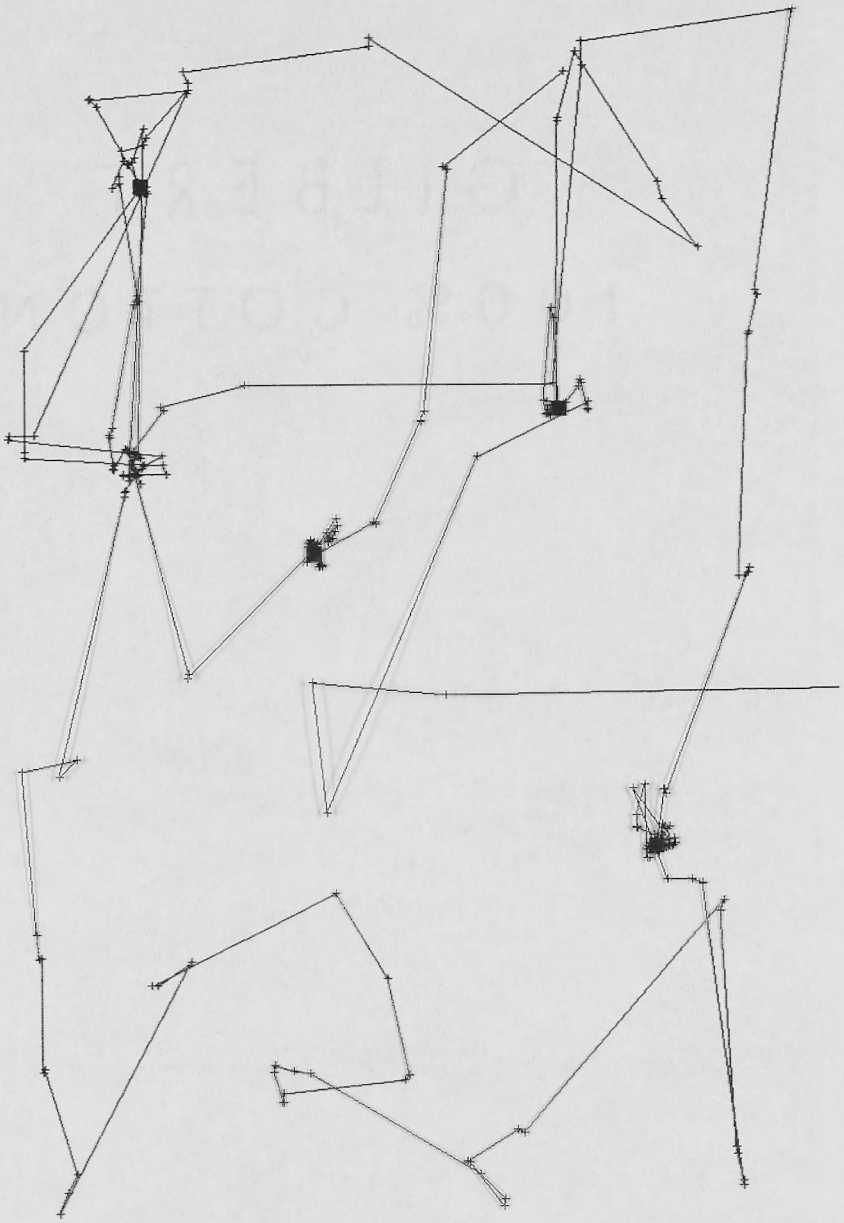


Figure E.20: Observer a's fixations for image stimulus G.



Figure E.20: Observer a's fixations for image stimulus G.

Figure E.21: Image stimulus G.



Figure E.21: Image stimulus G.

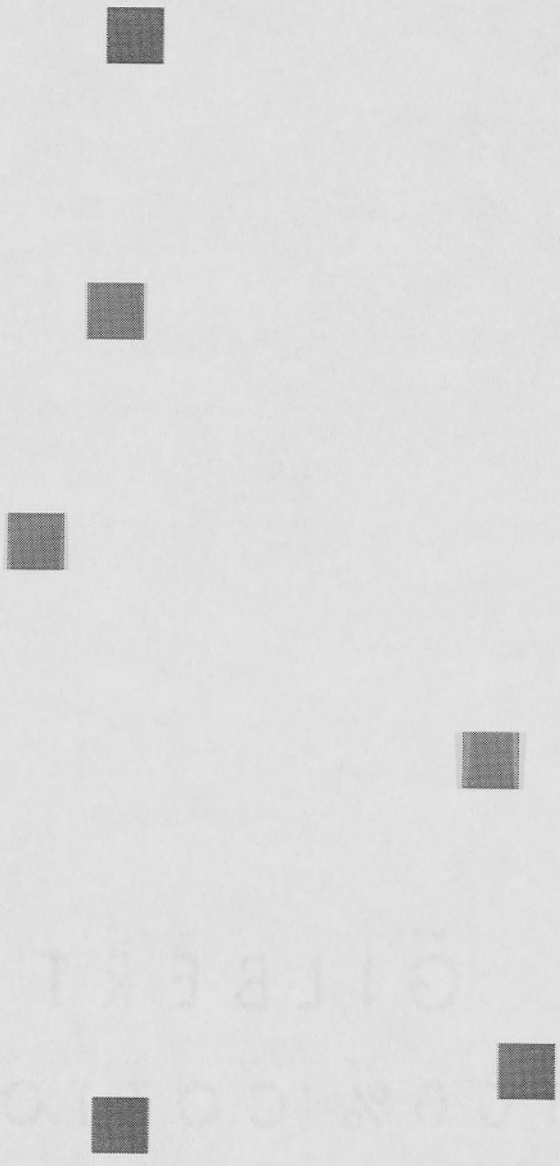


Figure E.22: Ground truth for image stimulus H.

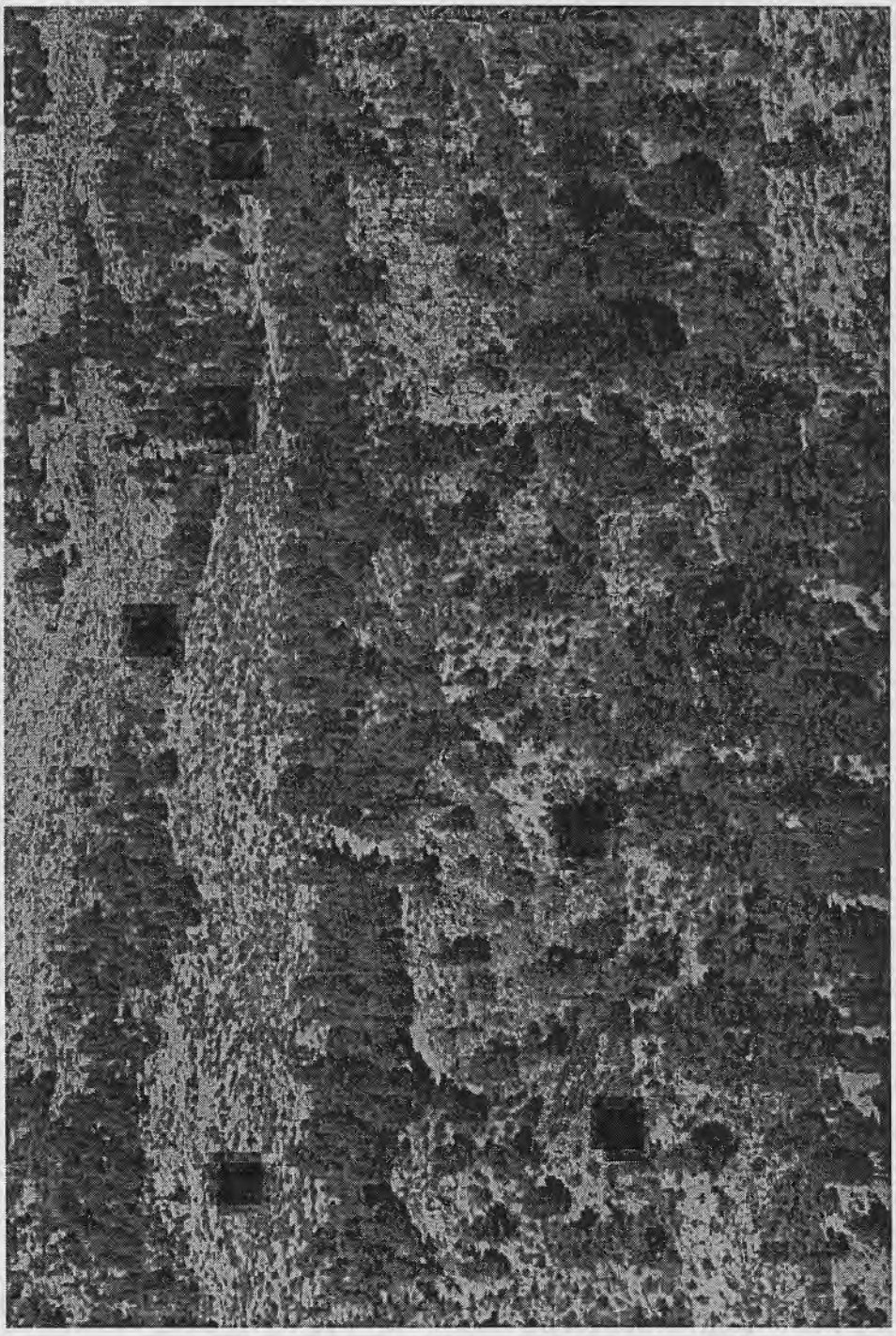


Figure E.22: Ground truth for image stimulus H.
Figure E.24: Image stimulus H.

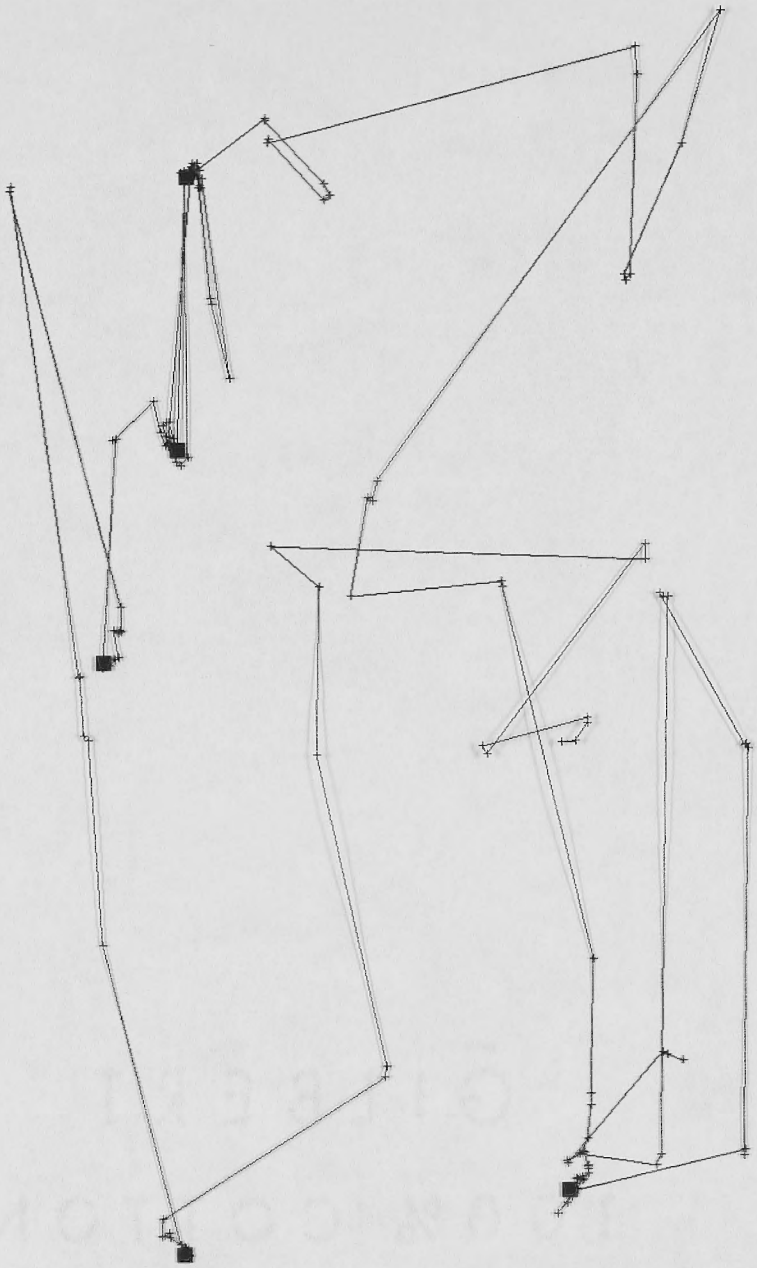


Figure E.23: Observer a's fixations for image stimulus H.

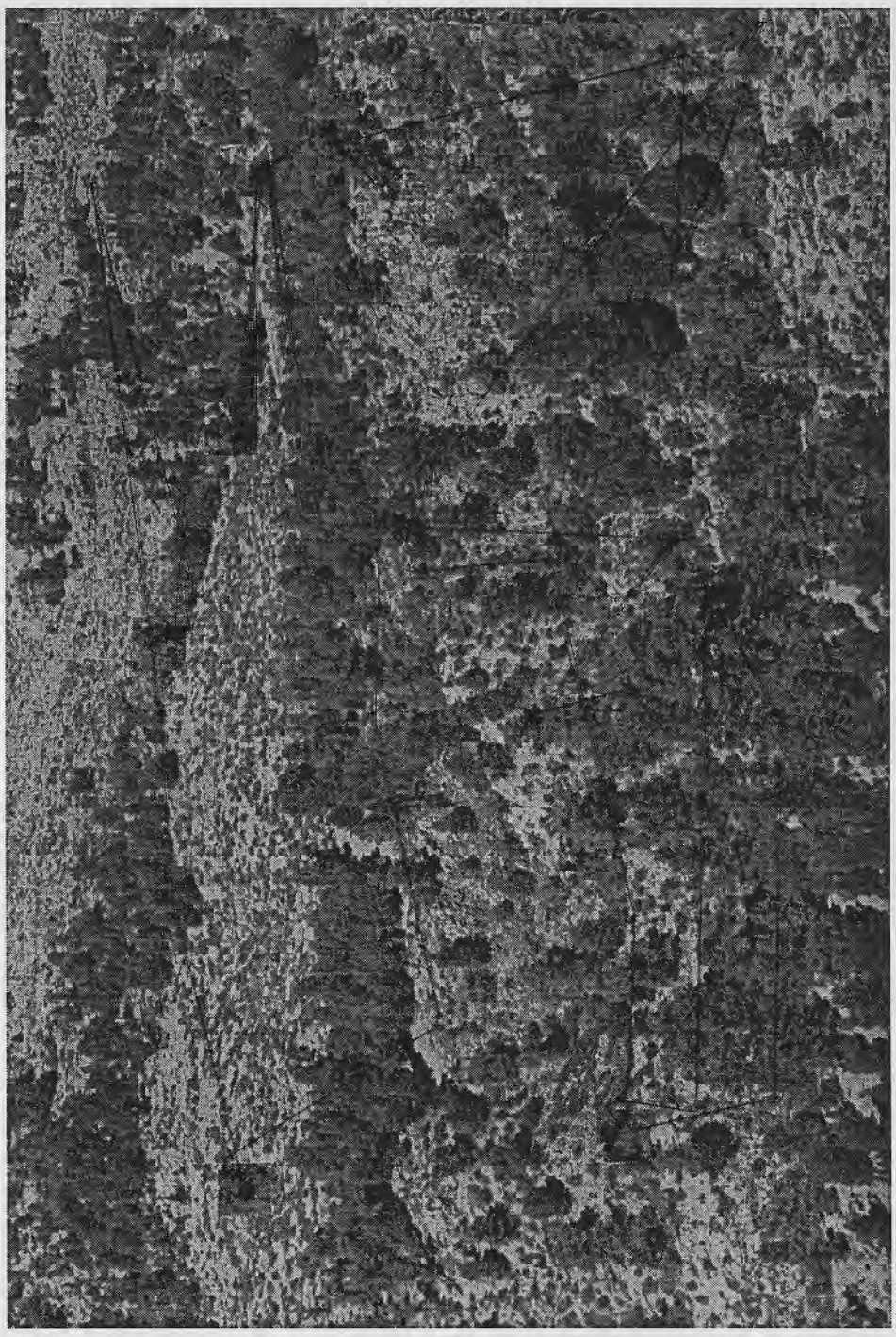


Figure E.23: Observer a's fixations for image stimulus H.

Figure E.24: Image stimulus H.

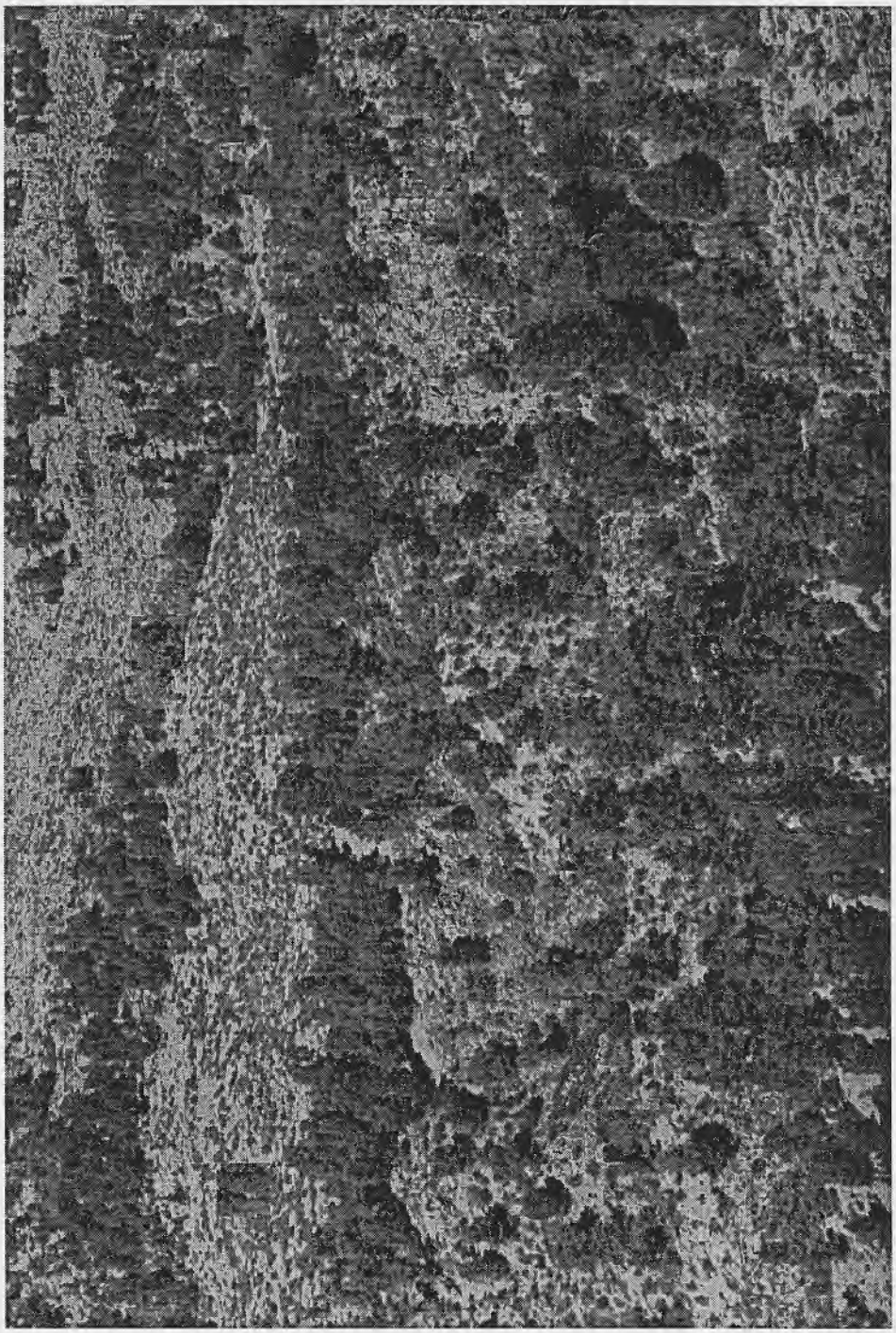


Figure E.24: Image stimulus H.

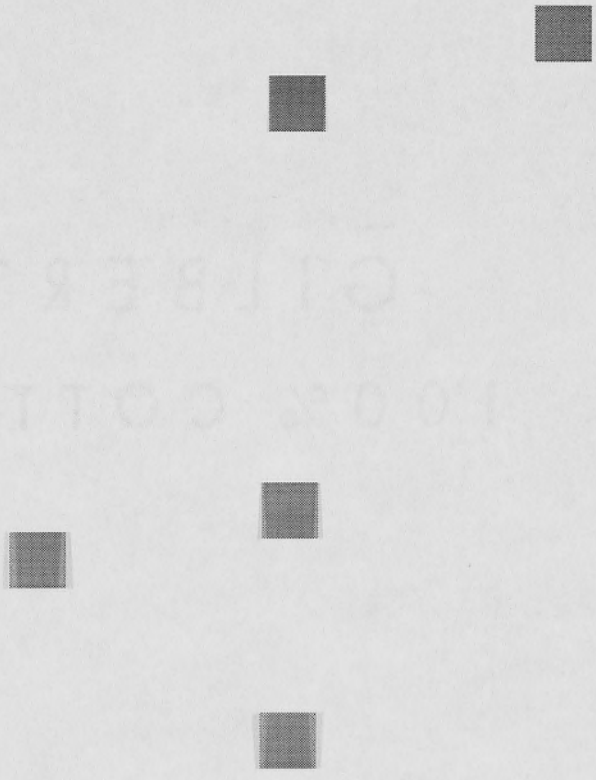


Figure E.25: Ground truth for image stimulus I.

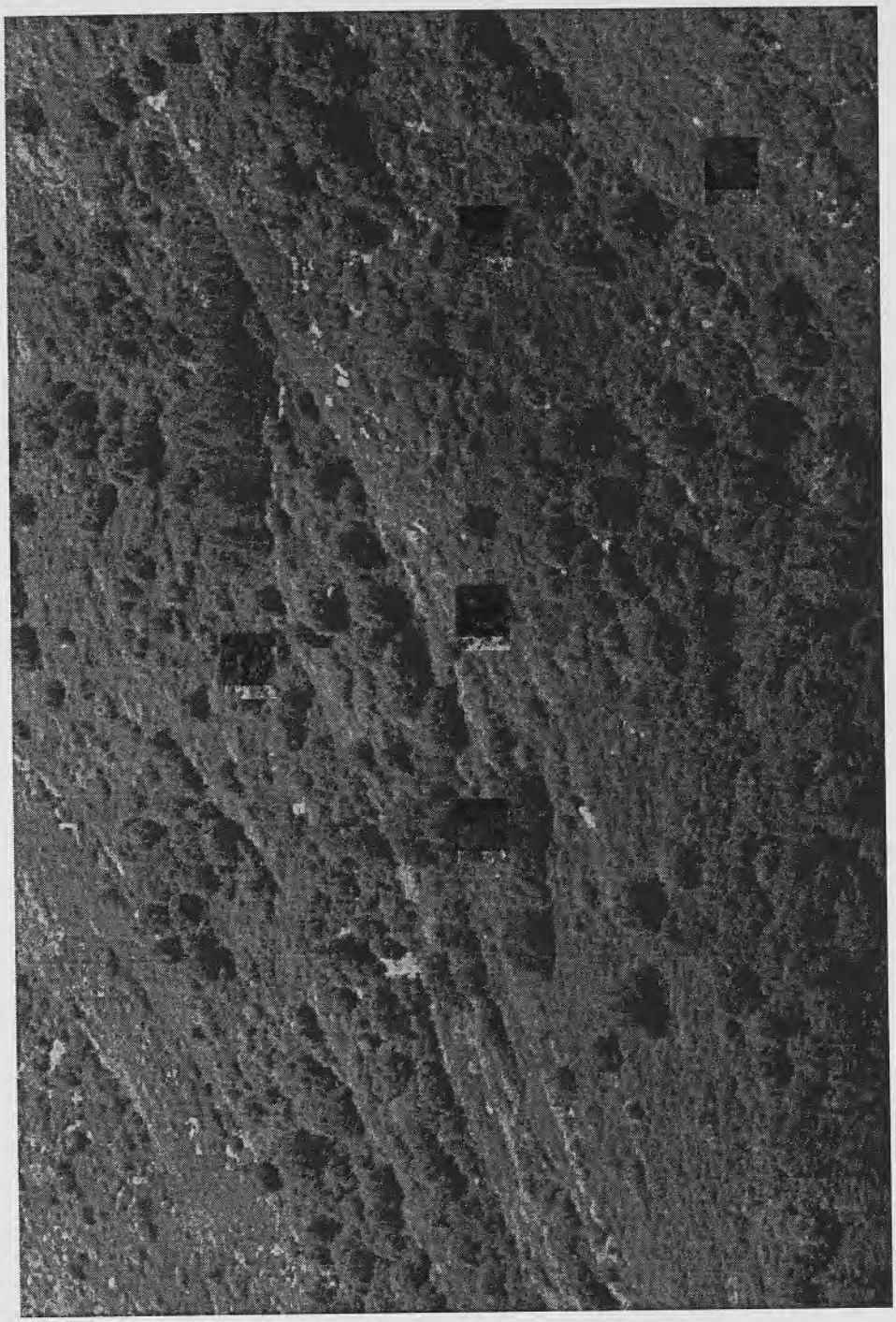


Figure E.25: Ground truth for image stimulus I.
Figure E.27: Image stimulus I.

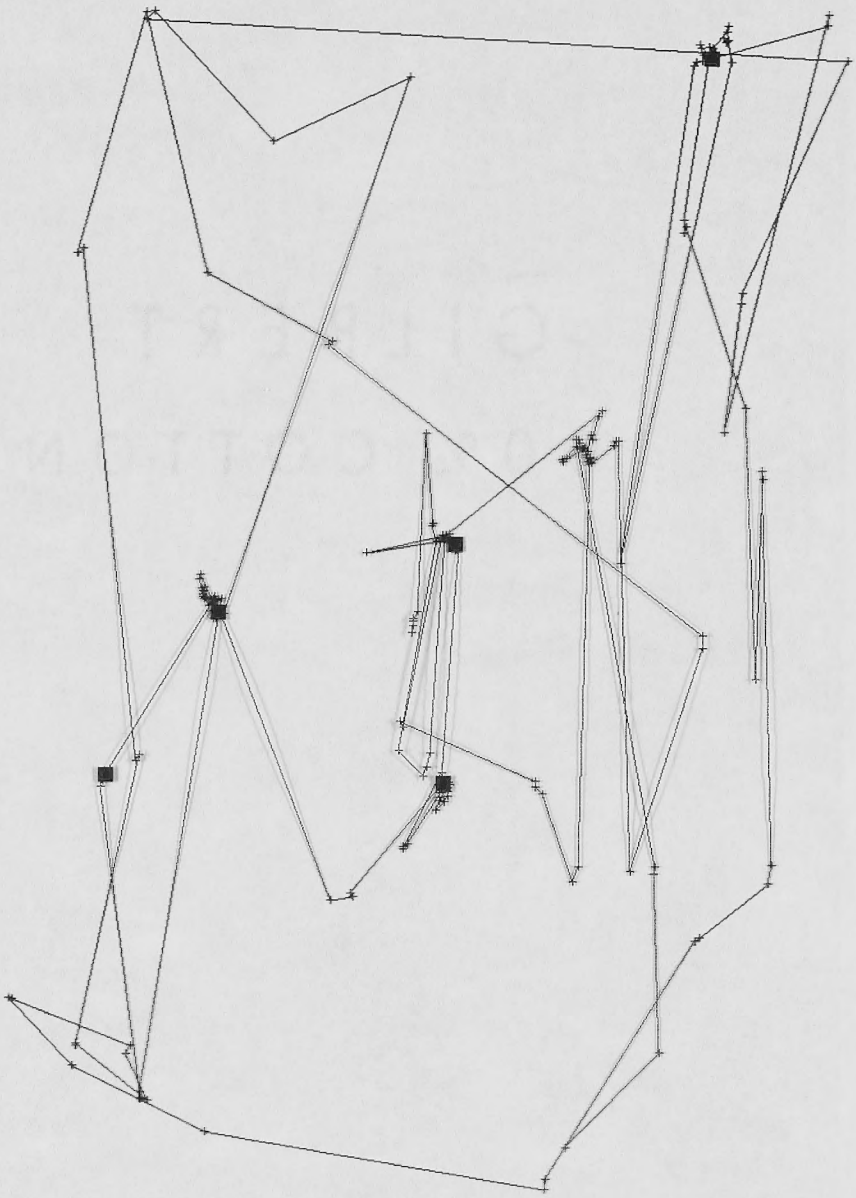


Figure E.26: Observer a's fixations for image stimulus I.

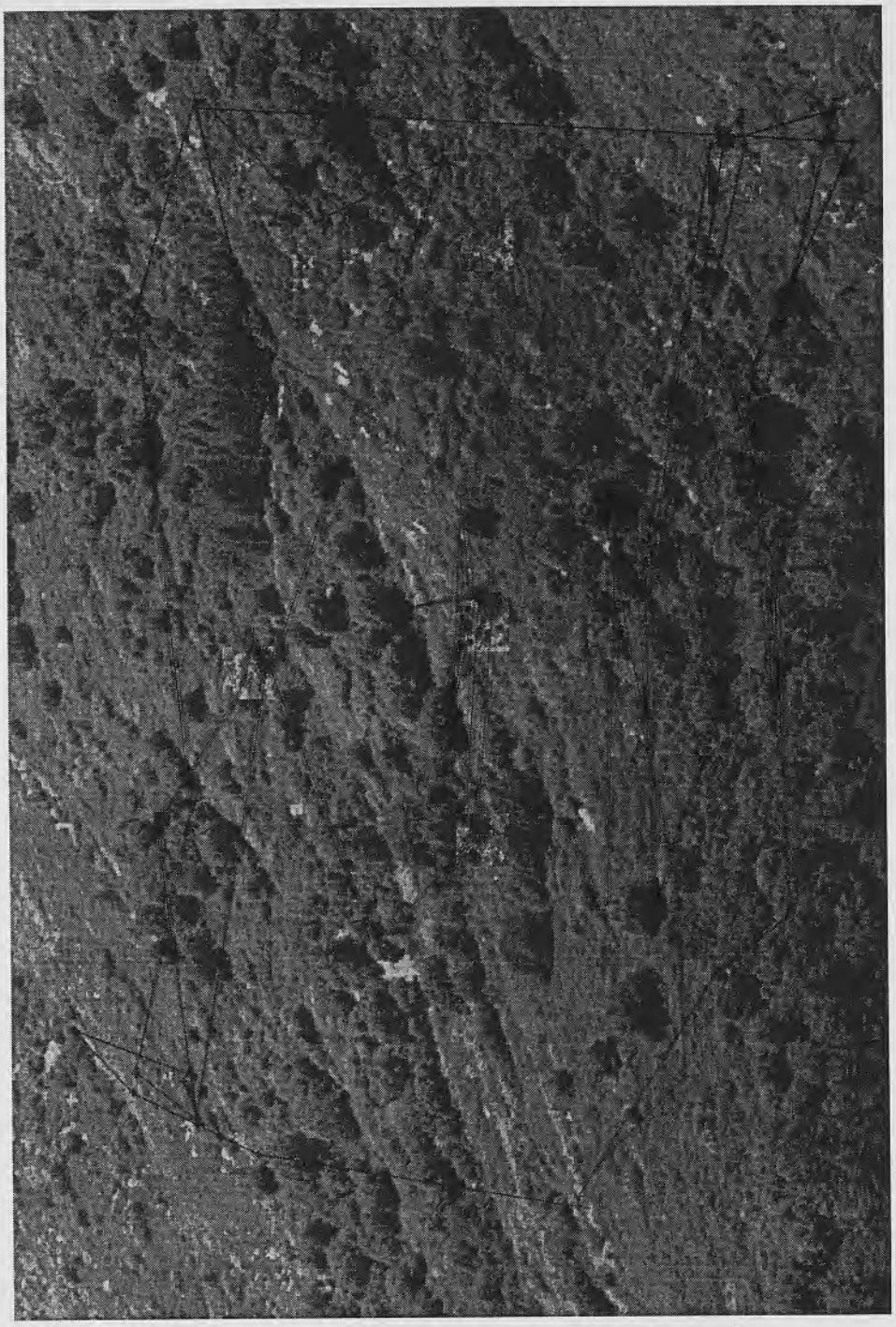


Figure E.26: Observer a's fixations for image stimulus I.
Figure E.27: Image stimulus I.

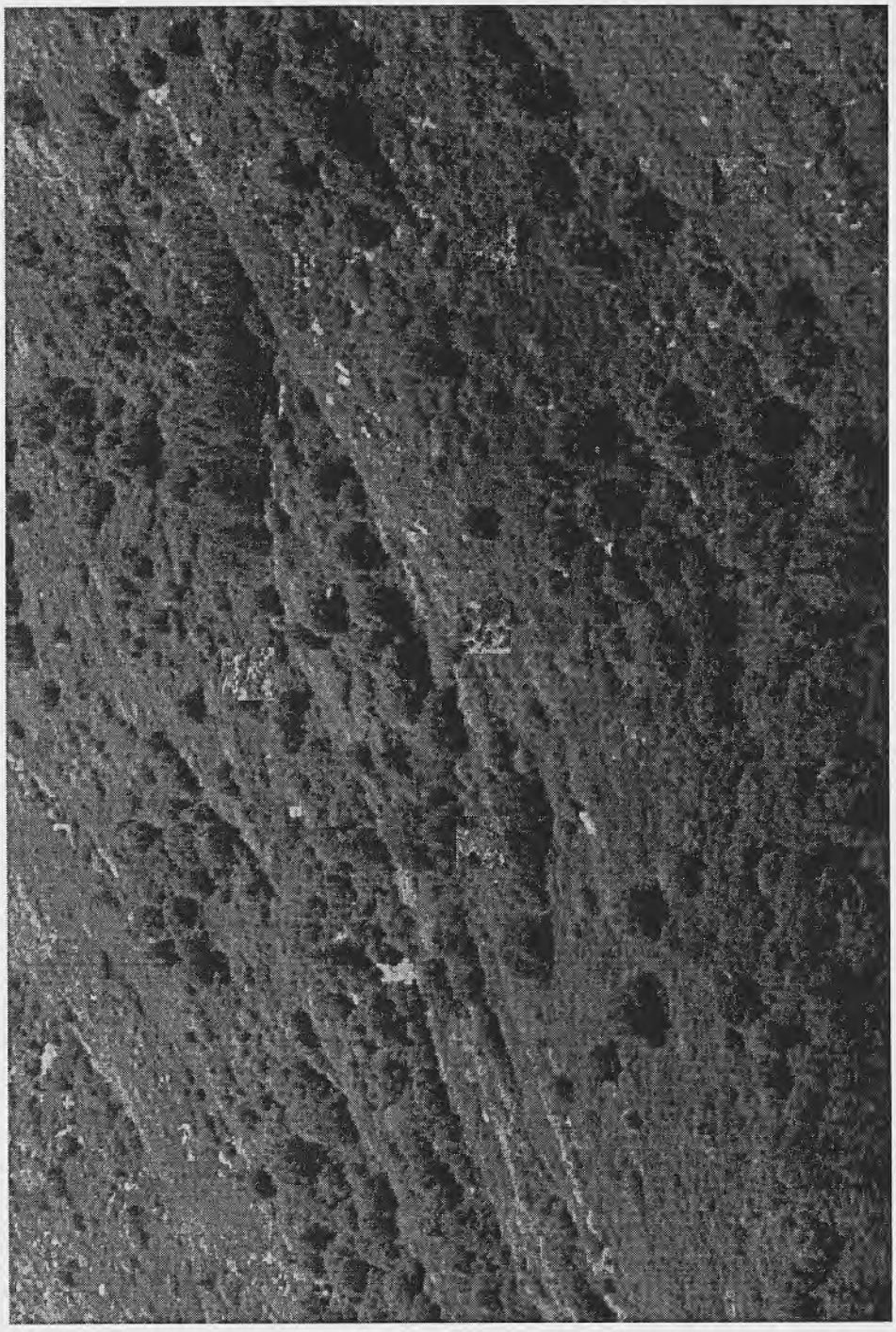


Figure E.27: Image stimulus I.

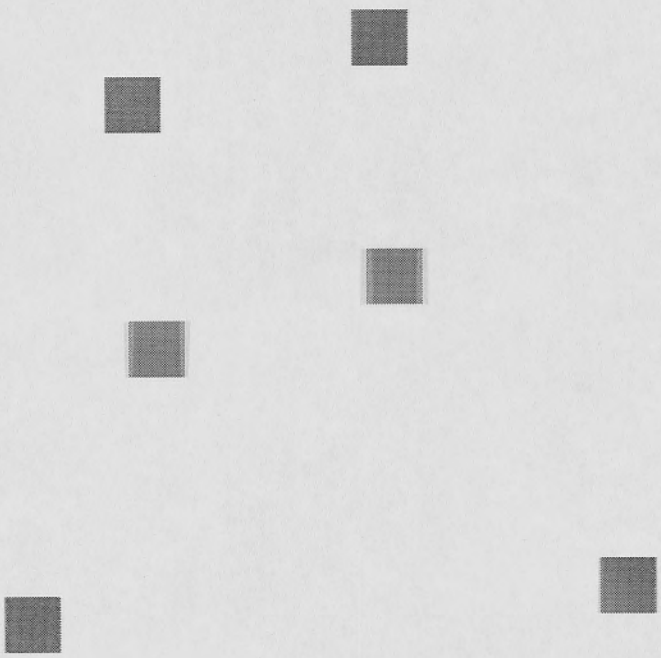


Figure E.28: Ground truth for image stimulus J.



Figure E.28: Ground truth for image stimulus J.

Figure E.30: Image stimulus J.

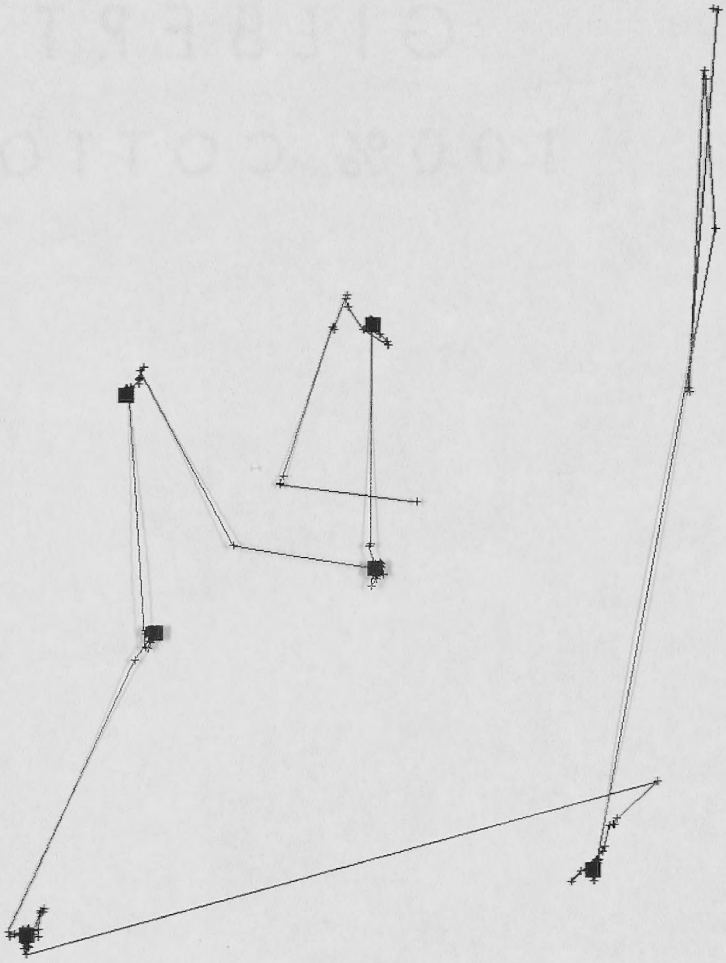


Figure E.29: Observer a's fixations for image stimulus J.

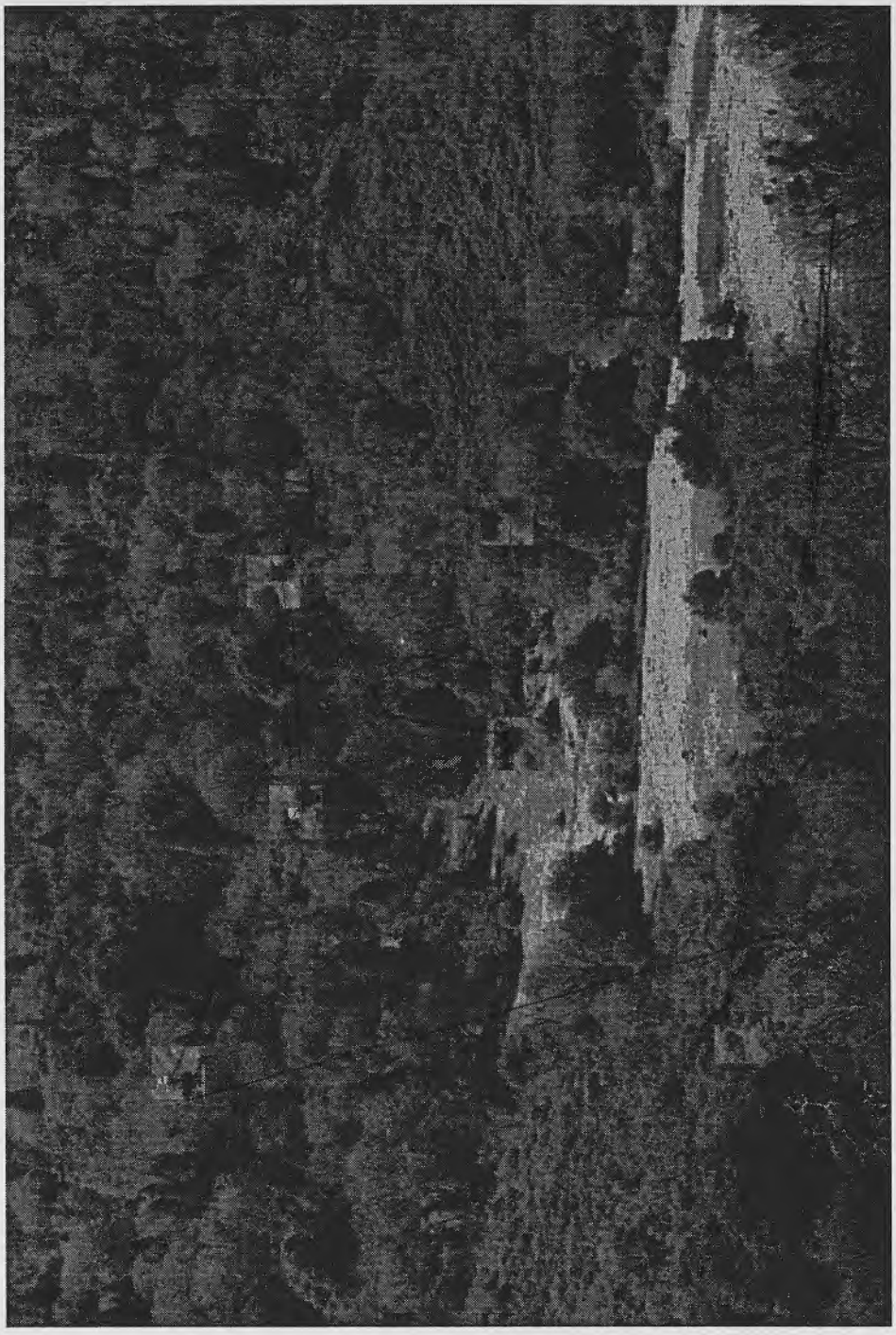


Figure E.29: Observer a's fixations for image stimulus J.

Figure E.30: Image stimulus J.

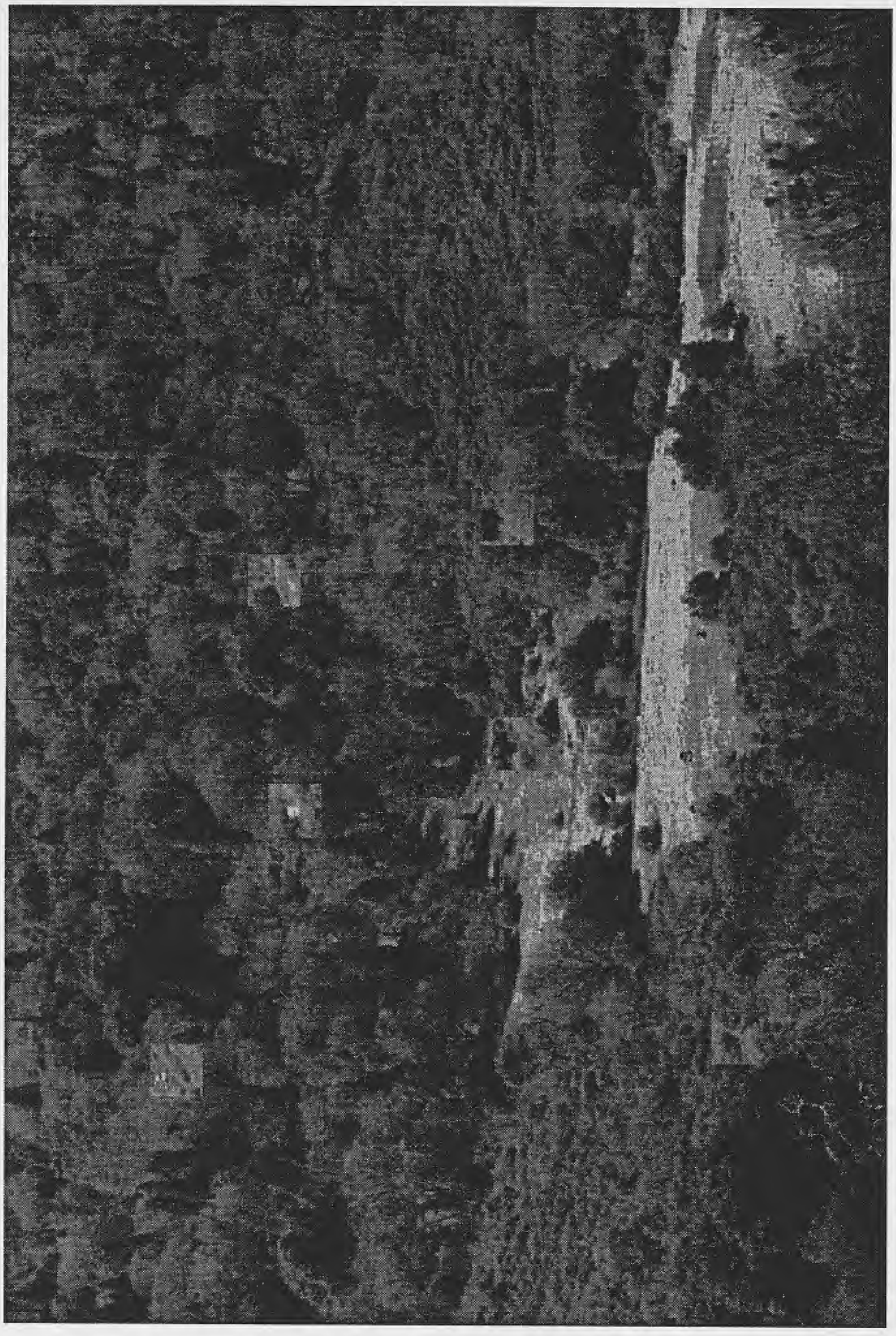
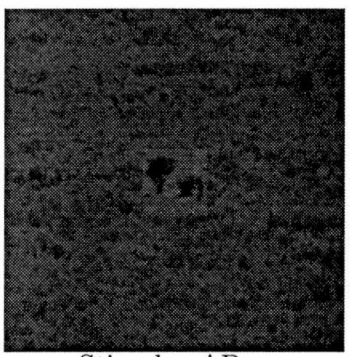


Figure E.30: Image stimulus J.

APPENDIX C

The 15 Natural Texture Experiment Stimuli



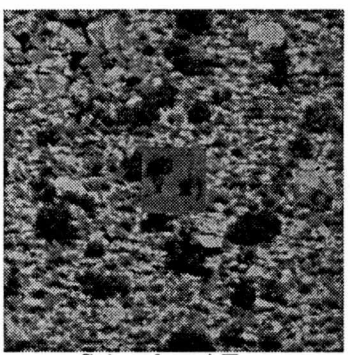
Stimulus AB



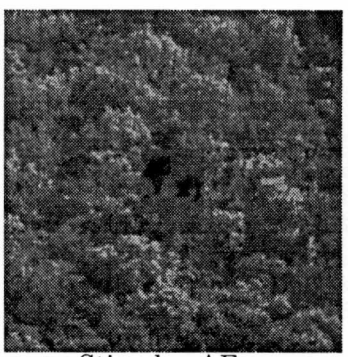
Stimulus AC



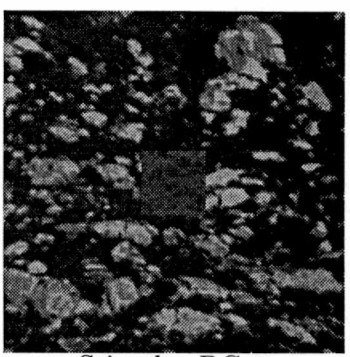
Stimulus AD



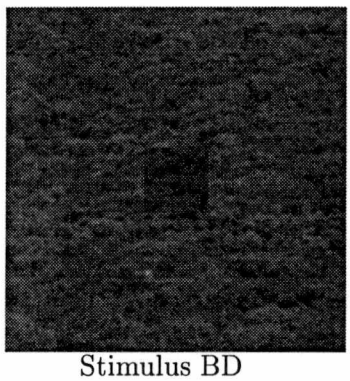
Stimulus AE



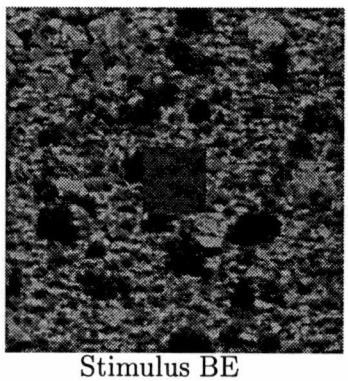
Stimulus AF



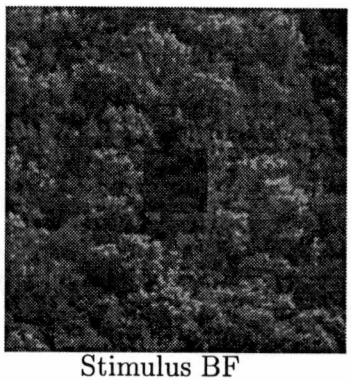
Stimulus BC



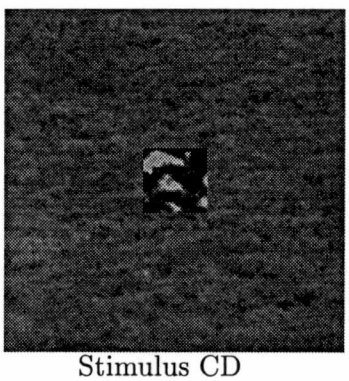
Stimulus BD



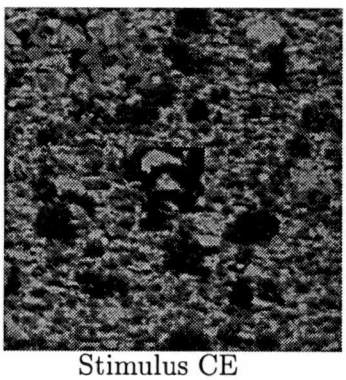
Stimulus BE



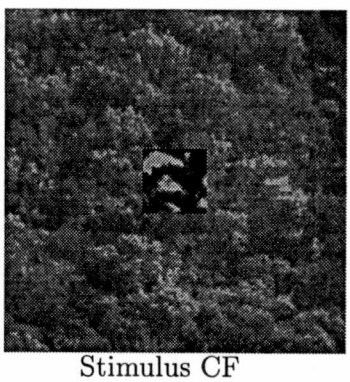
Stimulus BF



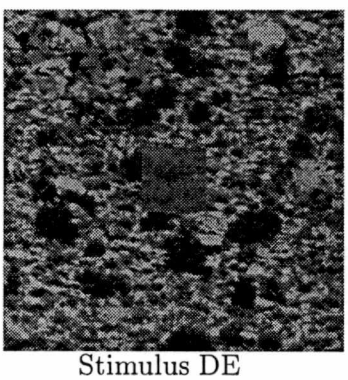
Stimulus CD



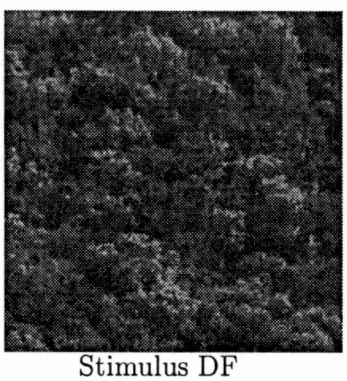
Stimulus CE



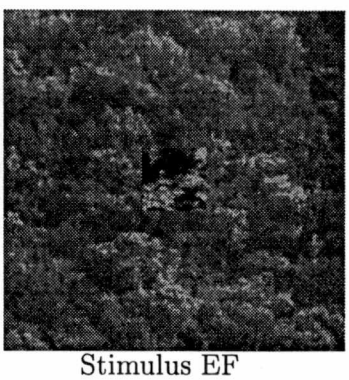
Stimulus CF



Stimulus DE



Stimulus DF



Stimulus EF

APPENDIX D

Target Locations for the Search Experiment Stimuli

Target	Top Row	Left Column
A1	301	696
A2	441	183
A3	515	883
A4	358	423
A5	649	648
B1	148	1045
B2	488	1034
B3	232	410
B4	222	847
B5	485	296
C1	274	469
C2	600	691
C3	575	496
C4	123	200
D1	124	362
D2	434	862
D3	251	620
D4	492	154
D5	503	626
D6	629	425
E1	243	438
E2	231	206
E3	582	986
E4	634	221
E5	450	772
E6	651	590

Target	Top Row	Left Column
F1	247	308
F2	426	944
F3	292	541
F4	495	415
F5	127	882
G1	322	657
G2	544	759
G3	613	377
G4	160	977
H1	535	151
H2	180	793
H3	502	415
H4	191	102
H5	115	597
H6	194	1027
I1	408	414
I2	642	1018
I3	410	610
I4	195	568
I5	416	958
J1	443	505
J2	132	207
J3	219	653
J4	642	239
J5	239	443
J6	431	710

APPENDIX E

Sample Fixation Data for the Search Experiment

APPENDIX F

Computed Statistics for the Search Experiment Targets

Target	P_{ID}	$\overline{T_{<ID}}$	P_{fix}	$\overline{T_{<fix}}$	$\overline{T_{fix}}$
A1	1.0000	6.2550	1.0000	4.4842	1.5996
A2	0.4167	18.3120	0.9167	16.8609	1.6440
A3	0.8333	13.8760	1.0000	9.3175	2.4549
A4	1.0000	5.5875	1.0000	3.3833	1.5330
A5	1.0000	8.3117	1.0000	6.4750	1.6551
B1	0.7500	20.2244	1.0000	10.7192	2.4549
B2	0.1667	15.9150	1.0000	12.6000	1.3774
B3	1.0000	3.4850	1.0000	1.5842	1.6774
B4	0.8333	10.3400	0.9167	7.6736	1.6329
B5	0.9167	10.1509	0.9167	6.5055	1.4219
C1	1.0000	9.2950	1.0000	2.4383	1.6774
C2	0.8333	19.9440	0.8333	13.1910	2.1883
C3	1.0000	13.3742	1.0000	9.9175	1.6551
C4	0.9167	3.0136	1.0000	1.4250	1.4441
D1	1.0000	7.3025	1.0000	5.2767	1.7662
D2	0.9167	8.7891	1.0000	5.8667	1.7440
D3	0.5833	18.8871	0.6667	14.6063	1.5107
D4	0.7500	19.9544	1.0000	14.8775	1.9440
D5	1.0000	5.9017	0.9167	4.3127	1.2552
D6	0.7500	12.5156	0.7500	9.0678	1.4219
E1	0.9167	15.1609	1.0000	10.1083	2.5882
E2	0.5000	16.5667	0.7500	21.1056	1.1331
E3	1.0000	7.3000	1.0000	5.2083	1.8551
E4	0.7500	13.5322	0.9167	12.3155	1.7996
E5	0.5833	20.9929	1.0000	10.6542	1.9884
E6	0.7500	19.2144	1.0000	9.8867	2.1550

Target	P_{ID}	$\overline{T_{<ID}}$	P_{fix}	$\overline{T_{<fix}}$	$\overline{T_{fix}}$
F1	0.0833	16.5500	1.0000	10.9017	1.4330
F2	0.5000	11.3100	0.7500	8.7267	1.5885
F3	0.0833	5.7400	0.9167	8.4155	1.1442
F4	0.7500	17.1689	0.8333	10.3940	1.8107
F5	0.9167	6.3727	0.9167	3.7809	2.1217
G1	0.1667	31.5500	0.9167	11.4464	2.5882
G2	0.5833	14.8386	0.9167	8.0200	2.5105
G3	0.2500	27.1600	1.0000	16.0208	1.7107
G4	0.9167	5.4973	1.0000	3.8550	1.4552
H1	0.9167	10.2182	1.0000	7.9750	2.1217
H2	0.6667	15.8700	1.0000	21.7933	1.7551
H3	0.1667	30.5650	0.5833	22.4986	0.7443
H4	0.9167	5.3355	1.0000	3.2142	1.8107
H5	0.2500	8.4933	0.9167	13.4591	1.3219
H6	0.5000	20.2483	0.9167	17.3436	2.2439
I1	0.5000	13.6500	1.0000	7.0575	1.8551
I2	0.8333	17.1010	0.9167	12.7527	2.0773
I3	0.9167	4.8255	1.0000	2.9692	1.5663
I4	0.9167	10.7955	1.0000	8.7000	1.7773
I5	0.7500	14.1489	0.9167	9.7764	1.7996
J1	0.7500	10.2767	1.0000	7.6492	1.5552
J2	1.0000	4.2975	1.0000	2.6458	1.5552
J3	0.9167	7.2582	1.0000	5.7200	1.4108
J4	0.9167	13.6964	1.0000	11.1133	1.6663
J5	1.0000	5.6875	1.0000	3.4492	1.3774
J6	0.9167	9.5091	0.9167	8.4273	1.2775

APPENDIX G

Computed Target Signature Metrics for the Search Experiment Targets

Target	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
A1	3.20e+1	3.46e+1	1.90e+2	1.46e+0	2.79e-1	9.33e-1	1.30e-1	2.67e-2	1.01e-1	2.24e-4	4.14e+1	6.41e+0
A2	4.32e+0	6.48e+0	0.00e+0	4.88e-1	9.22e-2	3.28e-1	2.30e-2	4.61e-3	2.05e-2	7.43e-5	1.50e+1	2.70e+0
A3	1.68e+1	1.68e+1	3.64e+2	9.90e-1	1.88e-1	7.05e-1	9.09e-2	1.87e-2	7.65e-2	9.17e-5	2.48e+1	4.89e+0
A4	2.88e+1	3.01e+1	0.00e+0	1.18e+0	2.22e-1	7.74e-1	1.02e-1	2.05e-2	8.32e-2	1.47e-4	2.86e+1	4.64e+0
A5	3.11e+1	3.54e+1	0.00e+0	1.24e+0	2.39e-1	8.56e-1	1.03e-1	2.10e-2	8.38e-2	2.50e-4	4.73e+1	8.80e+0
B1	1.10e+1	1.46e+1	1.00e+0	6.13e-1	1.13e-1	3.07e-1	1.01e-2	1.97e-3	6.76e-3	3.03e-5	4.46e+1	4.59e+0
B2	4.37e+1	5.09e+1	0.00e+0	1.24e+0	2.31e-1	6.63e-1	8.27e-2	1.72e-2	7.14e-2	5.57e-5	5.26e+1	5.58e+0
B3	5.05e+1	5.13e+1	8.05e+2	1.51e+0	2.85e-1	8.46e-1	6.99e-2	1.42e-2	4.84e-2	5.18e-5	5.09e+1	5.29e+0
B4	2.80e+1	3.19e+1	0.00e+0	1.02e+0	1.93e-1	5.36e-1	3.80e-2	7.55e-3	2.40e-2	1.30e-4	3.77e+1	5.16e+0
B5	4.40e+0	9.32e+0	0.00e+0	6.78e-1	1.29e-1	3.27e-1	1.05e-2	2.23e-3	6.45e-3	1.02e-5	4.66e+1	4.67e+0
C1	2.01e+1	2.10e+1	4.68e+2	8.79e-1	1.66e-1	4.45e-1	2.01e-2	4.06e-3	1.24e-2	3.77e-5	2.96e+1	2.79e+0
C2	1.10e+1	1.24e+1	1.30e+1	7.24e-1	1.32e-1	3.83e-1	1.77e-2	3.16e-3	1.17e-2	9.02e-6	2.90e+1	2.86e+0
C3	2.49e+1	2.51e+1	0.00e+0	1.01e+0	1.90e-1	6.16e-1	1.20e-1	2.47e-2	1.13e-1	1.01e-5	4.35e+1	4.57e+0
C4	3.20e+1	3.38e+1	1.05e+3	1.29e+0	2.43e-1	6.85e-1	4.54e-2	9.21e-3	2.93e-2	6.50e-5	5.06e+1	5.10e+0
D1	2.28e+1	3.38e+1	0.00e+0	1.53e+0	2.92e-1	9.35e-1	9.93e-2	1.97e-2	7.49e-2	1.19e-4	3.52e+1	3.94e+0
D2	2.13e+1	2.24e+1	6.82e+2	9.98e-1	1.88e-1	5.26e-1	3.13e-2	6.08e-3	2.04e-2	3.09e-5	3.42e+1	3.91e+0
D3	1.58e+1	1.64e+1	5.60e+1	8.84e-1	1.66e-1	4.54e-1	1.85e-2	3.67e-3	1.13e-2	9.79e-6	3.28e+1	3.21e+0
D4	7.55e+0	1.32e+1	4.71e+2	8.94e-1	1.72e-1	4.79e-1	2.97e-2	6.11e-3	1.99e-2	2.30e-5	2.68e+1	3.42e+0
D5	2.49e+1	2.68e+1	9.17e+2	1.26e+0	2.39e-1	6.65e-1	5.02e-2	1.01e-2	3.29e-2	2.73e-5	3.30e+1	3.66e+0
D6	2.50e+0	3.99e+0	1.86e+2	6.70e-1	1.28e-1	4.23e-1	2.78e-2	5.55e-3	2.39e-2	8.15e-5	2.71e+1	3.65e+0
E1	5.33e-1	7.17e+0	4.00e+0	5.50e-1	1.02e-1	3.13e-1	1.41e-2	2.66e-3	9.76e-3	5.57e-5	3.10e+1	3.88e+0
E2	5.54e+0	5.78e+0	4.80e+1	6.36e-1	1.19e-1	3.60e-1	1.97e-2	3.89e-3	1.28e-2	1.69e-4	2.13e+1	3.00e+0
E3	2.64e+1	3.97e+1	0.00e+0	1.58e+0	3.02e-1	1.04e+0	1.35e-1	2.70e-2	1.11e-1	1.06e-4	6.07e+1	8.54e+0
E4	1.47e+1	2.58e+1	0.00e+0	1.34e+0	2.53e-1	7.31e-1	5.74e-2	1.15e-2	3.81e-2	3.43e-5	2.89e+1	3.16e+0
E5	5.34e+0	1.97e+1	0.00e+0	1.06e+0	1.98e-1	6.01e-1	4.51e-2	8.95e-3	3.31e-2	9.09e-6	3.17e+1	4.03e+0
E6	1.82e+0	1.30e+1	0.00e+0	9.03e-1	1.71e-1	5.03e-1	3.07e-2	6.46e-3	2.20e-2	2.90e-6	2.35e+1	2.92e+0

Target	Δ T	Doyle	Eff_POT	ACE	WACE	AWCE	ASCE	WASCE	AWSCE	VOE	ABS	RABS
F1	9.65e+0	1.85e+1	0.00e+0	1.01e+0	1.90e-1	5.82e-1	5.05e-2	9.90e-3	3.63e-2	2.53e-4	2.13e+1	2.71e+0
F2	2.67e+0	6.13e+0	2.70e+1	8.49e-1	1.58e-1	5.00e-1	2.78e-2	5.54e-3	2.17e-2	4.66e-5	3.78e+1	4.22e+0
F3	1.60e+1	1.79e+1	4.00e+0	7.93e-1	1.49e-1	4.57e-1	2.65e-2	5.32e-3	1.84e-2	1.55e-4	2.66e+1	3.49e+0
F4	1.72e+1	1.72e+1	4.00e+0	7.76e-1	1.46e-1	4.07e-1	1.74e-2	3.62e-3	1.12e-2	5.48e-5	4.16e+1	4.36e+0
F5	2.78e+1	2.95e+1	0.00e+0	9.32e-1	1.78e-1	5.06e-1	3.01e-2	6.39e-3	2.07e-2	1.58e-4	3.39e+1	3.80e+0
G1	1.02e+1	1.06e+1	1.77e+2	6.05e-1	1.11e-1	3.08e-1	8.75e-3	1.67e-3	5.47e-3	3.13e-6	3.70e+1	3.64e+0
G2	3.97e-1	7.65e+0	3.50e+1	5.54e-1	1.04e-1	3.54e-1	2.44e-2	4.81e-3	2.20e-2	2.13e-5	4.56e+1	5.21e+0
G3	6.31e+0	6.31e+0	1.90e+1	3.73e-1	6.92e-2	2.08e-1	6.31e-3	1.24e-3	4.21e-3	2.41e-5	2.89e+1	3.24e+0
G4	2.57e+1	2.81e+1	6.36e+2	9.71e-1	1.82e-1	4.81e-1	1.95e-2	3.89e-3	1.17e-2	1.73e-5	4.08e+1	3.81e+0
H1	2.42e+1	2.47e+1	2.16e+2	9.91e-1	1.83e-1	5.15e-1	2.43e-2	4.81e-3	1.53e-2	8.87e-6	3.61e+1	3.57e+0
H2	1.20e+1	1.20e+1	1.18e+2	6.17e-1	1.14e-1	2.98e-1	9.77e-3	1.87e-3	5.30e-3	1.36e-5	4.26e+1	4.51e+0
H3	1.88e+1	2.22e+1	0.00e+0	8.11e-1	1.51e-1	4.05e-1	1.53e-2	2.94e-3	8.88e-3	3.14e-5	2.95e+1	2.95e+0
H4	3.40e+1	3.54e+1	2.72e+2	1.26e+0	2.38e-1	6.39e-1	4.00e-2	7.87e-3	2.38e-2	6.19e-6	4.03e+1	3.94e+0
H5	3.51e+1	3.51e+1	7.01e+2	1.32e+0	2.49e-1	6.83e-1	4.33e-2	8.56e-3	2.56e-2	1.40e-5	3.59e+1	3.46e+0
H6	8.28e+0	9.73e+0	2.53e+2	5.06e-1	9.62e-2	2.73e-1	1.07e-2	2.22e-3	7.85e-3	2.83e-5	2.98e+1	3.32e+0
I1	1.77e+1	1.78e+1	2.29e+2	8.05e-1	1.50e-1	4.60e-1	2.56e-2	5.08e-3	1.84e-2	5.74e-5	3.14e+1	4.40e+0
I2	2.33e-1	7.11e+0	3.67e+2	5.39e-1	1.02e-1	3.09e-1	1.51e-2	3.10e-3	1.11e-2	1.44e-4	2.96e+1	4.30e+0
I3	4.59e+0	6.65e+0	1.95e+2	7.92e-1	1.53e-1	4.59e-1	2.51e-2	5.33e-3	1.90e-2	8.52e-5	2.66e+1	3.75e+0
I4	2.07e+1	2.07e+1	4.86e+2	9.13e-1	1.72e-1	4.73e-1	3.31e-2	7.06e-3	2.18e-2	4.39e-5	3.30e+1	4.53e+0
I5	8.62e+0	1.11e+1	3.45e+2	5.41e-1	1.04e-1	2.82e-1	1.01e-2	2.14e-3	6.63e-3	1.13e-4	2.74e+1	3.46e+0
J1	2.09e-2	8.92e+0	0.00e+0	7.48e-1	1.36e-1	4.57e-1	2.69e-2	5.21e-3	2.10e-2	4.85e-5	2.74e+1	4.76e+0
J2	4.34e+1	4.35e+1	1.92e+3	1.84e+0	3.52e-1	1.27e+0	2.03e-1	4.21e-2	1.65e-1	9.86e-5	3.33e+1	7.23e+0
J3	3.75e+1	3.91e+1	1.68e+3	1.58e+0	3.02e-1	1.05e+0	1.49e-1	3.06e-2	1.19e-1	1.57e-4	2.89e+1	6.92e+0
J4	2.24e+1	2.34e+1	9.71e+2	1.11e+0	2.09e-1	7.30e-1	7.59e-2	1.52e-2	6.05e-2	5.17e-5	3.01e+1	5.89e+0
J5	3.26e+1	3.43e+1	1.43e+3	1.53e+0	2.90e-1	1.01e+0	1.27e-1	2.60e-2	1.03e-1	1.41e-4	2.78e+1	6.05e+0
J6	1.91e+1	2.08e+1	1.02e+3	1.25e+0	2.36e-1	8.49e-1	1.18e-1	2.33e-2	9.39e-2	1.64e-4	2.49e+1	5.47e+0

VITA

Anthony C. Copeland received the BSEE degree from the U.S. Military Academy at West Point, NY, in 1988, and the MSEE degree from the University of Tennessee, Knoxville, in 1993. He proudly served as a signal corps officer in the U.S. Army from 1988 to 1991. During his military service, he served as a communications node platoon leader for the 24th Infantry Division (Mechanized) based in Fort Stewart, GA. This service included participation in the liberation of Kuwait during the Persian Gulf conflict in 1990-91. He is currently a research engineer at U.C. San Diego. His research interests include image understanding, texture analysis, and human visual perception.