# OPTIMIZATION OF SINGLE AND LAYERED SURFACE TEXTURING

A Dissertation

by

ALETHEA SCATTERGOOD BAIR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2009

Major Subject: Architecture

# OPTIMIZATION OF SINGLE AND LAYERED SURFACE TEXTURING

A Dissertation

by

ALETHEA SCATTERGOOD BAIR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

| | |
|---|---|
| Chair of Committee, | Donald House |
| Committee Members, | Nancy Amato |
| | Ricardo Gutierrez-Osuna |
| | Colin Ware |
| Head of Department, | Glen Mills |

May 2009

Major Subject: Architecture

# ABSTRACT

Optimization of Single and Layered Surface Texturing. (May 2009)

Alethea Scattergood Bair, B.A., University of Illinois Urbana-Champaign

Chair of Advisory Committee: Dr. Donald House

In visualization problems, surface shape is often a piece of data that must be shown effectively. One factor that strongly affects shape perception is texture. For example, patterns of texture on a surface can show the surface orientation from foreshortening or compression of the texture marks, and surface depth through size variation from perspective projection. However, texture is generally under-used in the scientific visualization community. The benefits of using texture on single surfaces also apply to layered surfaces. Layering of multiple surfaces in a single viewpoint allows direct comparison of surface shape. The studies presented in this dissertation aim to find optimal methods for texturing of both single and layered surfaces.

This line of research starts with open, many-parameter experiments using human subjects to find what factors are important for optimal texturing of layered surfaces. These experiments showed that texture shape parameters are very important, and that texture brightness is critical so that shading cues are available. Also, the optimal textures seem to be task dependent; a feature finding task needed relatively little texture information, but more shape-dependent tasks needed stronger texture cues.

Subsequent, narrower experiments investigated specific texture parameters for optimal textures. Textures with significant directionality and structure were found to be useful on both surface layers. A range of viable top surface opacities was found, and relationships between texture sizes were investigated. A feasibility experiment was run to estimate bounds on the number of layers that can be effectively visualized. Finally, a single surface experiment was run to investigate causes of bias and error in principal direction and projected grid textures. Overall performances of principal direction and projected grid textures were not statistically different, but differences were found in the location of errors between the two texture types.

# ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. House, and my committee members, Dr. Ware, Dr. Amato, and Dr. Gutierrez-Osuna, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the Vizlab faculty and staff for making my time at Texas A&M University a great experience. I also want to extend my gratitude to all the students and faculty who were willing to participate in the experiments.

Finally, I offer many thanks to my mother, father, sisters and Sandy for their love and encouragement. I want to thank Christina and Eric for being wonderful roommates and the best of friends, Mayank for his kind friendship and willingness to discuss research problems at any time, and Brad for his support through both bad and good times.

# NOMENCLATURE

| | |
|---|---|
| ANOVA | Analysis of Variance |
| PCA | Parallel Coordinate Analysis |
| LDA | Linear Discriminant Analysis |

# TABLE OF CONTENTS

Page

Page

# LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

# LIST OF TABLES

# 1. INTRODUCTION

## 1.1 Motivation

Scientific disciplines widely use visualization as a way to show data. Although data can be shown in many formats, including numbers, visualization is a good choice for data representation because the visual system is the highest bandwidth sensory input in humans. As a result, images are an excellent way to show large quantities of data, as well as a natural way to show spatially-distributed data. Spatially-distributed data means data with inherent spatial properties, such as a medical scan of a human body, as opposed to data that is not inherently spatial, like nodal relationships in a graph. Often, it is necessary or desirable to show layered data. Figure 1 shows an example of some layered, spatially-distributed data: a scan of a human head. Two common visualization methods are shown: volume and surface visualization.

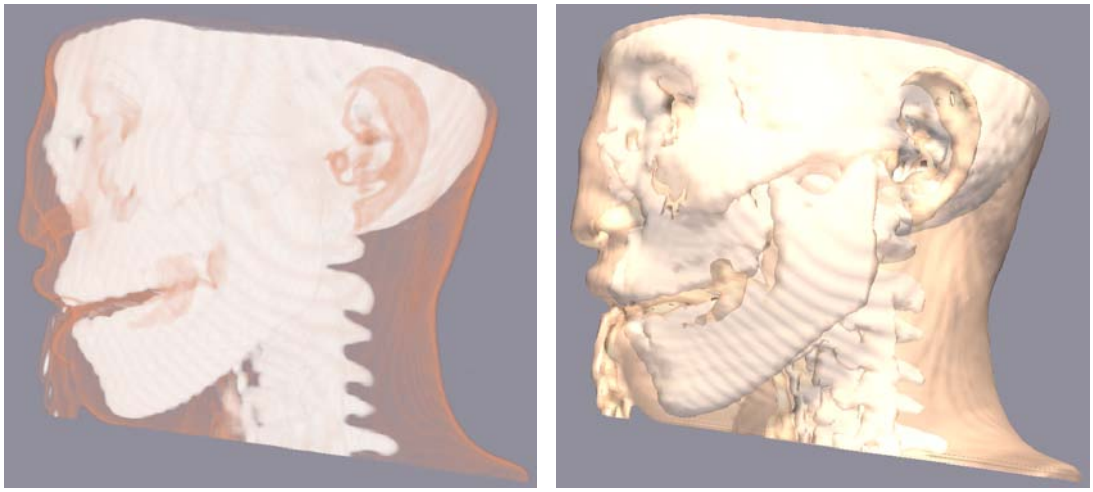


*Figure 1.Volume Rendering and Surface Shading Comparison. The data shown is from the Visible Human data set, [Squillacote 2006], of a human head, using volume and surface rendering. Layers are shown that correspond to skin and bone.*

---

This dissertation follows the style of *ACM Transactions on Graphics.*

Reasons for visualizing layered data are varied. For example, in the medical field, doctors must understand spatial relationships between body tissue layers for surgical planning. Geologists are interested in sediment layers for understanding rock formation as well as predicting locations of natural resources such as oil and gas. Technical illustrations of machinery, anatomy, or architectural drawings often need to use layered surfaces to show relationships between parts of an object. Finally, disciplines such as meteorology and oceanography use isosurfaces of variables like pressure or temperature to analyze the dynamics of complicated flow systems.

One important topic of research is the rendering of layered surfaces. When data is layered, there is an inherent problem when data on one layer obscures and conflicts with other data. The main challenge of visualizing layered data is showing each layer of data as clearly as possible, while minimizing interference across data layers.

One way to significantly improve surface visualization is to add texture. Figure 2 shows how the addition of texture can clarify surface shape in areas where shading provides no little or no information, such as the front of the hoof. However, care must be taken in the choice of textures because there can be a strong tendency toward visual conflict between the surfaces. Since the possible texture variations are enormous, finding optimal textures is not an easy task.

The approach used here is to conduct a series of experiments designed to find guidelines for using texture. Results from each experiment are synthesized with previous work in surface visualization and texturing. A simple model is hypothesized for how to optimize texture, which will be built based on expert knowledge, the structure of the human visual system, and previous experiments in rendering and texturing of surfaces. The model consists of guidelines for how to create good textures, and theories for why the textures are good. It explores what features are most important in layered surface texturing, and ways in which the features interact both within and between surfaces.

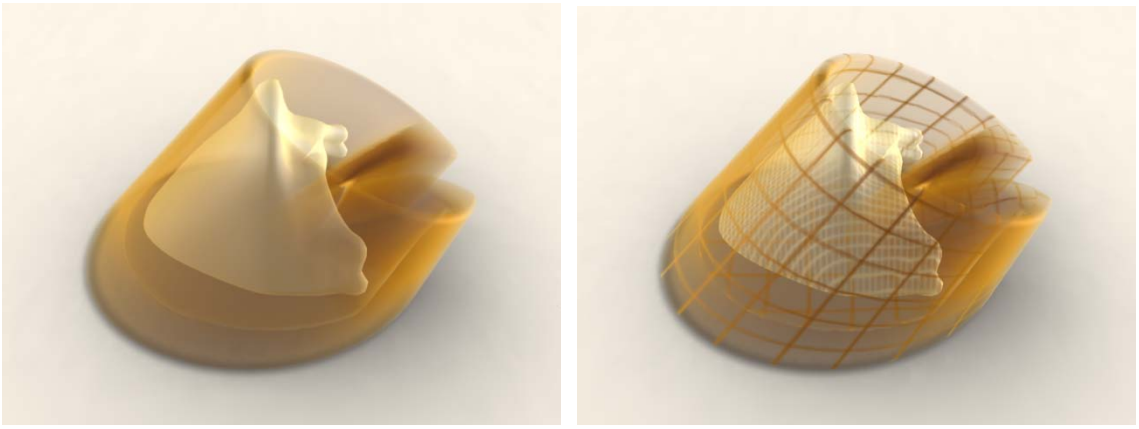


*Figure 2.Textured and Shaded Comparison. A horse hoof is shown with an outer wall and an inner coffin bone with shaded surfaces (left) and textured and shaded surfaces (right).*

Many factors go into creating a good surface visualization. Some examples might include lighting and shading, coloring and opacity, texture, shadows, motion, or non-photo-realistic rendering techniques such as edge enhancement. Some of the other perceptual cues are discussed in more detail in the background section. Many of these cues are commonly used in scientific, and others are not for various reasons. One perceptual cue that has a lot of flexibility is texture. Textures have a broad space of possibilities, and yet little research has been done on what factors make the best textures.

### 1.2 Research Focus

The focus of this dissertation is on texturing of surfaces. Texturing is one of the more powerful, and yet rarely used cues in surface rendering. It has been shown to be a useful perceptual cue both in single and layered surface visualization. However since it is seldom used in scientific visualization and has only occasionally been researched in perceptual studies, the huge space of texture possibilities is not well understood. One motivation for looking at texture is that it is a part of every real-world object. Even seemingly smooth objects nearly always have a very fine texture such as dust or scratches. These textures provide perceptual cues to how the surface is shaped. Figure 3 shows examples of small and large-scale textures that aid in surface shape perception.

*Figure 3. Texture Examples. A cloth's texture shows how it is folded. The texture of the rock shows how geological features are carved into the sediment layers in Big Bend National Park.*

Studies on optimization of textures have only been done for a few parameters. These generally studied the directionality of the textures. Comparisons have been made of isotropic (rotationally invariant) vs anisotropic textures [Sweet and Ware 2004 ; Interrante and Kim 2001], textures with multiple directions, and various heuristics for aligning the texture direction with surface characteristics. However, no evaluation has yet been done on stylistic aspects of texturing such as color, regularity and size. These parameters could be of critical importance for layered surface texturing, especially when interaction between the surfaces is considered.

Fully specifying the stylistic aspect of layered surface texturing requires many parameters and allows for many possible interactions among parameters. This type of problem cannot be studied with only a few controlled experiments because the number of variables is too great. Therefore, this set of studies uses two types of experiments combined with perceptual theory to triangulate results and develop theories for texturing of layered surfaces. Since little is known about the stylistic aspects of layered surface texturing, initial experiments were geared toward attaining general results about a broad spectrum of parameters. These experiments were not designed to yield statistically significant results for a large number of parameters because the amount of data required to sample a large parameter space is unfeasible. However trends were noted in order to guide the design of more specific experiments measuring the significance and optimal values for only a few parameters. These later experiments were designed to support,

disprove, and/or fine-tune theory based on the broader experiments. The results are heuristics and perceptual theories based on both broad and deep investigation into layered surface texturing. These theories will be useful to application developers that want to show layered surfaces for use in scientific analysis or explanatory visualizations.

### 1.3 Texture Parameterization

To explore the possible texture variations, it is necessary to define what a texture is, and find a set of parameters for creating textures. We consider textures to be made up of some finite set of marks. Marks have a number of features, which when varied can create any arbitrary texture. This set of features is shape, size, direction, placement, coloring and opacity, as shown in Figure 4.

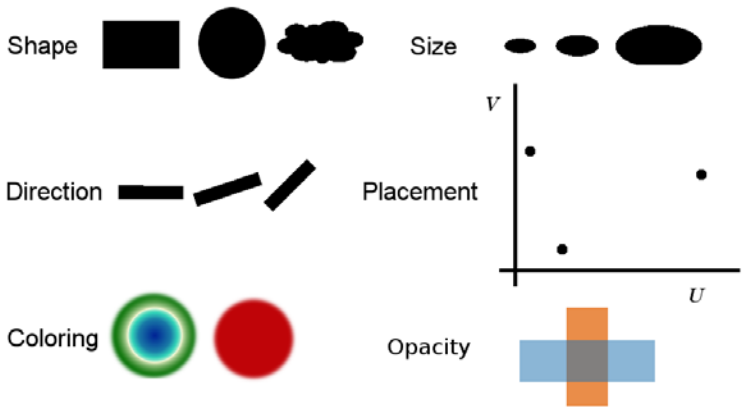



*Figure 4. Texture Mark Parameterization. Examples of possible variations in shape, size, direction, placement, coloring and opacity.*

Any of these features may be sub-parameterized for specific types of textures. For example, coloring might be parameterized as a hue, saturation and value. Shape might be parameterized by the major and minor axes of an ellipse, and placement might be parameterized by texture coordinates or shape parameters of the surface. Parameterization of mark features in this way sets constraints on the possible textures that can be created, as well as the distribution of textures.

In theory, each mark could vary each property independently, and the combination of all marks would create the visualization. However in practice, simplifying assumptions were made, and some attempt is made to account for the interactions between marks. In this series of studies, marks are grouped into sets with identical or similar shape, size and coloring, and the placement of each mark is not independent, but rather takes place in a general organization of marks. The background is dealt with separately as having a color and opacity. Also, the shape parameter is simplified to only a few simple cases, such as lines, curves, rectangles, ovals or dots.

Within these five broad parameters, quite a bit of interaction is possible. Following is a list of known interactions with examples. Size of marks interacts with their placing. Clearly, large marks cannot be placed very close together without overlapping; so optimal spacing will depend on the size of the mark. Mark shape interacts with mark direction because for a circular mark shape, direction does not matter, but for a line direction might matter quite a bit. Also, mark size and shape interact. Small details in a mark's shape can be lost if the mark is drawn very small. Finally, color interacts with several variables: size, shape and placement. The overall size of a mark determines how strongly its color is perceived [Stone 2003]. Similarly, thin lines may not appear to have the same color as round blobs of the same area, even if the physical color is identical. Lastly, placement of the marks affects the apparent color because of simultaneous color contrast [Albers 1975].

### 1.4 Methodology

The ultimate goal of this line of research is to develop theory about texturing of layered surfaces. To accomplish this goal we cycle through a three-step process repeatedly. The three steps shown in Figure 5 are design of experiments, analysis of results, and interpretation as theory. Each step must rely on the results of the previous step to proceed. The studies reported here go through this cycle a number of times in the process of gradually building up knowledge and theory behind layered surface texturing.

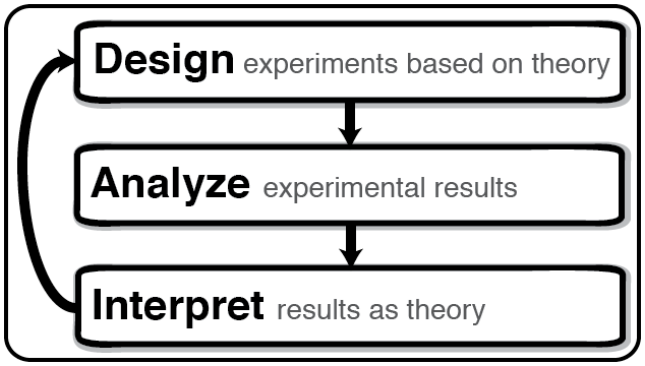



*Figure 5. Methodology. Diagram of the steps used in this research of layered surface texturing.*

Standard experimental design uses controlled experiments in which multiple measures are made for each possible level of one or more variable. For example, a variable might be ‘wood type’, levels might be ‘Oak, Ash, Spruce and Pine’, and the measurement might be strength. An experimenter might make strength measurements on several pieces of each type of wood, and then run statistical tests to see if the wood types had different strengths. This is easily done, but gets more difficult if more variables are included. The experimenter might also want to include information about the age of the tree, size and shape of the block of the wood, signs of rot in the wood, direction of the cut relative to the grain, and any treatments applied to the wood. The number of measurements required to test all combinations of all variables quickly becomes intractable. Running separate experiments for each variable will only work if the variables do not interact, and common sense dictates that unfortunately many variables used in layered surface texturing do interact.

Because little perceptual theory exists on the stylistic aspects of surface texturing, and a stylistic texturing parameter space is very large, initial experiments were designed with a novel methodology meant to broadly search a large parameter space. This search allows the experimenters to sparsely search the parameter space leaving the burden of finding structure in the space to the data analysis. Because the parameter space search is so sparse, these experiments would not provide reliably significant results, but

can be used to guide future controlled experiments by identifying relative importance of parameters and interactions between parameters. This is important so that time is not wasted running experiments on parameters that do not matter, or finding results that are only valid if other parameters are set to specific values. These multi-parameter experiments were designed to search an extremely large parameter space using user-guided heuristics. Although statistical significance is difficult to find with so many parameters, a variety of data mining techniques can be used to triangulate the results and gain insight into the parameter space. This multi-parameter methodology was introduced by House and Ware [2002]. Once the large parameter space has been analyzed and several hypotheses made, controlled single parameter experiments can be used to test these hypotheses. These studies varied only a few parameters, leaving others constant according to guidelines learned from theory and the previous multi-parameter experiments. Finally, results from these two-surface experiments are then applied to more than two surfaces, and analyzed to see if the two-surface methods are scalable.

The thesis is structured as follows. The background in Section 2 explores previous work in single and layered surface rendering. Work is looked at from the fields of art and perception, as well as controlled experiments related to surface rendering. Sections 3-8 describe the experiments that were run, the analysis and results. Section 3 describes a broad multi-parameter study designed to search the space of possible textures using 122 parameters to create textures. Section 4 is a similar, but slightly more focused study with only 26 parameters. Sections 5-6 document experiments varying small numbers of parameters, including size, opacity, and structure. Section 7 is an experiment designed to test the feasibility of scaling up the number of layers. Section 8 explores texturing of single surfaces to find correlations between texture, surface shape and viewing direction. Finally, the discussion and conclusions synthesize experiments and previous work into a set of guidelines with a theoretical model to motivate the guidelines.

# 2. BACKGROUND

## 2.1 Introduction

Knowledge on optimal rendering styles for single and layered surfaces can come from several sources. These include visualization experts, physiological models of the human visual system, and data from experiments. Experts bring years of practical experience working on specific problems, and have found optimal methods through trial and error, and learning from other experts. Artists are the experts of choice in the case of surface rendering. Artists have a wide variety of rendering techniques that have stood the test of time, and can be trusted. However, these techniques often depend on the purpose of the artist, and the tools used to create the rendering. They also offer little in the way of theoretical understanding of why some rendering solutions work well and other do not.

Models of the human visual system allow more general inferences to be made about the sorts of rendering styles that work best. These models are based on both human anatomy and human cognitive and perceptual studies. Much research has been done on human visual processing that can be applied to the topic of layered surface rendering. Unfortunately, these models are generally not detailed enough to allow accurate predictions for specific rendering styles, but can be used as a basis for further exploration.

Beyond expertise and theoretical models, experimental data is required to assess the quality of different rendering techniques. Experiments can take two directions. Controlled experiments generally give users a task to perform, and evaluate how performance on the task is affected by certain variables. Tasks are chosen to be as general as possible, so that results will hopefully be valid in other situations. Obviously, no one task can account for all the complexities of many different real-world tasks, but controlled experiments can be applied at some level to similar tasks. On the other hand, user studies try to determine optimal techniques within the framework of a particular application. These studies are particularly important when developing real-life applications because they can better account for the complex interactions that take place

when a human uses a complex system to solve non-trivial problems. Users are typically given more freedom to interact with the system, just as a real user would be able to do. Evaluation may not be limited to rigid definitions such as task-based accuracy or time, but might include user comments. These studies of surface rendering are aimed at developing general principles and guidelines, not for a specific application. Therefore, the focus is on controlled experiments rather than user studies. The following three sections present an overview of surface rendering techniques based on artistic methods, psychological models, and experimental data.

### 2.2 Artistic Surface Rendering

Artists have developed many techniques for showing surface shape. Here we discuss the techniques developed for use with pen and ink. For our purposes, ink drawings can be thought of as the simplest types of drawings because they only allow for two possibilities at any point on the paper: white or black. On the other hand, pen and ink drawings have a huge range of possibilities in rendering techniques. The fundamental unit of an ink drawing is a 'mark'. Aside from stippling techniques, these marks are some sort of line or curve. It is the artist's job to choose the line shapes and placements to achieve a desired effect. Artists typically draw lines for only a few purposes. These are to show silhouettes, ridges or creases, shading, texture, surface direction and curvature. Silhouettes are generally drawn with single, hard lines that represent the boundary of a surface and the background. Silhouettes may include edges within the figure, such as the profile of the nose against the cheek in a profile view of a head. Sharp ridges or creases are also generally drawn as hard lines. Figure 6 shows two examples of a drawing of a hand. The first, a) shows a drawing where lines are used only to show silhouettes and creases. The second b) shows the same hand drawn using shading. Instead of the emphasis on edges that the first drawing uses, the shading drawing seeks to show the volume of the hand with the varying shades of gray. Areas of the hand that are facing the light are brighter, and areas that receive little light are darker.

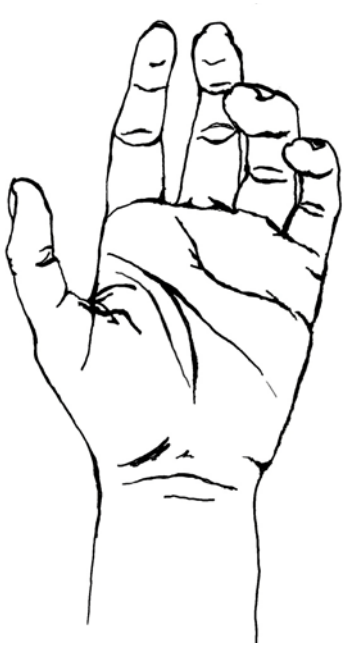
a)  b) 

*Figure 6. Silhouette and. Shading Artistic Examples. a) Hand drawn in ink by the author. b) Hand drawn in charcoal by the author.*

Artists generally seek to show the texture of the object they are drawing. This can be done in several ways, most often through the paper texture or through use of marks to represent the texture. Figure 7 a) shows how the same charcoal can look significantly different on different papers. An example of using marks to represent fur texture is shown in b), where the direction of the pencil strokes is aligned with the direction of the hairs in the coat of the leopards.

a)

b)

*Figure 7. Artistic Texture Examples. a) Charcoal on different paper textures. b) Fur texture drawn in pencil by the author.*

Cross-hatching is a technique that layers a second or more sets of lines at a different angle from the first, to darken an area further. The direction of the lines might relate to the direction of the surface, the direction of the light, or be chosen arbitrarily [Johnson 1982]. Crosshatched lines often have relative angles close to 90 or 45 degrees. Hatched lines are also used to show surface direction and curvature. Unlike shading hatching lines, these lines curve with the surface [Chapman 2006; Lohan 2004]. If the surface is singly curved, such as a cylinder, only one set of hatched lines is typically used, and they are aligned with the direction of maximum curvature. If the surface is doubly curved, two sets of lines may be used, but it is not clear if the directions chosen are always the directions of maximum and minimum curvature. By definition, the maximum and minimum curvatures of a surface are perpendicular, but artists do not always draw the curved lines with 90-degree relative angles. Figure 8 a) shows hatched lines used to show shadow and light direction. b) shows an example of using cross-hatching in the direction of surface curvature.

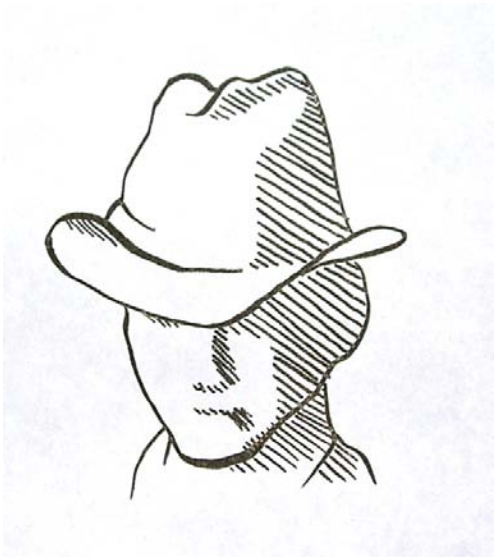
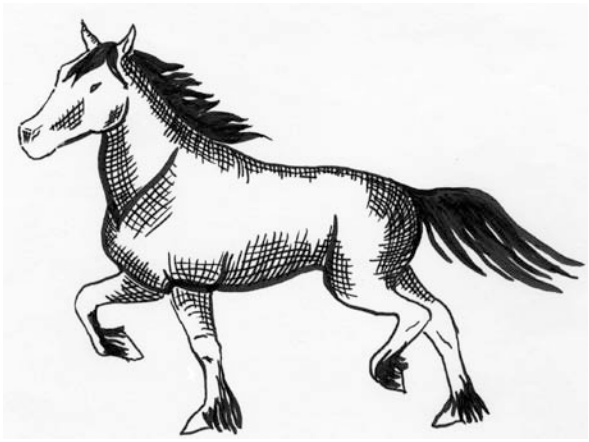
a)  b) 

*Figure 8. Artistic Lines for Lighting and Curvature. a) A drawing that uses lines to show shadow and light direction by the author. b) A drawing that uses cross-hatched strokes to show the curves of a horse by the author.*

The artistic techniques of contour lines, ridge and crease lines, textured surface lines and hatching to show both shading and curvature are worth considering for layered surface renderings. Algorithms for automating some of these techniques already exist, but most have room for improvement, and some are still unsolved problems.

### 2.3 Perceptual Surface Cues

Now we look at perceptual models of the human visual system. The anatomy of the human visual system is quite complex. Light that enters the pupil is focused by the lens and the cornea onto the back of the eye (retina), which contains photoreceptors sensitive to different intensities and frequencies of light. The different kinds of photoreceptors allow humans to see in color, as well as in drastically different illumination levels on the order of a $10^8$ difference in brightness. The receptors also vary in density, giving very fine-scale vision in the center of the field of view (fovea), and coarse-scale vision on the periphery. Finally, layers of neurons perform filtering on the visual signal within the eye itself, before the signal is transmitted along the optic nerve

to the brain. These filters have been found to act as simple parallel image processing operations. One type of filter that exists is called the center-surround. These filters work by combining the weighted activation of many photoreceptors. They are tuned to output a signal only if the central receptors are receiving lots of light, and the surround receptors are receiving little light. If either all receptors get a lot of light, or little light, the positive and negative weights cancel, and the ganglion cell does not fire. Off-center cells take into account the opposite case of a dark object on a light background. Clearly, brightness contrast is an important cue to the human visual system. Further along in the visual system, filters have been found that are attuned to specific directions. Features that the visual system is sensitive to through 'built-in' mechanisms are called perceptual cues. The main cues used by the visual system for shape perception are shading, reflection, shadows, occlusion, occluding contours, stereo, motion, lines, and texture. The cues that are relevant to this particular research are described in detail in the remainder of this section.

Shading is the color seen at any point on an object, and is based on the material properties of the object, the lighting environment, and the position of the viewer. The Bidirectional-Reflectance-Distribution-Function (BRDF) is used to describe material properties. The simplest BRDF model is Lambertian, which assumes equal reflectance in all directions. This model is well suited to simulate the shading of dull surfaces such as clay. The human visual system interprets shading as 3D shape whenever possible, as demonstrated by the apparent concave and convex shapes of the shaded circles in Figure 9. Most people will see the left circle as convex and the right as concave. This is because humans tend to assume a single lighting scheme for all objects in the scene, and are biased to assume a light direction from above. However, since exact light location and material properties are generally not known, shape-from-shading is an under-constrained problem. As a result, other shape cues tend to augment or override shading information. This will be discussed in more depth later [Zhang et al. 1999; Ramachandran 1988; Nefs et al. 2006].

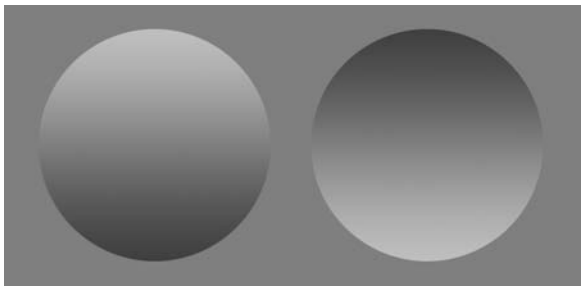


*Figure 9. Shape from Shading. Shape from shading demonstrating convex and concave bumps.*

Boundary contours are strong shape cues because along the contour the surface normal must be in the viewing plane, and perpendicular to the contour. This defines surface normals all along the contour, which can be interpolated using other shape cues [Barrow and Tenenbaum 1981]. Boundary contours are powerful enough that they tend to override shading cues, as in Figure 10. Both surfaces have the same shading information, as can be seen by covering the top half of both images. However, when the contours are visible, not only do the perceived surface shapes change from cylindrical to wavy, but also the lighting direction seems to move from directly above to the side.

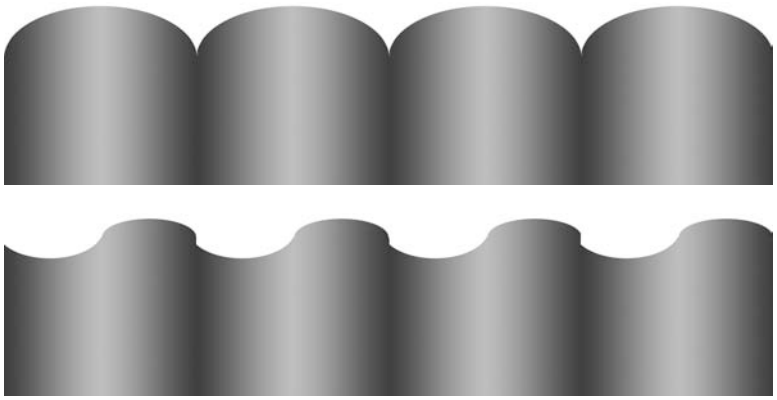


*Figure 10. Shape from Contours. Contours that override shape cues.*

Humans have two eyes, separated by a small distance, allowing us to use stereo cues to depth. Stereovision requires that corresponding points be identified in the right and left images. This is a difficult problem for computers, but humans do it almost effortlessly [Poggio 1984]. Once corresponding points have been found, humans can use the disparity between the points to judge distance. Disparities will be relative to whatever point the eyes are converged on, with closer objects having negative disparities and more distant objects having positive disparities. Stereo acuity from disparity is about 10 arc seconds, which is a little under 1mm of depth difference at 1m. Disparity is a strong cue up to about 1 meter, and is relevant up to about 30 meters, where depth differences have to get larger and larger to maintain the same angle difference. At thirty meters, a difference of about 70cm is required for a 10 arc second acuity to distinguish distances. Disparity is a cue that does not require movement of the eyes, however as the eyes move about a scene, some depth information can also be gotten from eye convergence (the angle that the eyes have to turn toward each other to focus on an object). Points that require more convergence are closer to the eye than points requiring less convergence. Comparison of distances at different points on an object and between objects can be a powerful cue to shape. A cross-eyed stereo pair of images is shown in Figure 11. These images were photographed simultaneously from different camera positions so that crossing the eyes to make the two images overlap causes the jets of water to be perceived in front of the bushes and steps in the background. A final depth cue is lens focus: physical bending of the lens to increase or decrease optical power of the eye. Unfortunately, focal depth that changes depending on where the viewer is looking is difficult to simulate in a computer environment because it requires eye-tracking and real-time depth of focus calculations [Ware and Frank 1996].

*Figure 11. Shape from Stereo. Cross-eyed stereo pair. Used with permission, [Grey 2007].*

Motion parallax can also be a powerful indicator of depth. It works similarly to stereo in that corresponding points are mapped between at least two images. Then the optic flow field can be used to estimate surface structure. Human abilities to correctly estimate surface shape and motion from only a small number of points and images suggest that humans employ various assumptions such as rigidity of objects to constrain the set of possible shapes and motions [Hoffman 1983]. Computational algorithms for obtaining structure from motion have been presented. [Koenderink and van Dorn 1991; Zhang et al. 2003]

Lines can represent silhouette edges, surface boundaries, creases, reflection and shadow edges. Interpreting silhouette edges is the same as occluding contours in that the normal is well defined and can be used to interpolate throughout the surface. When interpreting the surface from 2D lines, humans constrain the surface to fit with the local line characteristics, and global constraints on curvature. They also use global characteristics such as symmetry and parallelism and context to construct 3D interpretations of lines. Figure 12 shows how very similar sets of lines can be interpreted as boundaries in the top case, and silhouettes in the bottom case [Barrow and Tenenbaum 1981].

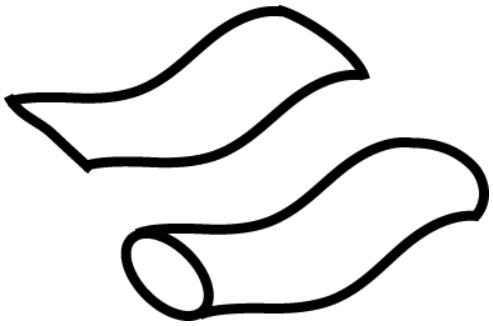


*Figure 12. Shape from Silhouette and Boundary Lines. Surface shape can be interpreted from boundary lines and silhouette lines. Context determines whether lines are interpreted as boundary or silhouette.*

Silhouettes alone can be a strong shape cue. Figure 13 shows a car rendered only using silhouette lines, where silhouettes are defined by surface normals within the image plane. The drawing is quite successful at showing the vehicle's shape without any other perceptual cues.

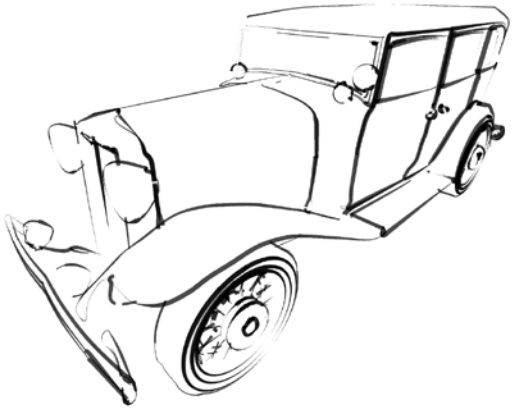


*Figure 13. Silhouette Algorithm Example. A '32 Dodge car rendered showing how silhouette edges can show surface shape. Used with permission, [Singh 2004].*

Homogeneous (translationally invariant) textures give several shape cues; these include compression, size gradient, density gradient, compression gradient, and perspective convergence [Saunders and Backus 2006]. Examples of these cues are

shown in Figure 14. Mark size and size gradient can show relative depth when using perspective projection to estimate the distance to the surface. Density and density gradient can also be a clue to distance when using perspective projection. If the density is only changing in one image direction, then this is a clue that the surface is curving. Perspective convergence is a strong depth cue when perspective projection is used [Saunders and Backus 2006]. Mark compression is a cue to the tilt of the surface. Finally, mark compression gradient is the dominant cue to changing direction of a surface [Saunders and Knill 2001].

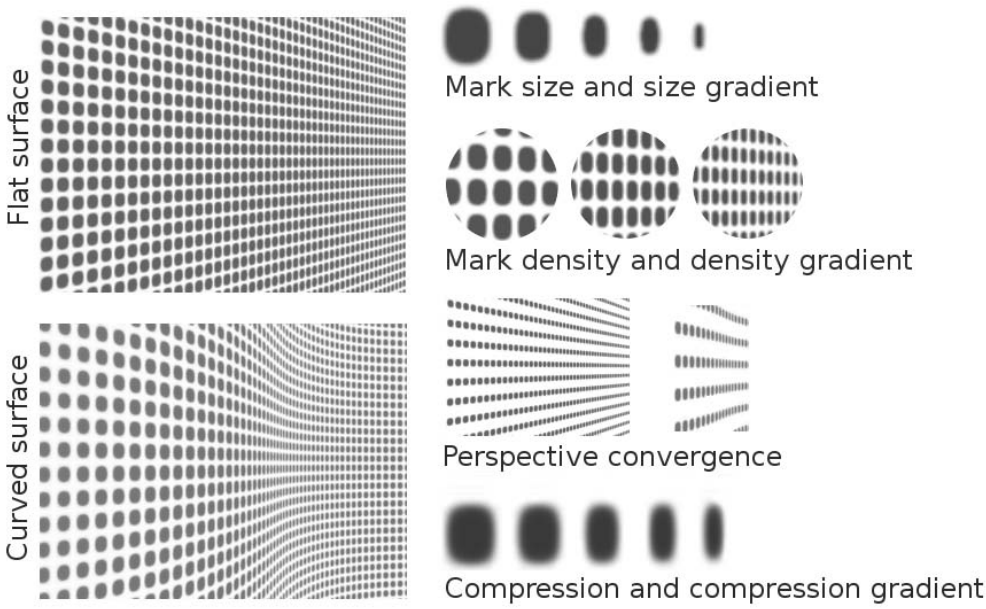



*Figure 14. Shape from Texture. Texture shape cues that can be derived from any homogeneous texture.*

Textures can be homogeneous, isotropic, both or neither. Homogeneity refers to the property of being independent of translation. That is, wherever on a surface that the texture is sampled it has the same characteristics, such as size and frequency. Figure 15 shows four combinations of homogeneous and isotropic textures. The isotropic texture in a) is the same in all directions, while the anisotropic texture in b) has elongated lines in the horizontal and vertical directions. The inhomogeneous textures c) and d) have larger

marks toward the center of the texture, and smaller ones toward the edges. Because this texture inhomogeneity is very similar to texture gradient cues for homogeneous textures, these textures may appear to bulge outward. For this reason, using inhomogeneous textures can be dangerous since they may conflict with other shape cues.

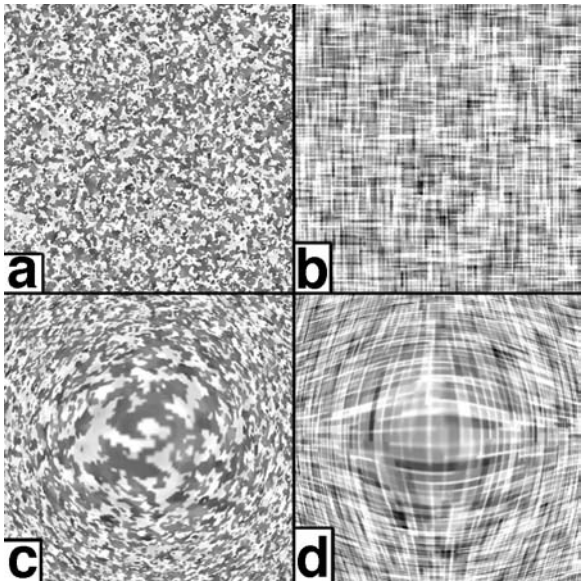



*Figure 15. Texture Isotropy and Homogeneity. a) shows an isotropic, homogeneous texture. b) shows an anisotropic, homogeneous texture, c) shows and isotropic, inhomogeneous texture, and d) shows an anisotropic, inhomogeneous texture.*

Anisotropic textures such as lines or grids give strong shape cues aside from compression and gradients. Contour lines can be interpreted with a parallel projection assumption to improve the interpolation of surface normals from occluding contours and surface boundaries [Stevens 1981]. An example of contour lines is shown in Figure 16. These contour lines are in fact inhomogeneous as well as anisotropic. Because the lines are parallel surface contours, they are more widely spaced on slanted areas of the surface then on level areas. Yet this and other inhomogeneous textures seem to work well even though many texture cues cannot be guaranteed [Todd and Reichel 1990]. One

explanation for why projected textures might work is that they are identical to patterns made from carving an object with a grain such as wood, or stretching something with a pattern, such as spandex cloth. Therefore, projected textures have a geometrical, logical basis, and exist in real life to provide our visual system with examples.

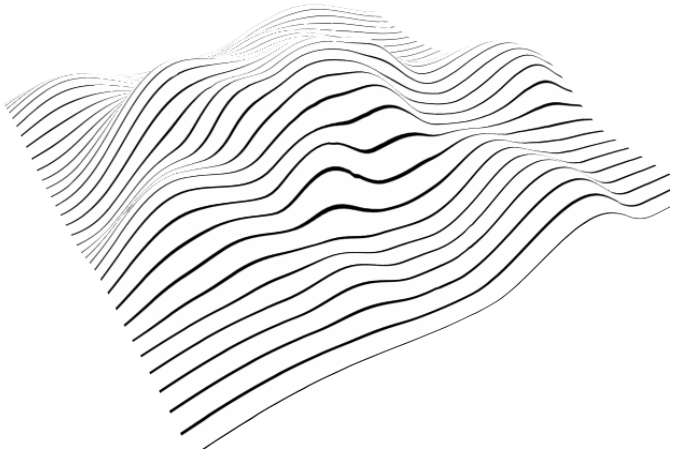


*Figure 16. Contour Lines. Parallel projected contour lines show the surface shape.*

**2.4 Surface Visualization Studies**

This section provides an overview of the various studies on surface visualization. It explains the tasks used to measure surface perception, the perceptual cues included in each study, and findings for optimal parameter values to take advantage of those perceptual cues.

Finding a task to accurately measure the accuracy of surface is not immediately obvious. The most common methods used are as follows. Estimating the curvature of cylindrical subsections [Todd and Mingolla 1983], picking the correct local shape (ellipsoids, cylindrical, saddle, flat) of quadratic surfaces [De Vries et al. 1993; Kim et al. 2004] require choosing from a finite set of possibilities for a restricted set of surfaces. Other researchers used accuracy of ‘slant’ and ‘tilt’ to measure comprehension of surface shape [Mamassan and Kersten 1996; Koenderink et al. 1992]. Slant is defined as

the angle between the surface normal and the viewing vector, while tilt is the angle of the surface normal projected into the image plane. This allows for arbitrary surfaces, but requires multiple measurements across a surface to gauge overall surface comprehension. Mamassan found that observers tend to underestimate surface slant, but have larger variance on the surface tilt. The bias was less strong on surfaces that were locally egg-shaped than for surfaces that are cylindrical or saddle-shaped. Recent work has simply used the surface normal angular error, because effect of tilt error depends on the slant angle [Interrante and Kim 2001; Sweet and Ware 2004]. Another possibility to measure surface perception is a feature finding task, such as identifying the location in which two similar surfaces are different [Interrante et al. 2002].

Sweet and Ware found that an oblique viewing angle had the least error for projected textures on Gabor surfaces, and a 45-degree angle minimized error bias [Sweet and Ware 2004]. We believe this might be because an oblique angle gives the viewer silhouette contour information as well as texture and shading information. Also, field of view of the camera has been shown to affect the perceived slant of surfaces [Todd et al. 2004]. A lighting position within a range of 45 degrees from overhead was found to give the most accurate results for curvature estimation of spheres [Curran and Johnston 1996].

In most applications, multiple visual cues are used. Therefore, it becomes important to know the relative importance of these cues when used in combination, and how any conflicts are resolved. Combinations of many visual cues have been shown to improve perceptual accuracy. Improvements ability to read large connected graphs have been found for the combination of stereo and motion over using either cue alone [Ware and Frank 1996]. This argues strongly for using these 3D cues together when possible. Also, texture combined with shading has been shown in many studies to be better than only shading for surface perception [Cummin et al. 1998; Sweet and Ware 2004; Kim et al. 2004; Todd and Mingolla 1983]. However, care should be taken that visual cues do not conflict, as the visual system must somehow resolve those conflicts, perhaps in unexpected ways. A previously seen light direction was shown to override cast shadow

cues in scene interpretation [Berhaum et al. 1983]. Shading information is also shown to override cast shadow information in the Mach card illusion where convexity is ambiguous [Mamassian et al. 1998]. On the other hand, lighting was found to have no overall effect on categorical shape judgment of superquadrics. In the same study shading was found to have a comparably small effect as well, suggesting that the silhouette contours of the superquadrics dominated the shape judgment. Results on the usefulness of specular highlight cues have been ambiguous; no effect was found for a recent study [Nefs et al. 2006]. However, specularity does seem to be useful when used with stereo disparity cues [Todd et al. 1997].

### 2.5 Surface Texture Studies

Now we look specifically at variation in the texture cues to surface perception. Motivation for using texture can be found in numerous studies that have found that texturing and shading a surface is better than simply shading it for a variety of tasks. [Todd and Mingolla 1983; Cummin et al. 1998; Kim et al. 2004; Sweet and Ware 2004]. Figure 17 shows a comparison of shading and shading combined with texturing on a cylindrical surface.

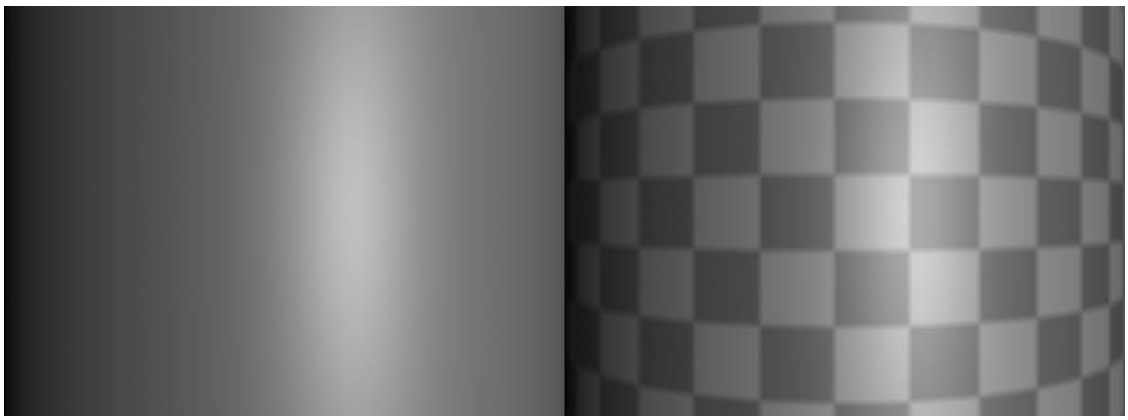


*Figure 17. Shaded and Textured Cylinders. A shaded cylinder and a textured and shaded cylinder. Texture cues give clearer information about surface curvature than simple shading.*

Anisotropy and inhomogeneity in textures has been thought to hurt performance because it reduces the ability to compute simple texture cues such as compression and gradients in the texture. Projected textures have an inhomogeneity that tends to exaggerate compression and gradient cues when viewed from an oblique angle, and minimize them when viewed from the direction of projection. For this reason projected textures do not work well when viewed from the direction of projection. However, anisotropic textures like grids provide other strong cues like surface contours. Carefully chosen anisotropies seem to be helpful. Figure 18 shows examples from a study comparing various choices for anisotropic textures compared to an isotropic texture. [Interrante and Kim 2001]

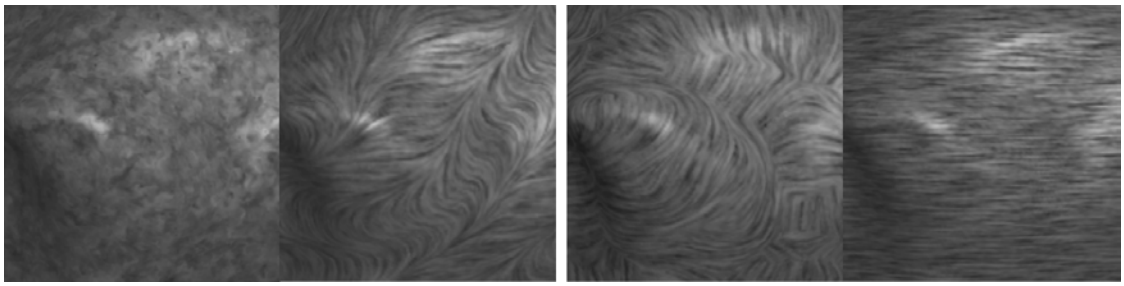


*Figure 18. Texture Direction. The left two textures (isotropic, and a random-direction anisotropic) performed poorly compared to the right two (anisotropic in the direction of principal curvature and unidirectional). Reprinted with permission from Interrante and Kim [2001].*

The preferred number of directions in an anisotropic texture seems to be two, with the exception of cylindrical surfaces, which seem to do better with only one direction [Kim et al. 2003; Sweet and Ware 2004]. Although no studies have used non-orthogonal directions for the 2-direction case, two orthogonal directions allow assumptions to be made about the surface normal at mark intersections. Figure 19 shows a comparison of 1 and 2 directional textures [Sweet and Ware 2004].

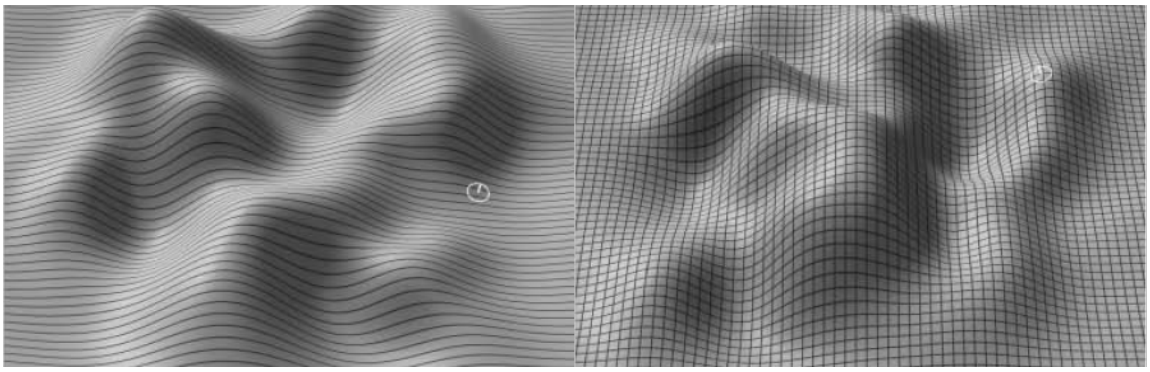


*Figure 19. Texture Lines and Grids. A grid of horizontal and vertical lines works better than either horizontal or vertical lines alone. Reprinted with permission from Sweet and Ware [2004].*

Preferred direction heuristics for anisotropic textures have been researched. For developable surfaces (surfaces with only one direction of curvature), several studies have shown that textures that align with the direction of curvature are better [Li and Zaidi 2001; Knill 2001]. Figure 20 shows an example of two line directions, in which the maximum curvature direction much more clearly shows the surface shape.

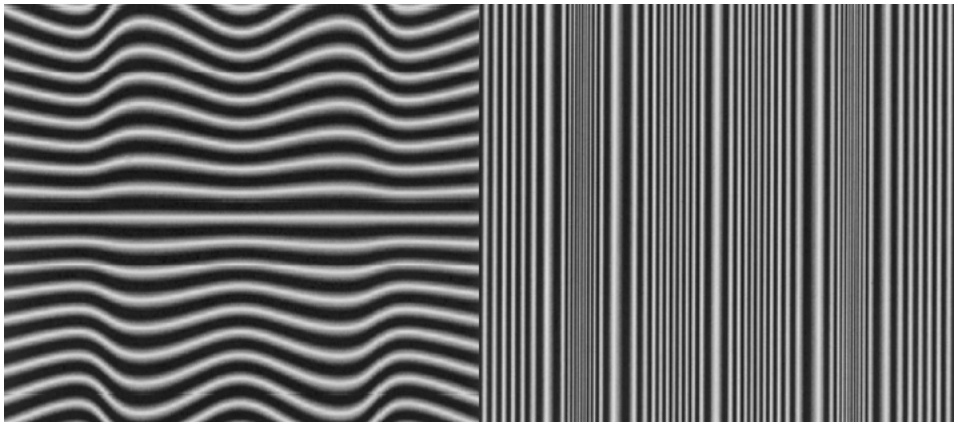


*Figure 20. Texture Principal Directions. Texture lines in the direction of maximum curvature and minimum (zero) curvature for a sinusoidal grating. The leftmost works far better.*

However, for the case of doubly curved surfaces (most everyday surfaces are double curved), even generating a texture that aligns with the direction of curvature at all points is difficult. Methods used have included placing ellipsoid marks on the surface [Interrante et al. 1997], and line integral convolution [Interrante and Kim 2001]. Recently texture generation by seaming together nearly planar surface patches [Gorla et al. 2003].

Interrante and Kim [2001] performed studies testing optimal directions for anisotropic textures. Textures in the direction of principal curvature were shown to be on average better than isotropic or anisotropic random textures. The study conditions included both monocular and stereo viewing, specular and diffuse shading, with no visible silhouette contours. However, the study showed no decisive difference between principal direction textures and constant direction textures. Errors in the case of principal direction textures tended to happen at texture discontinuities. In a follow-up study [Interrante et al. 2002], principal direction textures performed better than constant direction textures for a feature-finding task under monocular, static viewing, with diffuse shading, random directional lighting and both oblique and top viewing directions.

The previous research has all been done on single surfaces. Considerable research has been done on multi-parameter visualization [Acevedo and Laidlaw 2006; Hagh-Shenas et al. 2006; Urness et al. 2005; Kirby et al. 2004; Taylor 2002; Laidlaw et al. 2001]. However, almost no work has analyzed multi-surface visualization, with the notable exception of Interrante's work showing that opaque texture marks significantly improved distance judgments between two surfaces [Interrante et al. 1997]. Results for surface shape, viewing angle, shading, and silhouette cues probably apply to two surfaces as easily as one. Even so, many aspects of texturing on 2 surfaces remain to be studied, and no studies have evaluated texturing of more than 2 surfaces. Therefore this work seeks to analyze stylistic aspects of two-surface texturing, and see if these heuristics can successfully apply to more than two surfaces.

## 2.6 Alternative Layered Surface Rendering Methods

The previous sections on artistic techniques, perception and experimental analysis of surface rendering, do not refer much to the problem of layering surfaces. Significant research has been done looking at various ways of layering information, but with the exception of Interrante's work no studies were found that dealt with the layering of surfaces. Transparent layered surfaces are also not commonly used in scientific visualization at this time. This is likely the combination of two factors. First, texturing is not commonly used in scientific surface visualization. This is probably for many reasons, among them that textures were computationally expensive on older machines, fear of adding 'frivolous' patterns to scientific images, and a lack of algorithms for texturing general surfaces [Gorla et al. 2003]. The second factor for the rarity of layered surface visualization is the observation that layering of untextured surfaces simply does not work very well. Although a single surface without texturing can usually be understood fairly well, layering those surfaces causes far more difficulties. When the only cue to surface shape is shading, layering of surfaces makes it difficult to determine which surface the shading patterns belong to. Since most software designers or scientists likely did not consider using texture to improve the layering, the base case layered visualization would tend to fall flat. In reaction to the poor performance of basic layered surface views, a number of other techniques were developed for layered rendering. Among these are cutaways, cross sections, and exploded views.

Cutaway views simply remove the top surface where it is important to see the bottom surface. Various types of transitions are used between the part of the top surface that is transparent and the part that is opaque. Figure 21 shows a sharp transition between a transparent and opaque top layer. Naturally, the cutaway method does not work well if the user loses valuable information about the top layer in the area of the cutaway. In this example, the cutaway works well because the gray outer layer serves basically as context. Also, the black edge lines allow the viewer to interpolate and guess the shape of the outer surface well. However this will not be the case for all surfaces. A cutaway generally only works well if the outer surface serves mainly for context, and has

a simple shape. For outer surfaces with complicated shapes that are important to the viewer, a cutaway is not an ideal solution.

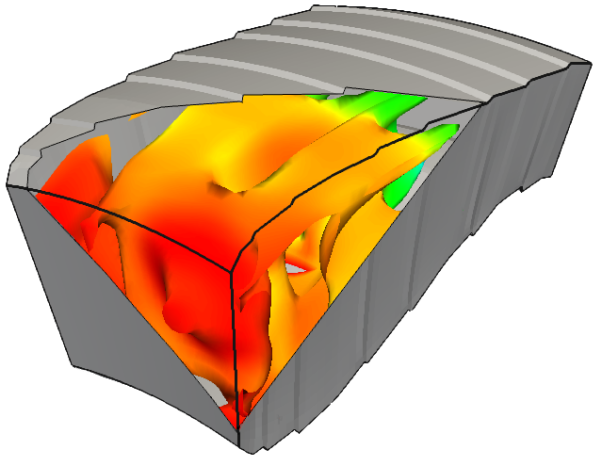


*Figure 21. Cutaway Example. A cutaway shown on the ParaView comb dataset [Squillacote 2006]. The inside has been contoured on density and colored by temperature. A cutaway has been used to make the inside data visible in the context of the outer dataset shape.*

Cross sections separate data into planar slices, giving clear, detailed images for each slice. Slices can then be compared to see how data varies in three dimensions. One typical use of cross sections is architectural plan views of a building. One view might show a single floor of a building as viewed from above, and another might show the height of the building from the side or front. Figure 22 shows cross sections used on flow data through a contained area. Each cross section is color-coded, in this case according to density. A pattern can be seen in the data, and the viewer can interpolate across the sections to estimate a shape at various levels of density. The drawback to using cross sections is that mentally interpolating these slices is somewhat difficult and prone to error. Getting fine-grained surface information requires a lot of slices, which tend to occlude each other, making seeing the data as a whole more difficult.

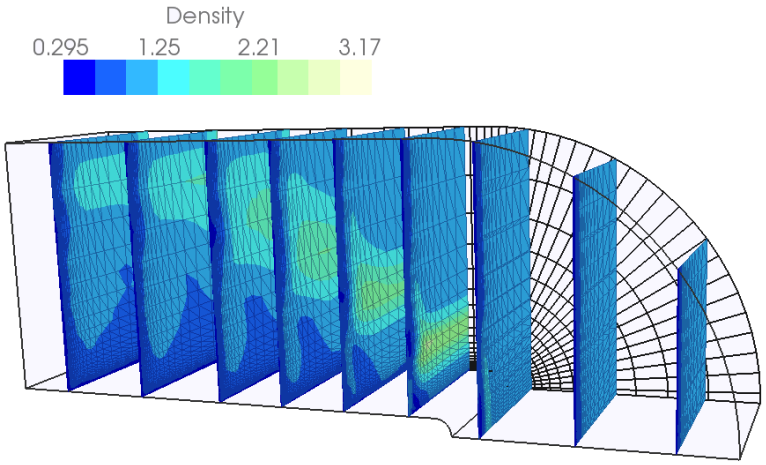



*Figure 22. Cross Section Example. Cutting planes of the Paraview Bluntfin dataset, [Squillacote 2006], show variations of density within the dataset.*

The exploded views technique shows all layers in three dimensions and fully opaque. It deals with occlusion by moving the pieces in such a way that all pieces are visible, and it is as clear as possible how the pieces would fit back together. Figure 23 shows an example of this method for sediment layer visualization. The sediment layers in are exploded vertically and to the right. The visualization clearly shows the 3D shape of all of the important components as well as the global structure of the assembly. Other methods for exploded views include radial movement of the pieces, or techniques that use deformation as well as transformation, such as 'peeling' back of parts. The exploded view can work quite well in an explanatory capacity. However, the method is a trade-off between maintaining reasonable distance relationships and minimizing layer occlusion; and one or the other must usually be sacrificed. In this exploded view, several of the layers are occluded, which is not a problem for showing the overall shape and layering of the sediment layers. However, for some data-analysis tasks both distance relationships and minimal occlusion are important, so exploded views are simply not an option.

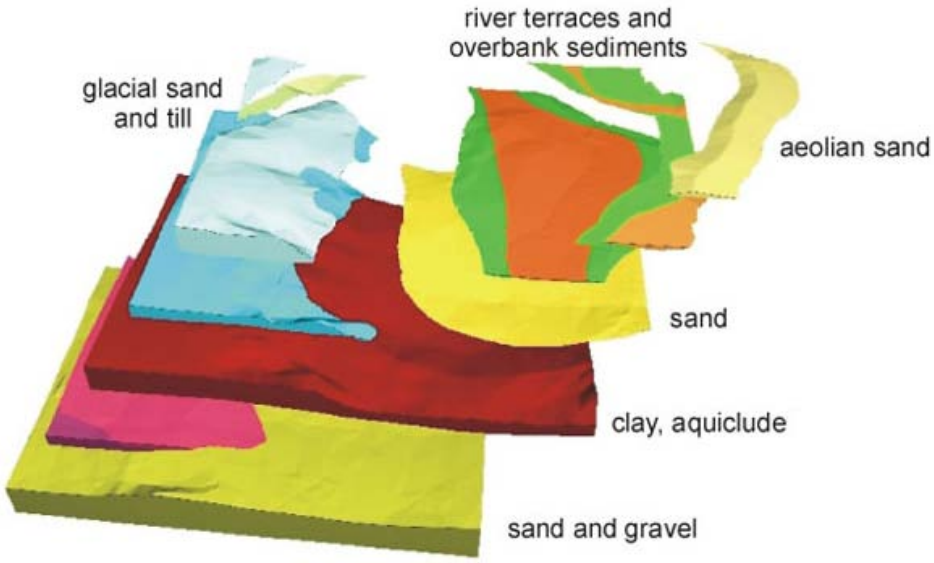



*Figure 23. Exploded View Example. Exploded view of sediment layers. Reprinted with permission, [Neber 2007].*

Cutaway, cross-section and exploded views are all extremely useful in different circumstances, but none solve the problem of simultaneously showing more than one surface without losing information or distorting spatial relationships. This brings the problem back to rendering layered surfaces.

## 2.7 Additional Visualization Techniques

Two other visualization techniques are discussed: importance-driven visualization and interaction. These techniques are not exclusive to the layered rendering method, but rather can serve as supplements to aid perception and usability of the visualization.

Importance driven visualization is similar to the cutaway technique in that the top surface is drawn transparent in areas where the bottom surface has important features, and opaque elsewhere. Hand-drawn technical illustrations are often excellent examples of importance-driven illustrations. For example, consider the task of a technical illustrator trying to show how a car's engine and interior components fit within the outer

car body. Important parts of the body such as the outside edges could be shown opaque, and in areas where the interior or undercarriage are important the outer body could be left transparent. Adding opaque edges and highlights can show the overall body shape clearly even when the rest of the body is drawn nearly transparent. Therefore, the most information about the layer is drawn with the least coverage on the paper (or screen).

Work has been done to automate importance-driven visualization [Viola et al. 2004], and other techniques used by artists to show significant features like silhouettes and edges [dos Ries Rivotti et al. 2007; Costa Sousa and Prusinkiewicz 2003], and shading [Gooch et al. 1998]. Importance driven rendering works extremely well for cases such as a car, in which the outer surface has relatively few important details. However the case remains to be solved of how to visualize multiple surfaces that are both detailed and important. How to optimally display the necessary surface features of two equally important surfaces is not well understood, and is the topic of this research.

Interaction gives a wide range of possible improvements to the comprehensibility of visual data. Aside from interacting with viewpoint, lighting, color and so on, software can easily allow a user to turn surface layers on and off, as well as having adjustable opacity and clipping planes. While this at first glance seems to solve the problem, it is simply another tool. Using clipping planes to cut off one surface is only useful if the outer surface only serves as context, and is not important to the analysis. Switching between surfaces can be helpful, but for each visual comparison made between surfaces the user must manually switch back and forth; this can prove quite irritating. Using a variable opacity again begs the question of how best to display a semi-transparent surface overlaying another surface. Interaction should be used as a supplement to good rendering techniques for layered surfaces.

# 3. MULTI-PARAMETER PILOT EXPERIMENT *

## 3.1 Experimental Design

Initially, little was known about layered surface texturing beyond extrapolations made from single surface texturing research and common sense. Certainly it seemed clear that using texture was a good idea, and that having some directionality to the texture was preferable, but many variations were completely unexplored. Therefore, the first experiment included as many stylistic variations as possible to try to get a broad outlook at the texture space. Non-texture shape cues, specifically stereo, shading and motion, were included at the discretion of the researchers with the main goal being to make the shape as clear as possible.

The experiment used a human-in-the-loop methodology first presented by House and Ware [2002]. A diagram of this methodology is shown in Figure 24. The methodology starts by defining a visualization problem and method. The specifics used for this experiment are described in detail below. The parameterization is the choice of texture parameters based on the problem and visualization method. A search of the parameter space must be conducted in which sets of texture parameters are rated. The rating should be done by a human being, either using some heuristic, or by measuring performance on a task. It is the human rating that gives the technique its name: human-in-the-loop. The parameter space search produces a database of rated textures, which can be data-mined for various purposes. Simply sorting the database by rating can produce example good solutions, and more complicated data analysis can generate more general design guidelines and perhaps theories.

---

* The images and text in this section are reproduced in part with permission from previously published material. See [Bair et al. 2005; House et al. 2005; House et al. 2006].

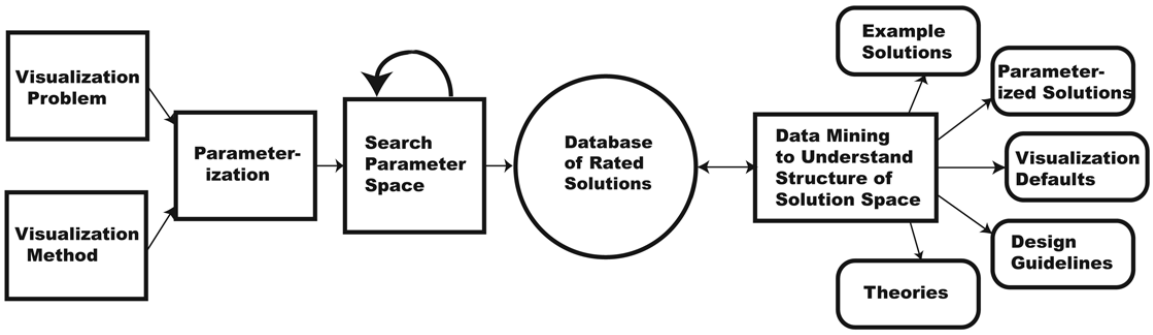



*Figure 24. Human in the Loop Diagram. Methodology from Visualization Problem through data mining and results. [House et al. 2006].*

In this experiment, the visualization problem is showing layered surface data. The visualization method is rendering a semi-transparent textured surface above another surface. Specifically, a display of two textured overlapping surfaces was presented in stereo (using a polarizing filter and glasses) and visually rocked about the screen central vertical axis to provide motion parallax. There was no attempt to represent focus cues, however texture parameters providing low-pass filtering on each layer were included, simulating one or both of the surfaces being out of focus. Both the stereo and rocking motion provided very strong depth cues. The two surfaces used were smoothly varying height fields, and the top surface had less detail than the bottom. The surfaces were shaded with a diffuse shading model and shadows were not included. Figure 25 shows the two surfaces that were used. They are presented side-by-side for clarity, rendered with shading but without any texture.

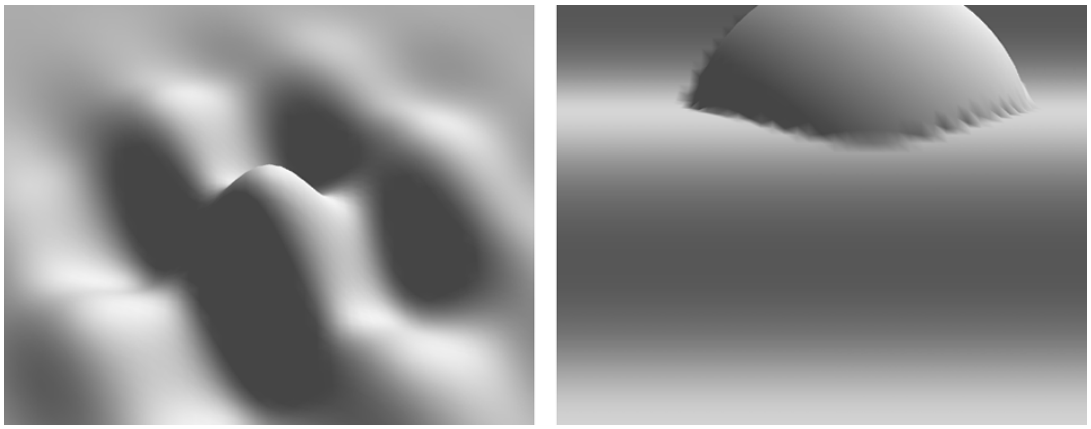


*Figure 25. Bottom and Top Surfaces. Surfaces were identical throughout the experiment.*

An oblique viewing angle of 45 degrees was used so that occlusion contours were available, and because studies have shown that it might be preferable over other viewing angles [Sweet and Ware 2004]. Projected textures were chosen for a number of reasons. Although they are inhomogeneous, they do allow for global features such as surface contours without any discontinuities, they are not surface dependent and so can be tiled and reused on other surfaces, and are far easier and faster to compute than surface-dependent or viewpoint-dependent textures.

The parameterization used 61 parameters, or 122 total parameters for a bottom-top pair of textures, in an attempt to cover the space of possible textures as completely as possible. Each texture was made up of four layers: a background and three layers of features composited over the background. Two of the feature layers were lines, and the third layer consisted of dots. Lines provided the ability to create crosshatching and linear structure, while dots provided the ability to create a high frequency, mottled look. The marks were arranged on the texture in a grid pattern to allow for global patterns such as contour lines and grids. The textures where mapped onto the two surfaces so that they made an eight-by-eight tiling on each surface. The parameterization includes overall surface parameters and parameters specific to each mark layer. Each surface had four overall parameters. These were the background HSV color, the opacity (though the bottom surface was always opaque), the overall rotation (which rotates all the feature

layers together to angles between -45° and 45°), and the low-pass Gaussian filter width (blur ranging from 0-8 pixels). Each mark layer consisted of marks drawn on a grid, with a single drawing color and opacity, and a variety of shape and drawing parameters. Some of the parameters used to vary each mark layer are shown in Figure 26. Shown in a) is an example set of lines on a 4x4 grid. Line shape parameters shown in b), can be varied to change the line aspect ratio and the line size within the grid cell. Since these parameters are measured relative to the grid cell, they must be compared with the grid cell size and shape to recover the actual mark size and aspect ratio. Each layer grid was defined by two parameters: number of rows and columns. These could vary independently to anything from one to thirty-two rows or columns. Uneven numbers of rows and columns can create large-scale ordering of the features, like the 20x4 grid shown in c). Note that vertical lines are perceived, even though the actual mark lines are horizontal. The marks are also given a rotational offset between -90° and 90°. An angle of 45° is shown in d). This allows the mark angle to be independent of the grid angle. Horizontal and vertical offsets (not pictured) are also included so that marks on different layers would not overlap by default. Marks are randomized in several ways: rotational jitter is shown in e), and translational jitter in f). Horizontal and vertical jitter are separate parameters. A probability parameter, shown in g) at 50% is the probability that a feature is drawn at each grid cell. This allows randomness from the grid formation. Finally, h) demonstrates blurring, which is controlled by a parameter that adjusts Gaussian low-pass filter width. Dots and lines use the same parameters, except that dots use the width parameter as a diameter, and ignore length and rotational parameters. Again, all features are drawn with the length and width parameters interpreted as fractions of grid cell diagonal length. As a result, the actual feature size depends both on the grid spacing and the length and width parameters.

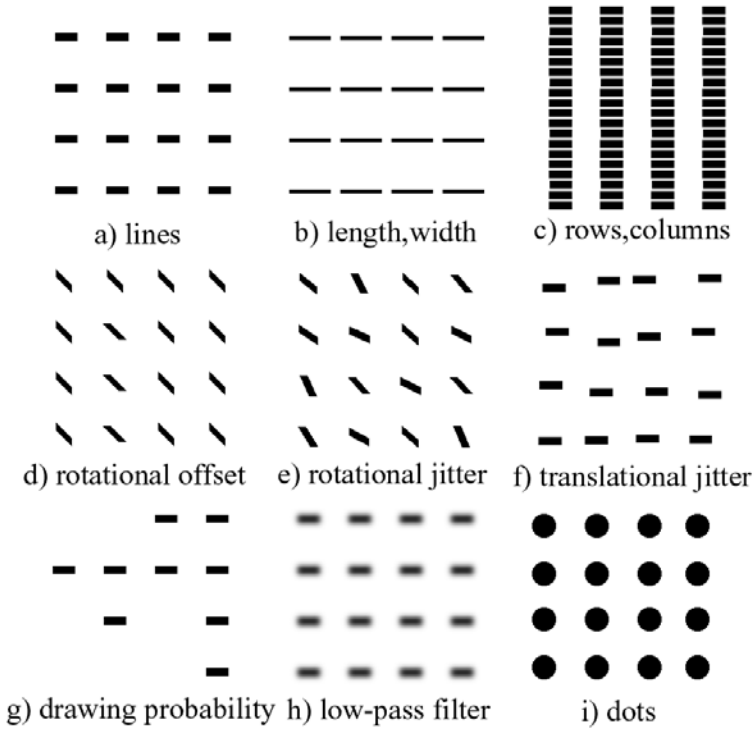



*Figure 26. Mark Parameterization. Parametric variations of marks for each feature layer.*

The method chosen to search the parameter space was chosen to be practical given the speed of human subjects and the size of the parameter space. 122 parameters with two or more levels per parameter are realistically impossible to search exhaustively. Genetic algorithms are traditionally used to search parameter spaces that are prohibitively large for even a computer to search. This study adapted a genetic algorithm to be used with a search by humans. A genetic algorithm generally starts with a generation of whatever is to be evaluated, in this case textured, layered surfaces. Once a full generation is evaluated, the next generation is produced by breeding between textures. The probability of a texture pair being selected for breeding was determined by the experimental ratings. In this analogy, the 'genome' of a texture is represented by the bits of its texture parameters. Breeding in this study was done using a two-point crossover approach, in which a random location on the genome is chosen and the candidate parents switch data around that point to produce two children. The idea of this

form of search is to bias the search toward texture parameters that produce good visualizations, but including some randomness. This meant that subjects spent more time evaluating good parts of the parameter space, while less useful parts of the parameter space were searched more sparsely. Related search methods might be gradient descent, or simulated annealing. Figure 27 shows the distributions of how textures were rated in the first generation (600 initial textures were generated randomly), and the ratings distribution of the complete dataset. The algorithm is successful in producing a bias toward highly rated textures.

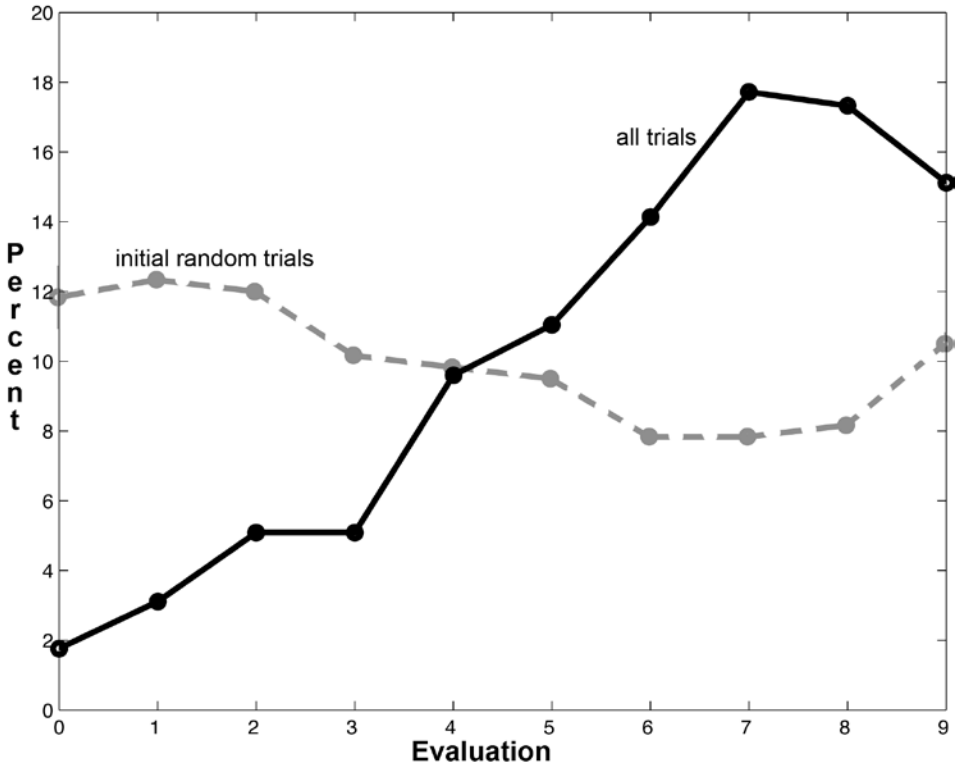



*Figure 27. Ratings Distributions. Initial random trials had even distributions while the entire dataset was biased toward high ratings. [House et al. 2006].*

Because of the speed of human reaction time, the number of textures in a generation was limited to 40, and only a few generations were evaluated in a given

session. Also, given the pilot study nature of the experiment, a simple task was chosen for the evaluation of the textures. Subjects were asked to rate how well they could see the two surfaces on a scale from 0-9. This number was typed using a keyboard stroke, and then the next set of textures was displayed. Five subjects ran through 3 sessions each, rating a total of 9720 texture pairs.

Unfortunately, due to the genetic algorithm approach, false positives are also possible in the data. False positives are parameters that appear to be important but in fact are not. Textures that were rated good by subjects were combined to produce new textures using a genome splitting mechanism with some random 'mutations' occurring. As a result, if a certain parameter is particularly important for good visualizations, then parameters near it in the genome are likely to be passed on to the child textures even if those parameters don't matter at all.

## 3.2 Analysis of Data

Because this method is likely to give some false positives numerous data techniques were used to triangulate and increase confidence in results. Since the parameter space had not been fully tested, the data analysis techniques could not be simple, conventional statistical tests. The techniques used here were more exploratory than evaluative. These included clustering, PCA, neural networks and hypothesis tests on parameter distributions.

### 3.2.1 Clustering

Clustering is a method of grouping points in a multidimensional space according to some distance metric. In our case, points represent textures in a 122 dimensional parameter space, where each parameter has been scaled to a range of 0-1. Textures that were rated highly (8 or better) were clustered using a thresholded minimum distance criterion where distance is measured using a Euclidean metric. This k-nearest-neighbors algorithm can form elongated shapes, unlike some algorithms that only form spherical

clusters. However, in practice, the members of all of our clusters were clearly part of a cohesive group.

Visual inspection of the clusters showed that most highly rated textures had a transparent background with semi-opaque to opaque features, as seen in Figure 28 a) and b). On the other hand, some had milky-transparent top textures like shown in c) and d). Also, most of the good clusters had structured shapes on the top surface due to the underlying grid structure. Some, like the upper textures in e) and f) have more than one set of structured lines. Since each set of marks was not rotated individually, having structure in two directions like this was a result of having mark grids with more rows than columns and mark grids with more columns than rows. A natural right-angle rotation resulted from this method.

Overall, the textures shown are not as convincing as some hand-crafted textures created by an expert might be. It became apparent through the course of running the experiment that the average texture generated randomly using this parameterization was not particularly good. However, given the pilot nature of the study, a very general parameterization was important. If the texture space was more limited, good but unpredictable areas of the space might have been neglected. Also, the textures shown here are smaller and not seen using motion or stereo, and all the images are much more easily understood under these conditions.

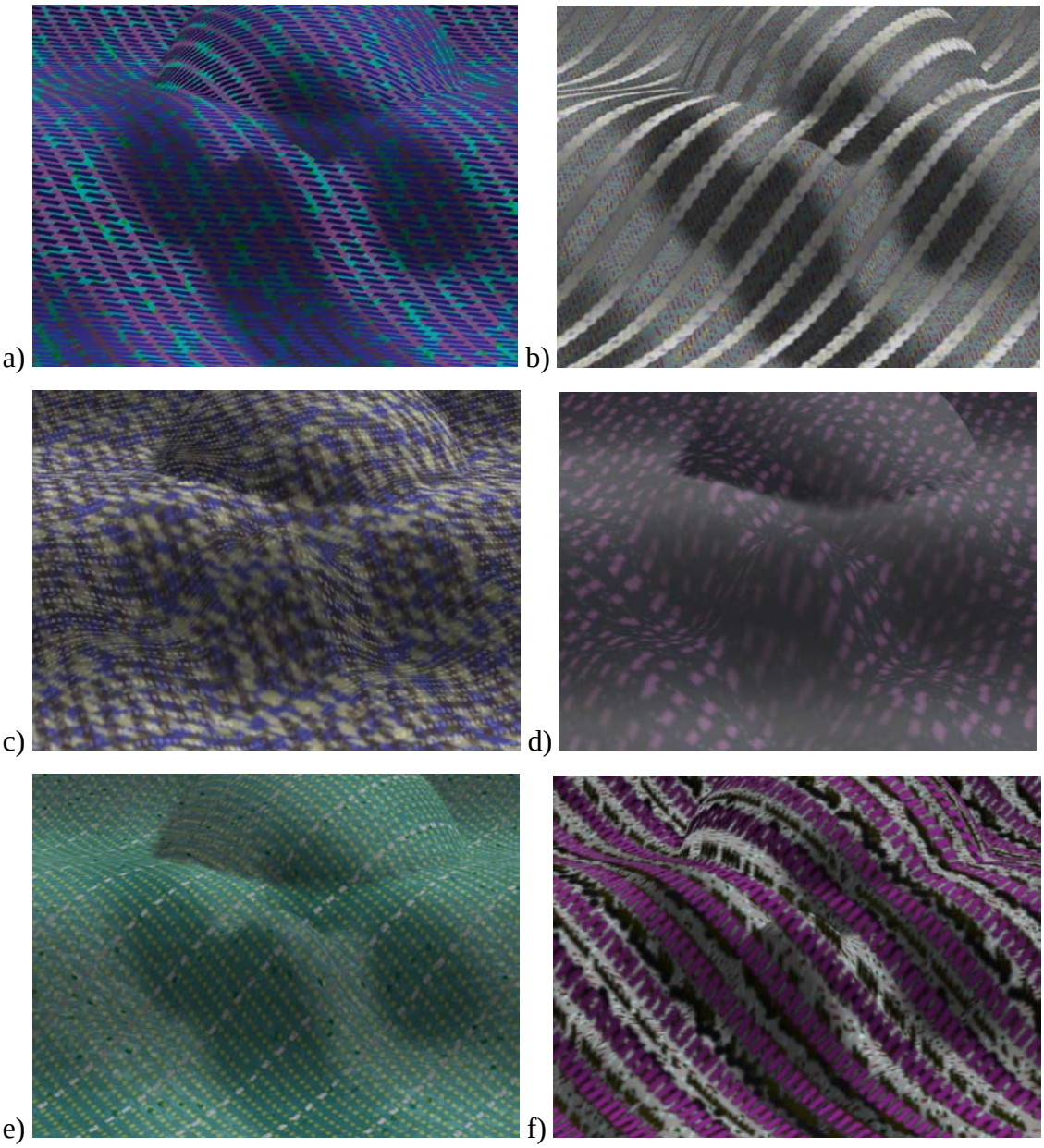



*Figure 28. Clustering Examples. Six texture combinations averaged from clustering results.*

### 3.2.2 Principal Component Analysis

Principal Component Analysis (PCA) is a method that finds directions of maximum variance in a high-dimensional dataset. It is often used to reduce the number of dimensions necessary to describe a dataset. In this analysis it was used to estimate

importance of parameters. Good data from all subjects (rated 8 or better) was analyzed using PCA on the 122-dimensional parameter space. Parameters that were found to contribute the most variance to the good data were assumed to be less important variables. Contribution to the variance was measured by the weighted sum of components in the high-eigenvalue eigenvectors. Note that it might have been simpler and better to simply compare the variance of each variable without performing PCA at all. Either method will give very similar results. The assumption is that if a parameter has high variance within a group of good textures, then the value of that parameter must not affect quality very strongly.

Rather than directly comparing the variance of each parameter, the variance was weighted by the parameter's number of levels. This was for two reasons. In general, a randomly distributed variable will have larger variance with an even number of levels than an odd, because the odd will have some proportion of the samples on the mean. Also, as the number of levels increases, the variance decreases for evenly spread distributions. In our experiment, different parameters had drastically different numbers of levels, anywhere from binary to a 32-bit floating-point number. As a result, variables that were found to have high variance were compared with their maximum possible variance before being judged unimportant. This method provides a fair comparison of the significance of individual parameters on texture rating. This does not rule out the possibility that an interaction between two or more parameters might be important, but is a good first pass to estimate parameter importance.

Figure 29 shows a cluster mean and two textures created by adding and subtracting the scaled principle eigenvector. The eigenvector changed the top rotation, as well as the hue, size, randomness and opacity of one of the top features. The texture on the bottom remains small and grainy, while the main features on the top, the lines, do not change at all. The cluster was rated an 8, and contained nine texture sets. Since cluster sizes were generally not large enough to run PCA for 122 variables, eigenvectors were found from the combination of several nearby clusters. It should be noted that if the scaling of the eigenvector puts the new texture set outside of the domain of the original

cluster, this analysis is no longer necessarily valid. Even so, in many cases it was found through informal experimentation that the eigenvector can be scaled far beyond these boundaries and the textures still have high quality. This gave confidence that the eigenvectors were a good measure of parameters that could be freely varied.

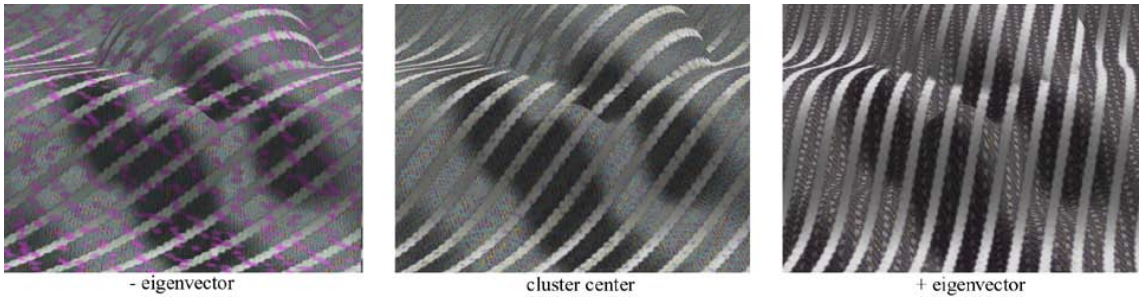



*Figure 29. Eigenvector-Based Textures. Textures were made by moving along the primary eigenvector from the cluster center.*

With the 122-dimensional data, the principle components have relatively large parameter components in many dimensions, making them difficult to interpret. Also, depending on the individual cluster, it requires around 85 of the 122 eigenvectors to account for 90% of the variance. This fits with the intuition that the texture space is highly non-linear, and the parameters interact in complex ways to make perceptually good textures. However, some information was extracted from the complicated eigenvectors.

First, the parameters were ordered according to which ones tended to have more average variation in all of the good clusters. The eigenvectors with the ten highest eigenvalues were selected. Then, within these eigenvectors, the parameters were ordered by the sum of each parameter's five highest magnitudes. This approach was chosen as being reasonable through inspection of the data. It kept only the five highest values for each parameter because most of the small values in eigenvectors are likely to have a lot of noise due to the nature of principle component analysis. Parameters with the highest

sums might therefore be useful as free parameters when constructing good textures. This method does not provide specific rules for making good textures, since variation across the parameter space is ignored. However, it does give an indication of which parameters are more important, and suggests areas that should be looked at more closely.

Several trends were clear. Comparable parameters always varied more on the top surface than the bottom surface. Also, with the exception of opacity, comparable parameters for the surface background varied more than those for the features. This implies that the bottom surface characteristics are more important than the top, and the element characteristics are more important than the background for creating good textures. The bottom-top importance difference is suspect, however, because the top surface had so much less detail than the bottom. This would have to be further explored to be a viable hypothesis.

The color variables hue, saturation and value, were parameters of interest. In all cases, the hue and saturation variables had more variation than value. Certain settings of the value parameter are likely to be much better than others for creating good textures. On the other hand, saturation and hue might be free variables that can be used to encode other information, or simply to change the visualization to aesthetic taste. Interestingly, the parameters that encode the shape of the elements, such as the number of rows and columns in the grid, size and shape of the elements, and randomness of the features, always varied less than the color parameters. Thus we can conclude that the features must have good placement, size and shape before parameters like color, rotation and filtering can have much of an effect on visualization quality. Finally, the opacity of the top surface background and features varied more than expected. This is probably because the actual coverage of the top surface is a complex function of the four opacities, the size, randomness, separation and probability of being drawn of each of the features. The actual top coverage is explored further in the hypothesis testing section. Finally, binary variables, such as those used to switch randomization on and off, also displayed very high variance, but this was considered to be a false positive since a binary distribution is strongly biased toward higher variability than a continuous variable.

### 3.2.3 Neural Networks

Neural networks were used as another check on importance of each parameter, but more importantly as a test to determine whether the values should be high or low. The type of network used was a fully-connected, back-propagation network with sigmoid transfer functions [Haykin 1999] [Craven and Shavlik, 1997]. Sigmoid transfer functions output a high value (usually 1) if the weighted input is above some threshold and a low value (-1) otherwise. The training of a neural network is an algorithm that allows the network to learn weights associated with each node that lets it 'classify' data. This means that given a new texture as defined by its parameter values, the neural network should be likely to correctly rate the texture as a human would. In a single layer network, the network training corresponds to finding a linear combination of the parameters that gives the best classification for each set of input parameters. Theoretically, using two layers with enough nodes is sufficient to reproduce any function. For this experiment, two network layers were used, with 122 inputs corresponding to the parameter space, 20 hidden nodes, and 10 outputs corresponding to the 0-9 texture ratings. This network structure is shown in Figure 30. Each texture is input as a vector of parameters, and the output is one of the classes (0-9).

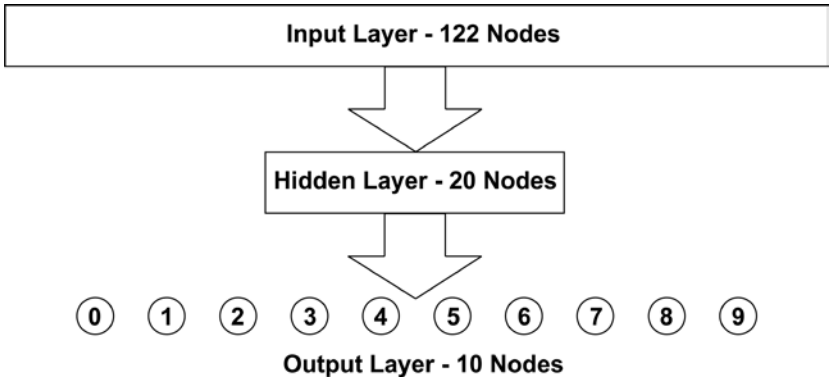



*Figure 30. Neural Network Structure. [House et al. 2006].*

Of the 9720 data points, 9000 were randomly chosen for training and the remaining 720 for testing. The 20 hidden units are a large data reduction, but the network learned to categorize with reasonable accuracy. Figure 31 shows the histogram of ratings given by the network for textures rated as a 9 by humans, showing that most are rated between 7 and 9. Histograms for the other rating groups had a similar spread, for example misclassifying a 4 as a 5, rather than misclassifying a 0 as a 9. It should be noted that the human ratings are subjective and almost certainly varied from subject to subject. Although a network could learn an exact mapping of textures to weights, it would probably not generalize well to another dataset. Therefore, care was taken that the network was not over-trained on the data, or given too many hidden nodes.

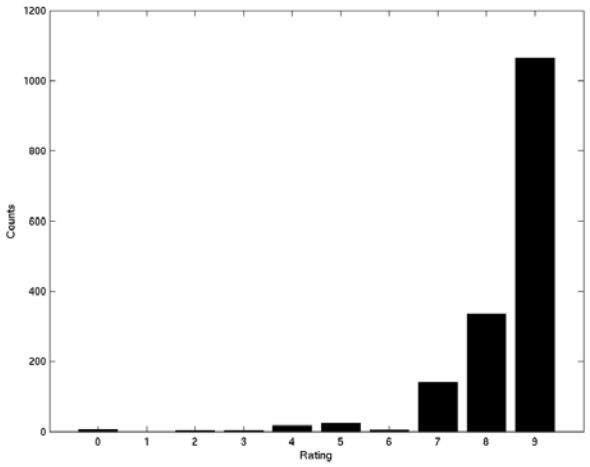



*Figure 31. Network Classification Accuracy. The network correctly classified most of the textures rated 9.*

Given a network that does a good job classifying the textures, understanding the meaning of the weights is still difficult. The non-linearity of the sigmoid function prevents a simple analysis of weight vectors. However, simply looking at the parameters with large magnitude weights proved interesting. Fortunately, many of the weights leading to the output layer were very small. If the weights leading to a specific output

node are positive, large values of those parameters will increase the chance of the class being selected. Likewise, large values with negative weights decrease the chance of getting a particular rating.

Analyzing the weights for textures rated as 9 produced the following guidelines. The top background should have a small alpha, little blurring, and a high rotation. The top lines should have a grid with few rows, and the lines should have small length and thickness, small horizontal and rotational jitter, but large vertical jitter. The second set of lines should not be drawn, and the top dots should have a low probability of being drawn, a small size, and on a grid with few columns. The bottom surface should have a bright background color, with bright, saturated lines with few columns, low saturated dots with a small radius, and a lot of vertical variance.

Figure 32 shows a texture that was created based on these feature characteristics, where parameters unspecified by the analysis were generally set to the median parameter value. It actually works well as a single image, but is especially effective when rocked so that motion cues are available. The anisotropic grid structure on top forms banding across the whole texture, and the high rotation on the top maximizes the comparative rotation between the perceived lines on top and bottom. Interestingly, the high vertical jitter on the top lines does not greatly increase the perceived randomness.

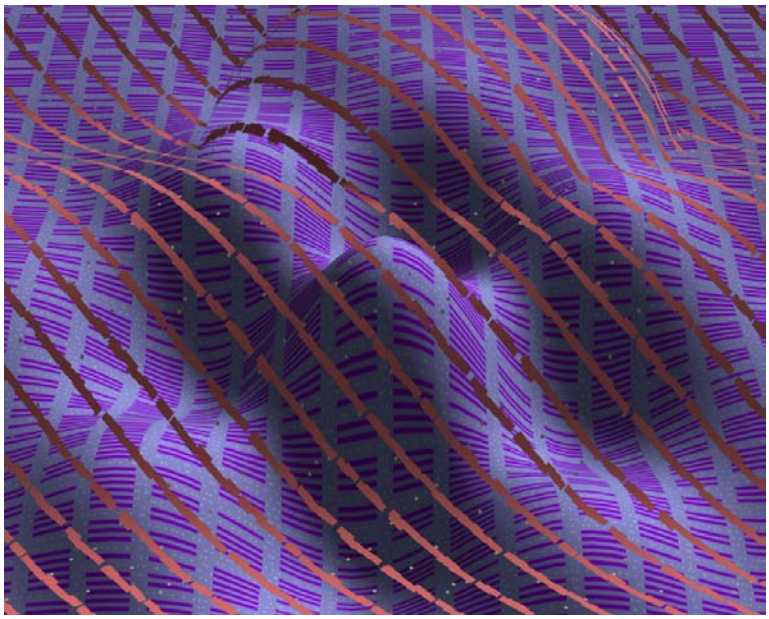


*Figure 32. Network Rule-Based Texture. Long lines and small dots on top and dashes on the bottom show both surfaces.*

The results for textures rated as 8 had similar weights to those weighted 9, except they showed a tendency for the bottom surface to have a blue-violet hue, and the top dots a red hue. Interestingly, the hypothesis tests discussed later actually show a slight trend in the opposite direction. Given the conflicting results as well as the lack of importance seen for hue in the PCA analysis, these trends are ignored as probably insignificant. The weight magnitude in the neural network also measured the importance of parameters. Similar to the PCA analysis, the shape parameters again have much stronger weights than the color parameters.

As a check, the weights were analyzed for bad textures – those rated as 0. The results made sense. The top background is saturated and opaque, both the top lines and dots are small and transparent, and the bottom surface has saturated lines. Simply put, bad textures have top textures that obscure the bottom layer, and with hardly any texturing visible on the top.

### 3.2.4 Hypothesis Testing

Given the results from the clustering, PCA and neural networks, as well as experience with layered surface texturing, various hypotheses were made and tested on the data through histogram comparison. Matlab was used for this portion of the analysis. Hypothesis testing usually takes the form of a statistical measure like a t-test. This type of test calculates the probability that the measured means of two sample groups are different by a given amount. The probability is based on the assumption that the samples all come from the same underlying distribution. For example, a t-test could measure whether oak is harder than maple by comparing the mean measured hardness of a group of oak planks with a group of maple planks. If the probability that the samples are from the same distribution is low enough, it is called rejecting the null hypotheses. In the context of this dissertation, a t-test be used on the texture data to measure if the texture brightness of highly-rated textures is higher or lower than for randomly-generated textures. However, the mean is only one statistical measure of a distribution.

To get at more fine-grained information, for example which values of brightness seem to work best, histograms of good data were compared with histograms of randomly generated textures. Parameter levels that are more common in good data than in random data are probably preferable to other parameter levels. This section discusses the various hypotheses that were tested in this manner. The good data used for these tests are all textures rated either 8 or 9, and contained a total of 3078 texture pairs.

In this experiment, colors of the surface backgrounds and features were constructed with HSV (hue, saturation and value) parameters. The PCA analysis showed that among the color parameters, value (brightness) was most important, saturation less so, and hue unlikely to be important at all. The neural network analysis showed a trend for high value background and marks on the bottom surface. Based on the human visual system's tendency to respond to changes, differences in the color parameters between the top and bottom surface were considered for hypothesis tests. Since the textures were made from three layers of dots and lines, color differences between surfaces were found by comparing the color of the most prominent features, i.e. the features that covered the

most area with the highest opacity and probability of being drawn. Value difference across the two surfaces is shown in Figure 33. The distribution of texture brightness for the good textures is compared to the distribution for random textures. Peaks are visible to be significantly higher than random at -0.4 and +0.4 brightness differences. The confidence intervals displayed are 95% based on the binned data around each point. The results show a preference for textures with about a 40% difference in brightness across surfaces, with the better choice being when the bottom surface is brighter.

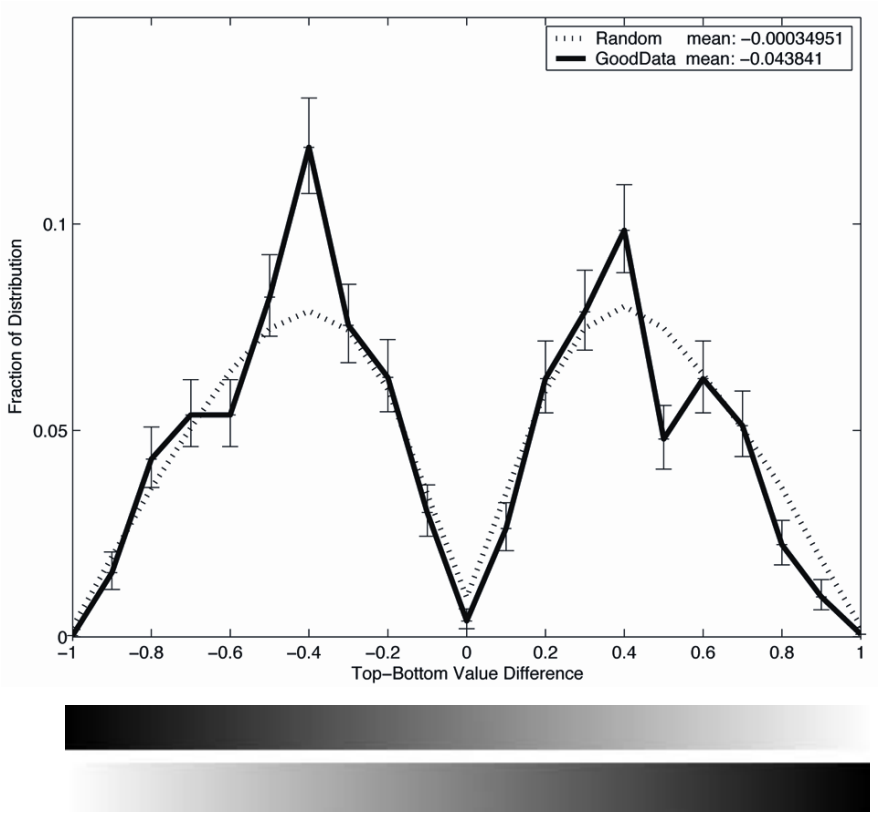



*Figure 33. Value Difference Histogram. Difference in value between the top and bottom surfaces.*

The optimal brightness values for both the top and bottom surfaces are also interesting questions. Rather than using the estimate of the surface brightness based on prominent features used above, the actual texture images were used, with the overall brightness found using averaging. Figure 34 shows the results for the bottom surface brightness. The plot was created by convolving a smoothing kernel with the brightness distribution of good data, and subtracting the distribution of the random data. Confidence intervals were based on data in bins of width equal to the size of the smoothing kernel. The results show a preference for a brighter-than-average bottom surface, with two peaks at about 45% and 80% brightness. A Kolmogorov-Smirnov test showed the distribution is different from random with high significance ($p=0.000068$). In contrast, the top surface distribution of values was not significantly different from the random distribution. This agrees well with the results from the neural network analysis. Based on these results, the preference for -40% difference in top-bottom brightness might simply be due to a preference for a relatively bright bottom. However, since both difference peaks were near 40%, it is safe to say that a difference in brightness probably has an effect.

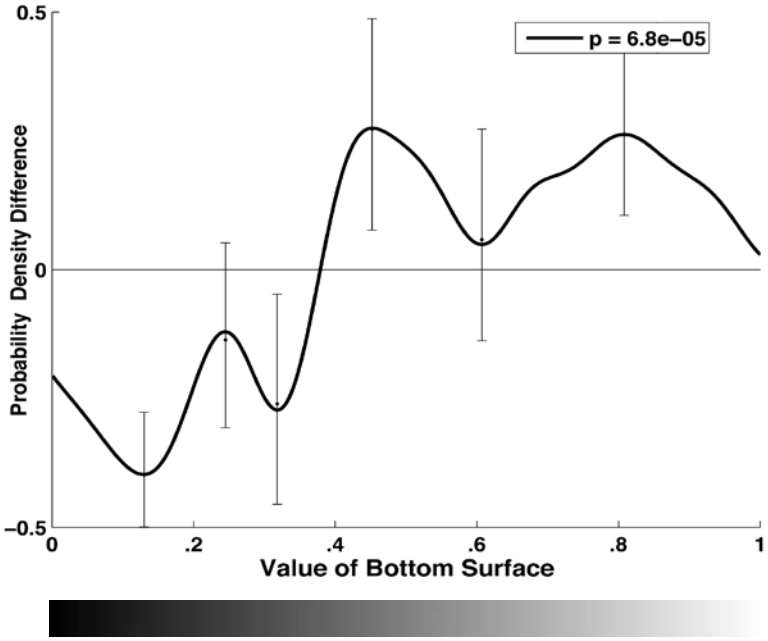



*Figure 34. Bottom Value Probability Density Difference. Bottom surface color value distribution shows a preference for high brightness. In this case the probability distribution of the brightness is shown relative to the random distribution, so that areas above zero are more likely levels for good data. [House et al. 2006].*

Although the colors were drawn using HSV color parameters, the perceptual difference in saturation is more complicated than simply subtracting the color saturation values, since value affects the amount of perceived saturation. In the extreme case of black, there is no difference between fully saturated and unsaturated color. Therefore we measured the difference in saturation between surfaces as

$$S_{difference} = (S_{Top} - S_{Bottom})V_{Top}V_{Bottom}$$

where *S* denotes saturation and *V* denotes value. The nonlinearity of this equation is what produces the spike at zero seen in Figure 35. As described in the value difference hypothesis test above, the saturation was computed based on the saturation of the most prominent texture feature. The most obvious pattern in the saturation distribution is a preference for textures where the bottom is about 70% more saturated than the top.

There is a corresponding lack of good textures where the top is more saturated than the bottom. It seems clear that a difference in saturation seems to be significant.

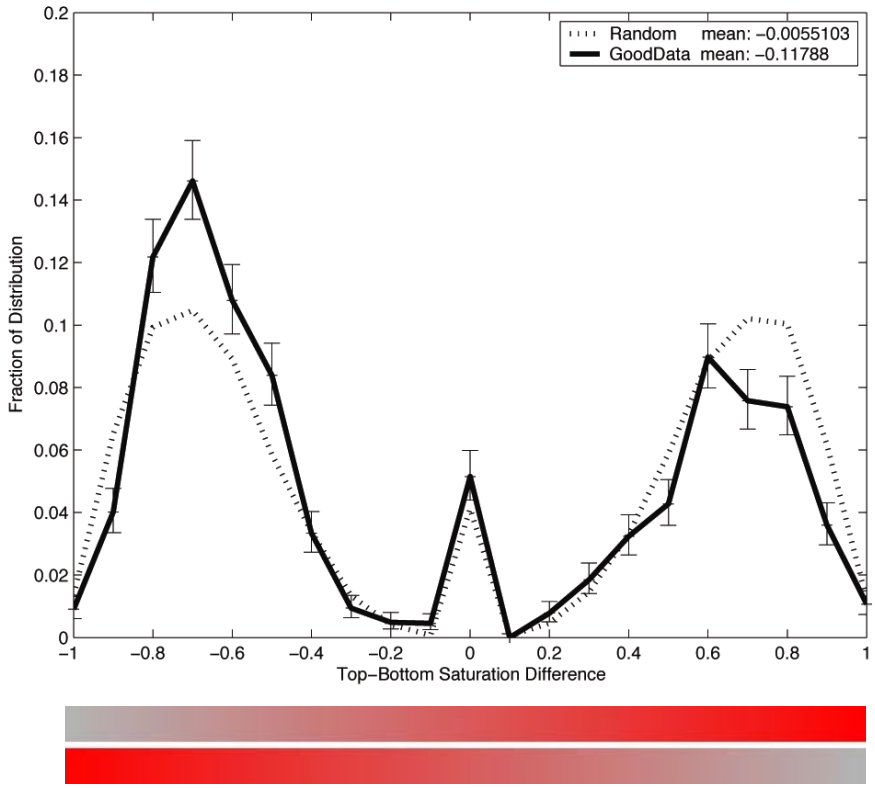



*Figure 35. Saturation Difference Histogram. Difference in saturation between top and bottom surfaces.*

The overall opacity of the top texture is likely to be a large factor in how well each surface can be seen. Therefore this parameter, which we will call 'coverage' is the object of the next hypotheses test. Because of the complex interaction between the top background layer and the three layers of textures, a measure of top coverage was estimated as follows:

$$
\begin{aligned}
&C = \alpha_{back} \\
&For \quad (i = 1 : 3) \\
&\{ \\
&\quad C_i = \alpha_i p_i A_i \\
&\quad C = (1 - A_i) C + A_i (C + C_i) \\
&\}
\end{aligned}
$$

Here, $C$ is the total coverage, $\alpha_{back}$ is the background opacity, $A_i$ is the area covered by a feature layer, $\alpha_i$ is the feature opacity and $p_i$ is the probability of being drawn for a feature layer. This estimate assumes that the features are drawn randomly. Note that for some cases, this measure of coverage is not accurate. If some marks consistently cover other marks, the actual coverage will be lower than the estimated coverage. For this reason, the average coverage is likely overestimated for high coverages and underestimated for small coverages. With this in mind, the peak in Figure 36 between 30-50% would probably move toward 50%, and be more pronounced. Interestingly, the top coverage mean for the good textures is almost exactly at 50% (0.5062). Therefore we estimate that an optimal top coverage is somewhere in a range between 40-50%. Only five bins were used because the opacity parameter was discrete and had five levels. One other point that should be made is that the random distribution of the top coverage is biased toward high coverage. About 35% of the textures are in the top 80% of the coverage. This could be important because overly high top opacity could bias other results. For instance, the importance of a bright bottom could be due to the need to see through an excessively opaque top.

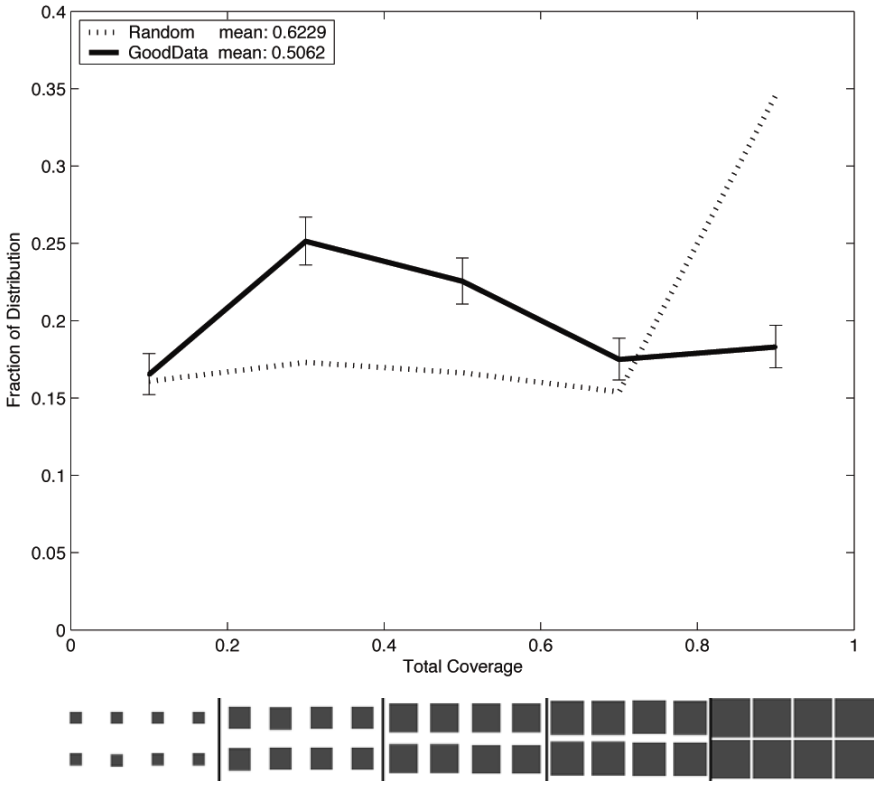



*Figure 36. Top Coverage Histogram. Estimate of total coverage on the top surface.*

To get a better estimate of the actual average top opacity, the texture images were used. Coverage was simply computed as the average image opacity of the top textures. Figure 37 shows a kernel-density estimate of the good texture coverage distribution after the random distribution has been subtracted. As before, in the image brightness probability distributions, the random distribution has been subtracted from the good-data distribution. Again, there is a peak near 50%, and another at low opacities of 20%. Both histograms support the idea that middle to low overall top opacities are best.

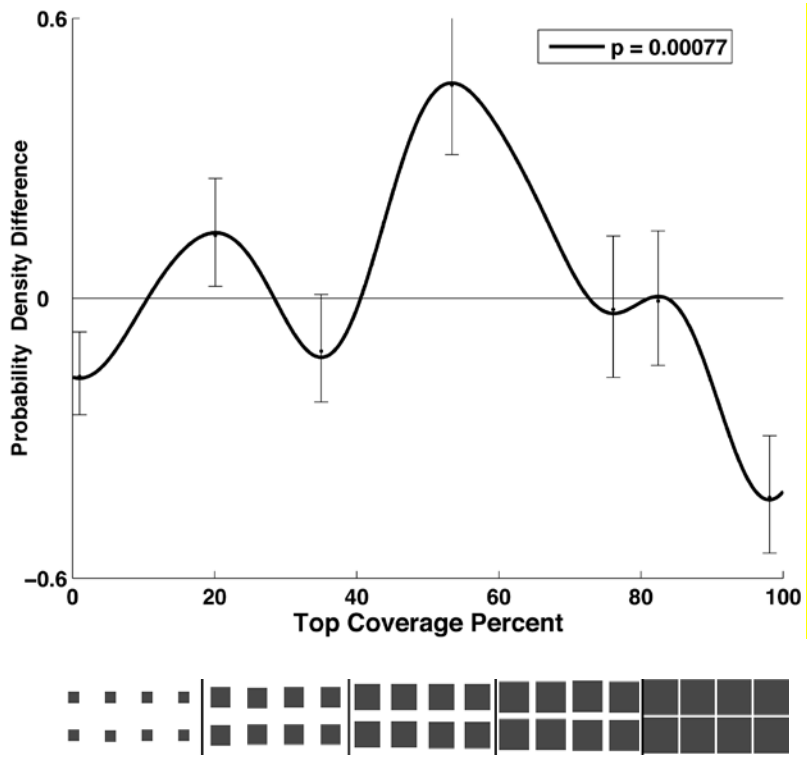



*Figure 37. Top Coverage Probability Density Difference. Distribution of average top texture opacity. The random distribution has been subtracted from the good data distribution. For more explanation, see Figure 35. [House et al. 2006].*

The next factors that seemed very important based on the previous analysis techniques were the various texture shape parameters. To review, the textures were built from layers of marks laid out on a grid. The grid could vary in numbers of rows and columns, and the marks fit within the grid cells. The marks were given a relative size and aspect ratio within the cell.

First we discuss the area of the features and the area of the grid cells, since the two both contribute to the total size of the features. Figure 38 shows the product of the number of rows and number of columns for grids on both top and bottom surfaces, with the random distribution for comparison. Both top and bottom surfaces show a preference for the relatively big grids, but the means show that the top in general should be less dense and the bottom more dense than random. Because marks are sized relative to the grid cells, this would produce larger marks on top and smaller on the bottom.

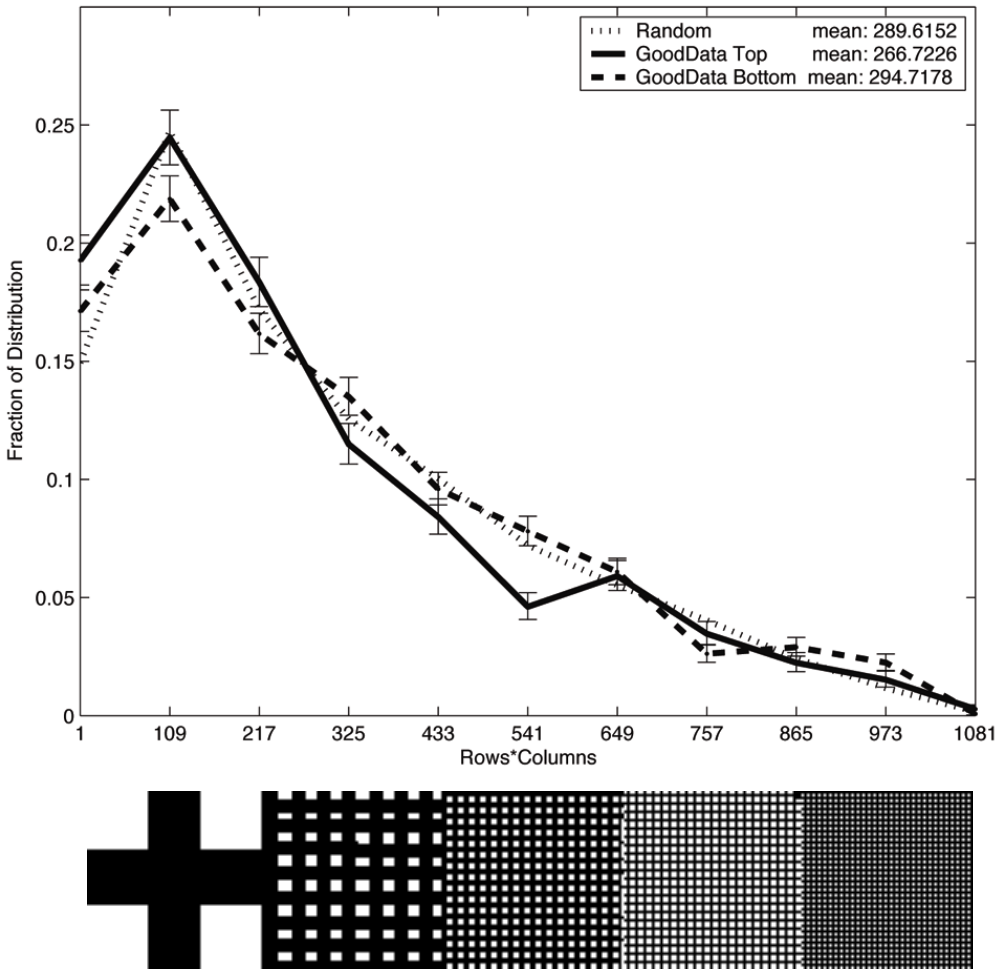



*Figure 38. Grid Density Histogram. Grid density (rows x columns).*

The marks had parameters of size and aspect ratio that determined the mark shape relative to the grid cell. The results shown in Figure 39, display the percentage of the grid cell area that is covered by the marks. As with the estimates of color on a surface, the percentage was estimated for each layer based on the most prominent layer of marks, with prominence being a combination of mark size, opacity, and probability of being drawn. The means show a preference for the top marks being larger and the bottom marks smaller than the random distribution. The main trend is for the top to prefer marks covering about 30% to very small marks covering only about 10%. Combined with the slightly larger average grid sizes for the top, it would seem that for

this case the top surface should have larger textures, and the bottom smaller textures. This conclusion must not be taken at face value, however, because the top surface used was much lower-frequency and simpler than the bottom surface. If the brain is using the texture to reconstruct the shape of the surface, and imagining using the Nyquist limit as a simple model, a low-frequency surface would require a less-detailed texture. However, here we note the size discrepancy as an interesting pattern to be looked at in further research.

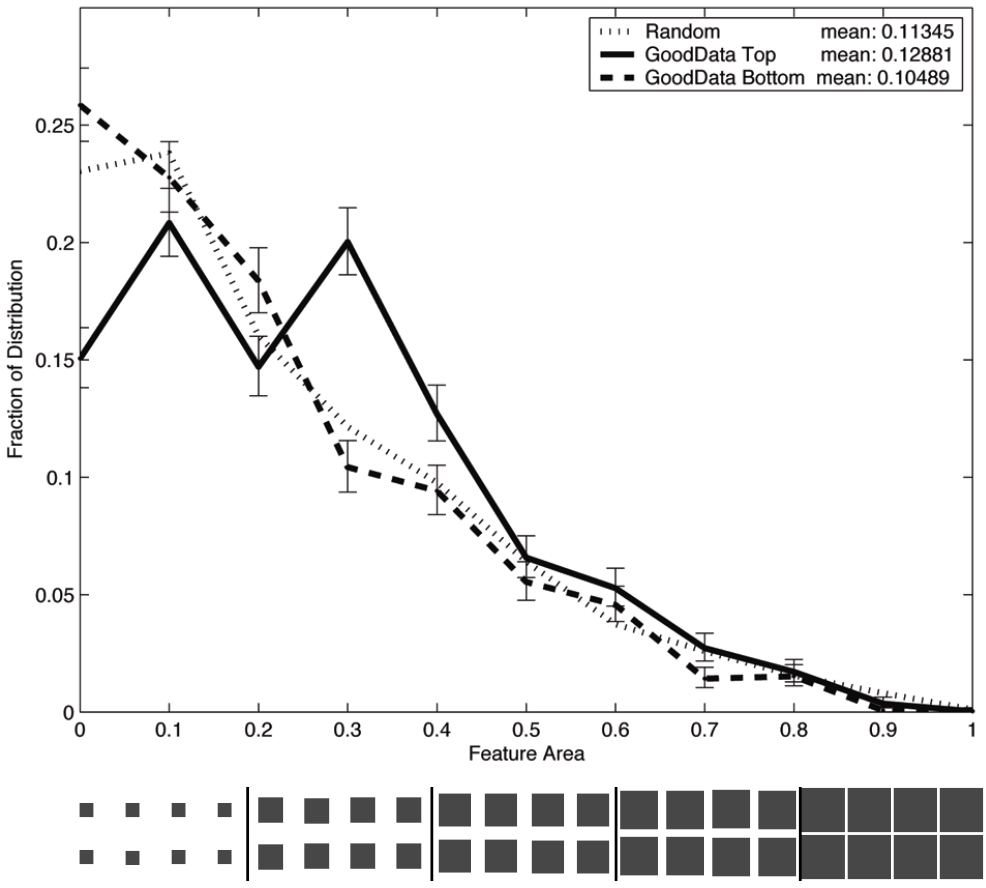



*Figure 39. Grid Cell Percentage Histogram. Percentage of the grid cell area that is filled by the marks.*

Looking at the cluster images, several examples included textures with uneven number of rows and columns so that the large-scale ordering of the marks created parallel lines along the surface. To measure if this was a significant factor in making good textures, the grid aspect ratio distributions were tested. Figure 40 shows the results, where the random aspect ratio distribution has been subtracted from both the top and bottom distributions so that the curves can be compared with the zero line. Both the top and bottom distributions were significantly different from the random distribution, and both had a bias against the square, equal row and column grids. It appears that the large-scale lines produced by uneven grid cells were useful for good textures. Another striking pattern was that the top and bottom surfaces seemed to prefer different aspect ratios. It is possible that as the genetic algorithm evolved, the two surfaces starting with random distributions co-evolved to tend toward different ratios. The grid size ratio shows a preference for the top having more pronounced lines of features; that is the top grids had a higher overall aspect ratio. This could either be the result of the top surface having lower frequency structure and thus showing up better with large-scale patterns and order, or simply because of a slight initial random bias in the distributions, which was accentuated by the observed modes.

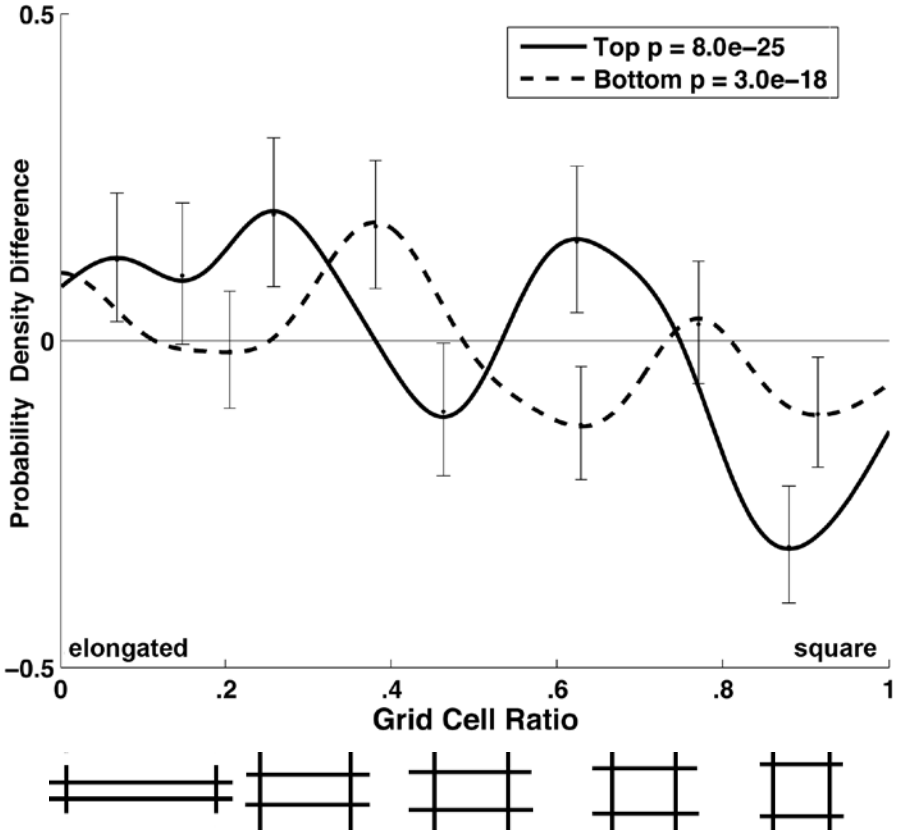



*Figure 40. Grid Aspect Ratio Probability Density Difference. [House et al. 2006].*

To test if the difference in top-bottom surface aspect ratios was real or if the specific ratio lobes seen above were significant, the distribution of the difference in aspect ratios is shown in Figure 41. Interestingly enough, the strong peak near zero shows a clear preference for the top and bottom surfaces to have nearly the same grid aspect ratio. From this, we assume that a difference in aspect ratios is not particularly important and may even be bad, but the top surface works better with more elongated grids on average than the bottom surface.

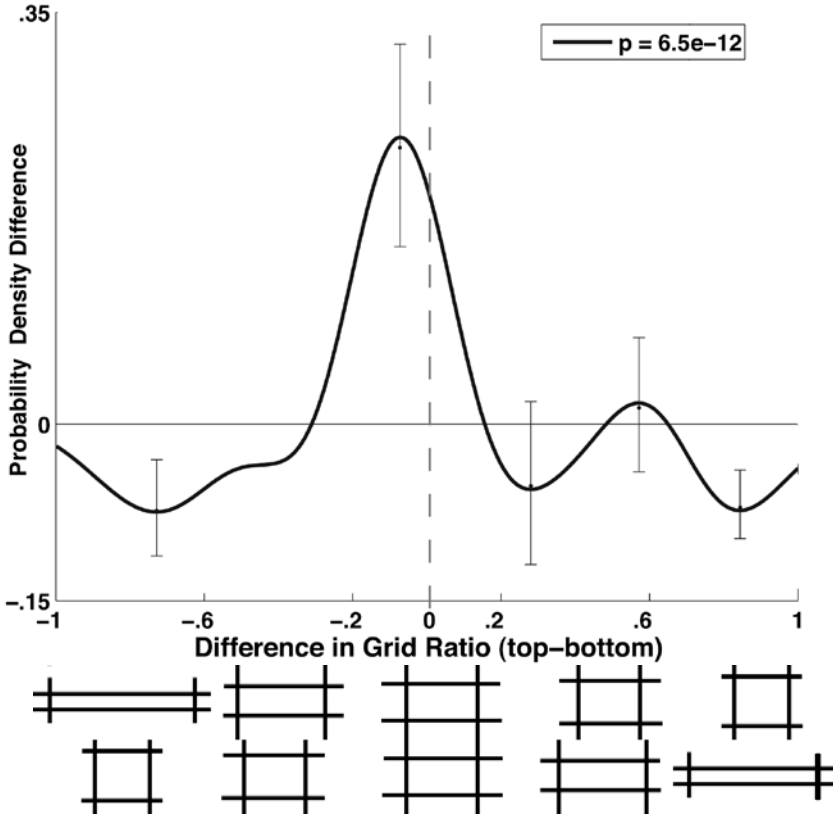



*Figure 41. Grid Aspect Ratio Difference Probability Density Difference. The top-bottom aspect ratio distribution, where the random distribution has been subtracted. [House et al. 2006].*

Next, we look at the rotation of the two surface textures. Figure 42 shows the distribution of rotations for the top and bottom surfaces. Note that in this case, since rotation of a surface is defined by only one parameter, the random distribution would be flat and so is not shown. Since the features are all arranged on grid structures, which tend to create both horizontal and vertical banding, it makes sense that some relative rotation might help to visually separate the two surfaces. We see a strong tendency for the top surface to have about a 25° rotation, while the bottom surface tends to be at -45°, 5-15°, or 45°. This suggests that about a 10-20° or a 70° difference in rotation across the surfaces might be optimal. Certainly, the strong peak at 25° for the top rotation provides global lines that go at an angle back and to either the right or left depending on the

particular grid aspect ratio. These diagonal lines might have advantages over lines going directly back or horizontally.

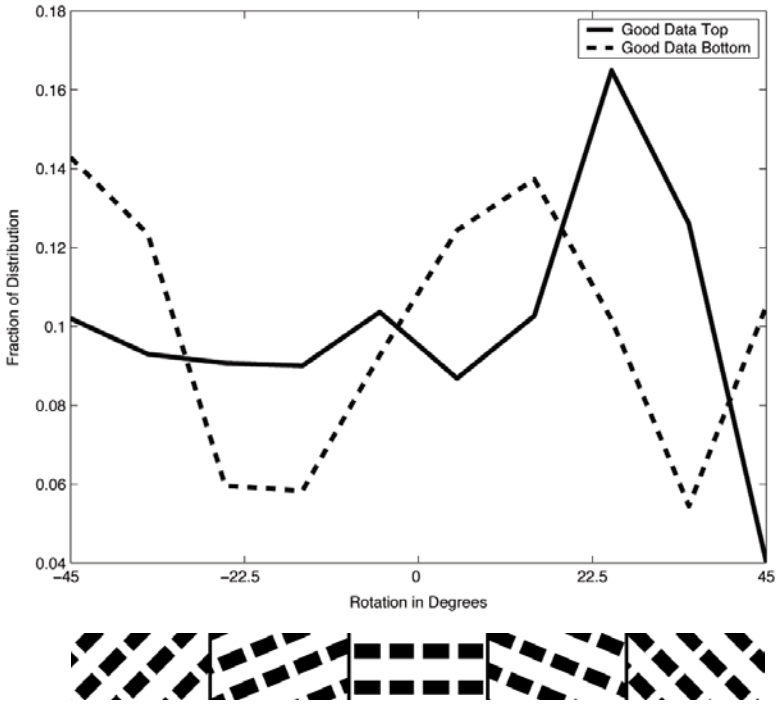



*Figure 42. Rotation Histogram. Rotation of top and bottom surfaces.*

The distribution of rotational difference shown in Figure 43 confirms that several rotational differences are preferred. Either 10°, 40°, 60° or 75° are best. Surprisingly, the differences near 90° do not seem to work well. Probably, this is because the 90° case is confounded with the 0° case when the grids have opposite aspect ratios.

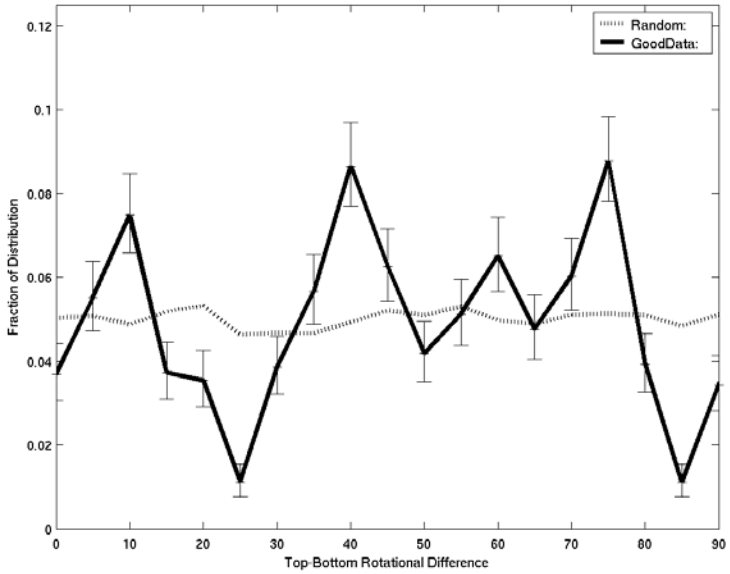



*Figure 43. Rotation Difference Histogram. Top and bottom surface rotational differences.*

Low-pass filtering was included both as a possible aesthetic aid, and to possibly simulate depth-of-field cues. Figure 44 shows the filtering distributions for the top and bottom surfaces. Note that filtering was turned on and off with a binary parameter, so the random distribution is biased to have half of the textures with no filter and an even distribution for the other pixel widths. The bottom surface rarely uses the filter, but has a high width when it is used. The top surface uses the filter more often, but uses a smaller width when it does. Because there are so many differences between the top and bottom surfaces already, such as the overall mark sizes, grid sizes and the way opacity interacts with color and blurring, these results are difficult to interpret for meaning. So we shall simply say that both surfaces had lower than random blurring, with the bottom surface using even less average blur than the top.

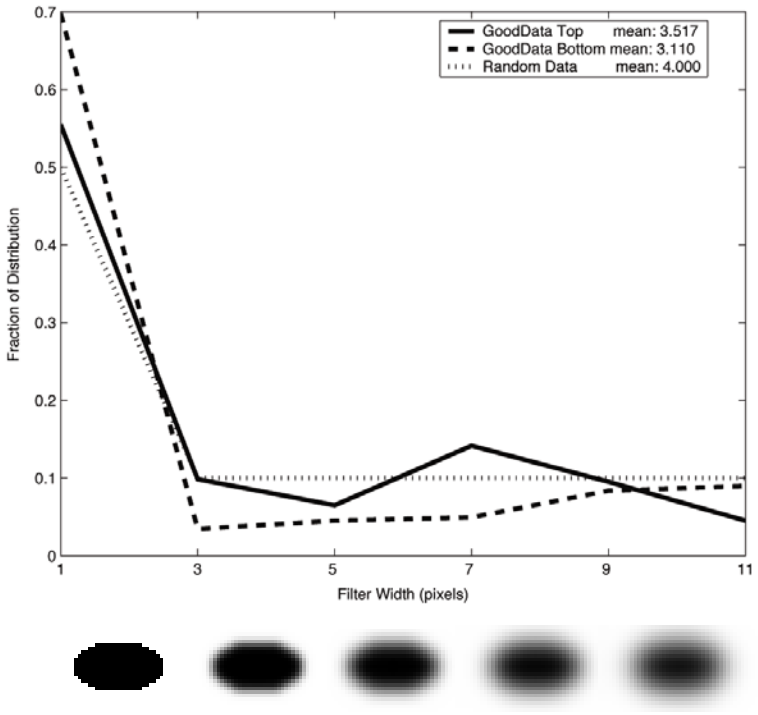



*Figure 44. Filter Width Histogram. Filtering width distribution on top and bottom surfaces.*

Finally, we look at the randomness of each surface, as parameterized by horizontal, vertical and rotational jitter. We analyze the translational parameters separately from the rotational because there was no way we knew of to find a correspondence between perceived randomness with translation or rotation. Like the color comparisons, randomness comparisons were made between the most prominent features on each surface. Figure 45 shows that larger differences in randomness are preferred, with a bias toward the bottom being more random than the top.

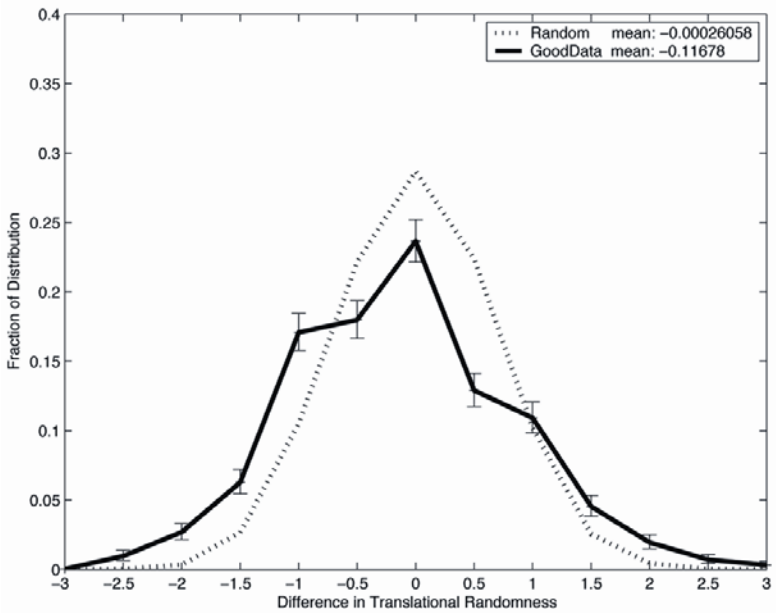



*Figure 45. Translation Jitter Difference Histogram. Translational jitter differences between top and bottom surfaces.*

The rotational jitter comparison shown in Figure 46 also shows a preference for larger differences, and this time a bias toward the top being more random than the bottom. However, the bias magnitude is only about 1/15th the size of the translational randomness bias. Unlike grid aspect ratio, it appears that an overall difference in texture randomness (translational and rotational) is good.

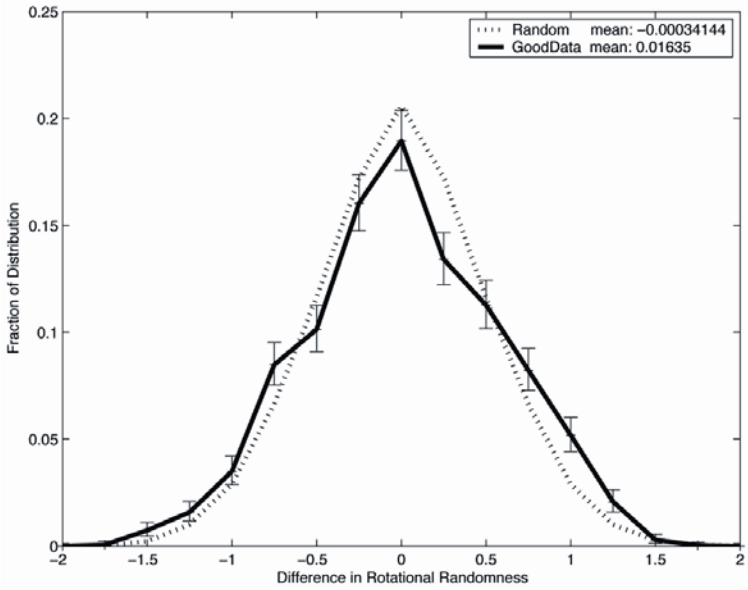


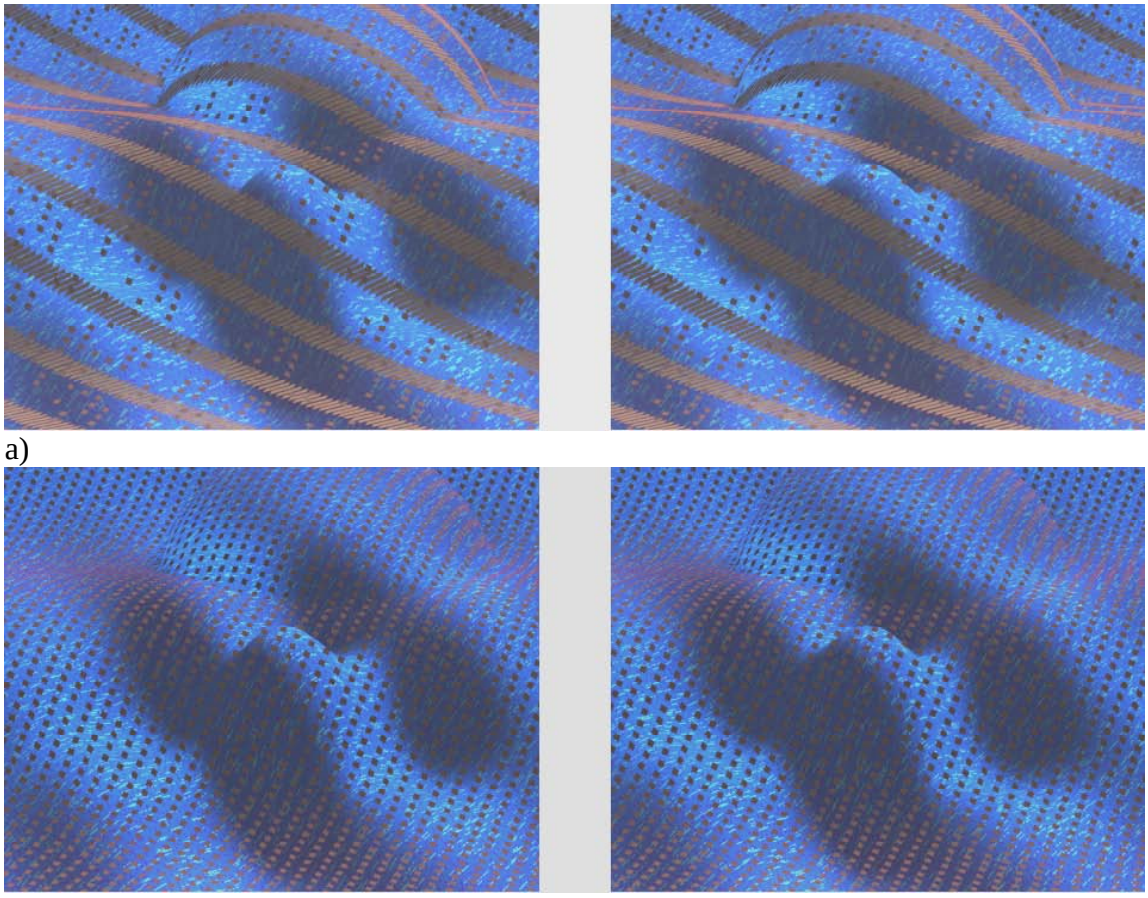

*Figure 46. Rotation Randomness Difference Histogram. Rotational randomness differences between top and bottom surfaces.*

### 3.3 Validation of Results

The results from the various data analysis methods were combined to suggest general guidelines for making layered surface textures. Textures were built based on these guidelines to demonstrate their effectiveness. Also, textures were built breaking several of the guidelines to show the effects. These are shown in Figure 47 as crossed-eye stereo pairs. These give a feel for how the images would look in the experiment, but their low resolution and lack of motion cues make them considerably weaker than they actually appear on screen. Nevertheless, they do convey the trends we describe below.

Figure 47 a) is a default texture hand-created according to our results. Each surface has two feature layers with appropriate saturations, values, randomness, and filtering. The grid sizes and feature areas were picked from the most common range in the distributions, and the grid and feature aspect ratios were picked from the peaks for each surface. On the bottom both feature layers were drawn with probability 100%, and on the top the first layer is drawn at 100% and the second at 40%. Figure 47 b) shows a version where the second feature layer on the top is drawn at 100% and the first not at

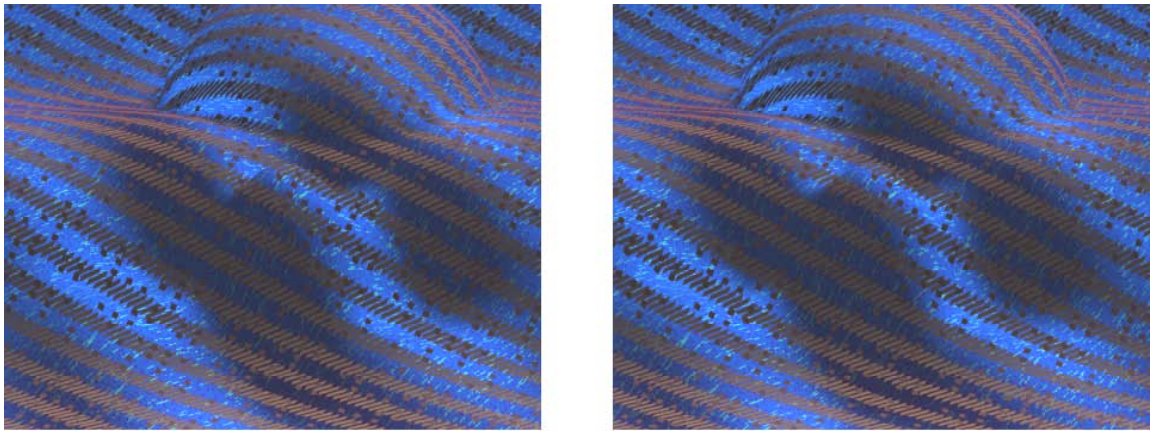
all. While not quite as striking as a), it still does a good job of showing both surfaces. Note that when we left the first layer being drawn at 40% the larger features were quite distracting. Although not one of our rules, it makes sense that randomized features should be smaller than regular ones to avoid this distraction. Figure 47 c) shows a variation in the top grid within the allowable range. The textures are still quite good. Finally, Figure 47 d) shows a version with the blue and red hues switched between the top and bottom. The effect is much more strident, but still visually very readable, demonstrating our finding that hue is freely variable.



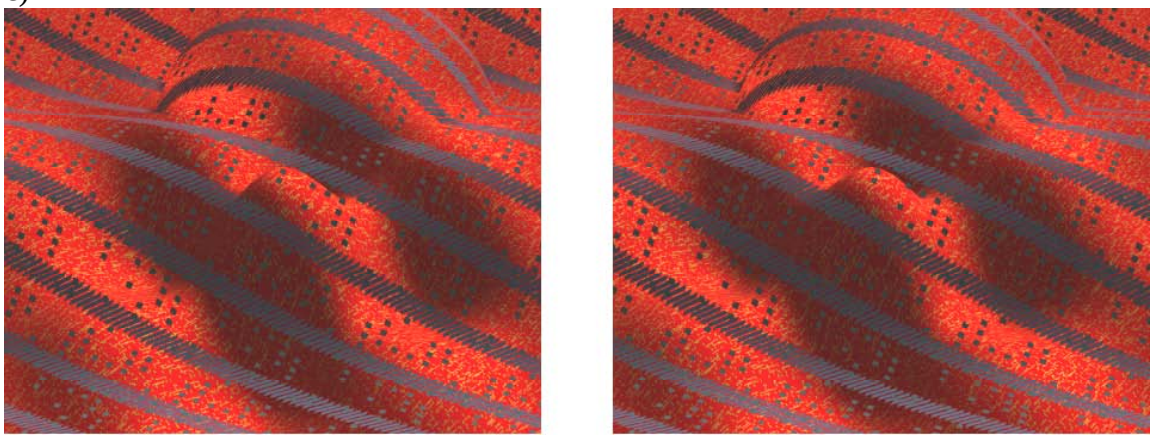
a)

b)

*Figure 47. Rule-Based Textures. Variations on a default texture within the boundaries of the guidelines.*



c)

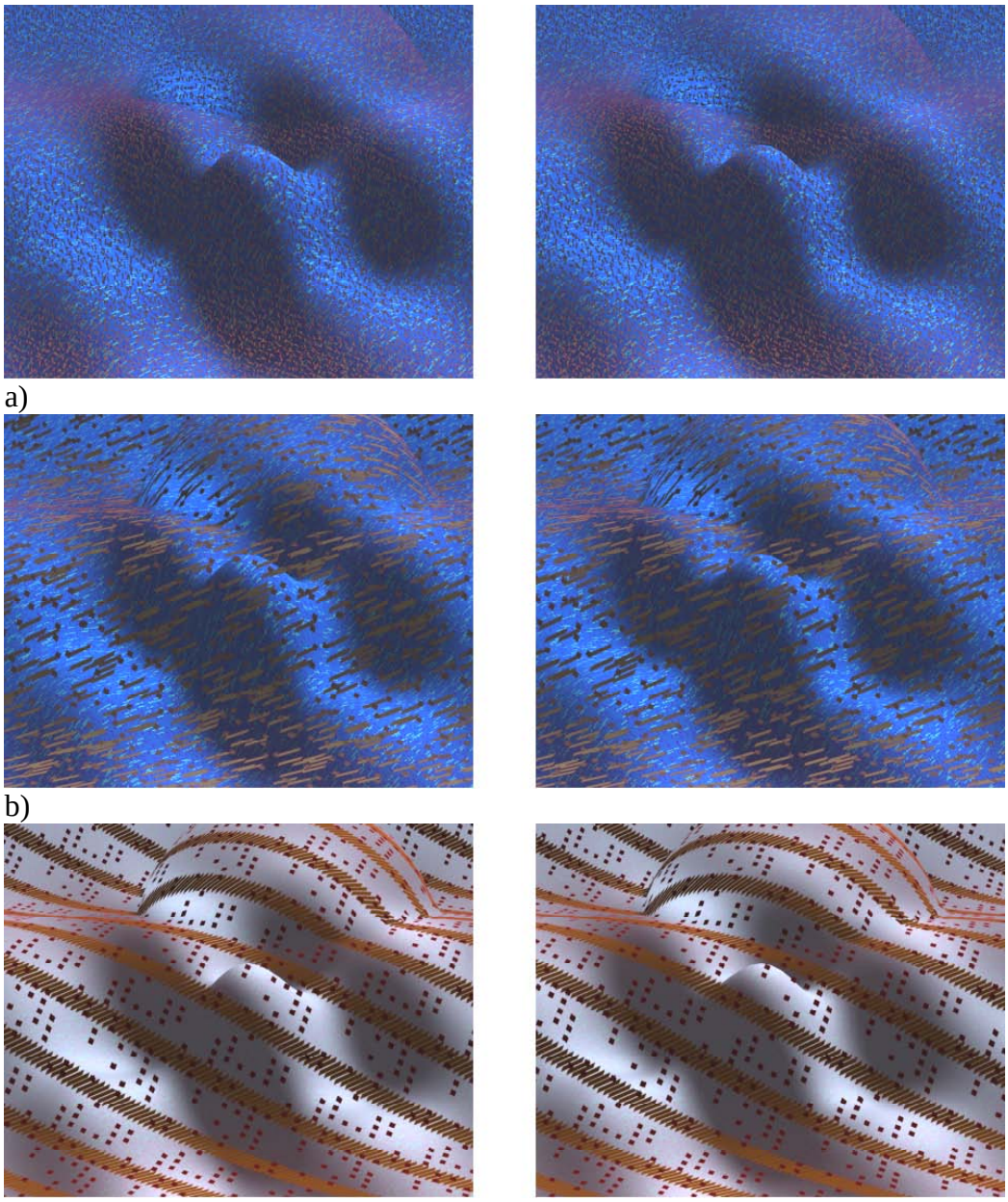


d)

*Figure 47. Continued.*

Next the rules are broken. Figure 48 a) shows the result when the top surface has a very fine grid similar to the bottom surface. Even though the same top area is covered, the fine texture on the top blends with the texture on the bottom and it becomes very difficult to see the shape of the top. Figure 48 b) increases the top translational randomness, making the top shape harder to pick out. Figure 48 c) flips the saturations of the top and bottom surfaces. With the bottom surface still at high value and now low saturation, it becomes nearly white, and it is hard to see the texture information. Figure 48 d) flips the top and bottom surface values, which makes the bottom surface too dark to easily see the texture or shading cues. Figure 48 e) increases the size of the bottom features and the bottom grid to be nearly the same as the top. This also amplifies the

bottom randomness to the point of being visually confusing. Lastly, Figure 48 f) flipped the randomness, size and grid characteristics of the two surfaces. This actually works reasonably well, except that the top surface is still difficult to see.

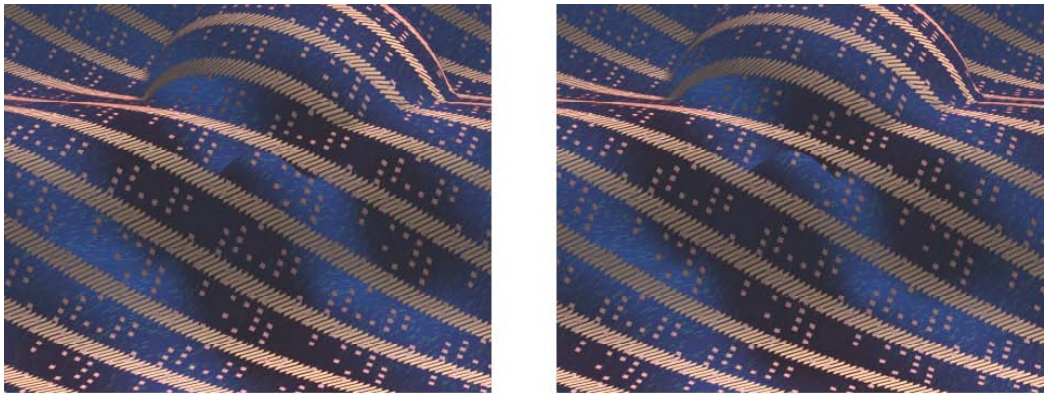


a)

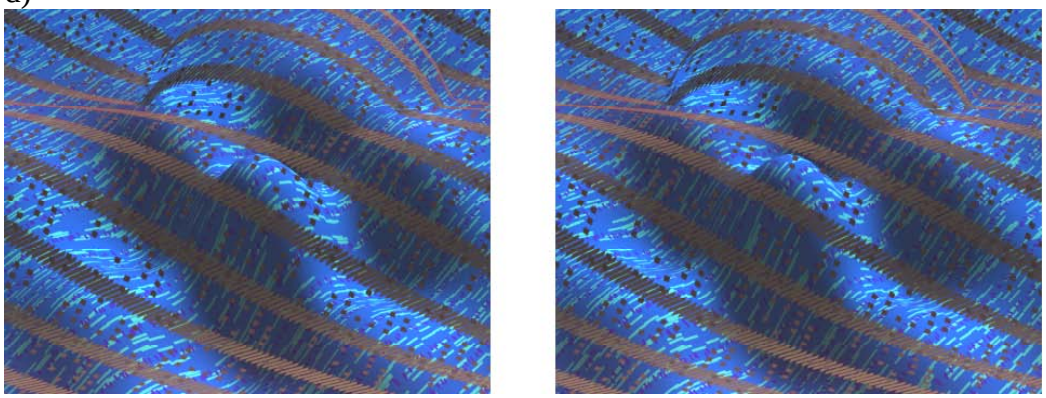
b)

c)

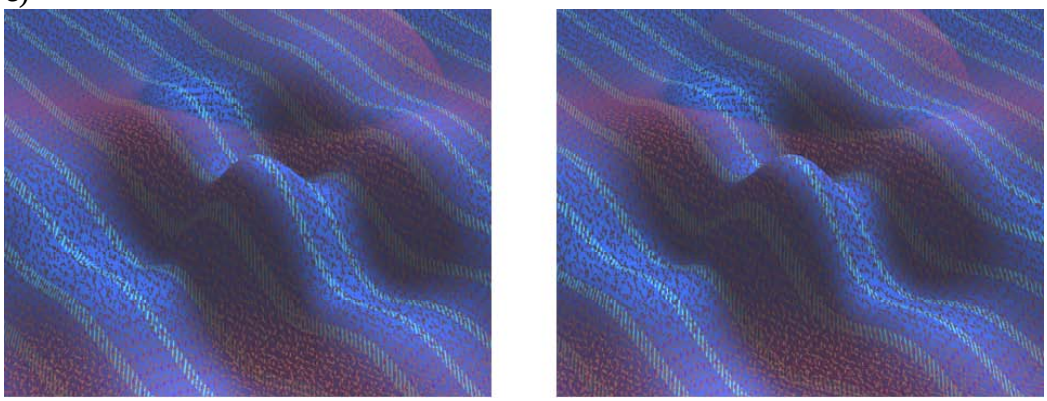
*Figure 48. Rule Breaking Textures. Examples of variations on a default texture that break specific guidelines.*



d)



e)



f)

*Figure 48. Continued.*

By subjective evaluation of the above images, it seems that the guidelines work reasonably well. The images that are supposed to work well seem superior to the ones

that break the guideline rules. This gives us confidence in our conclusions within the bounds of this experiment.

### 3.4 Discussion

The importance of several visual parameters became clear from the different analyses. Shape parameters, like grid and mark size and randomness were found to be more important than the color parameters. That is, the structure of the texture seems more important than the color. Within the color parameters, value is the most important, followed by saturation and hue matters very little. Again, this fits with perceptual research. Black and white photographs are interpreted easily given only a luminance channel, and color-blind people are often not aware of having a disability. Also, results suggested that only one or possibly two sets of marks might be necessary. More complicated textures might have the tendency to confuse rather than help shape perception.

Results suggested that marks and spacing should be bigger on the top surface. Although this means there are larger opaque areas to block view of the bottom surface, this also means that there are larger transparent areas to clearly see through the surface. However, for this experiment this result could also be because the top surface has less detail, and does not need fine-grained marks to show that detail. The top should also have low blur and randomness, meaning that it will appear more structured and precise. This fits with results that structured textures are better than random for shape perception, but it is interesting that this effect seemed much stronger for the top surface than the bottom. The bottom surface should be high in value, and a preferred difference in luminance between top and bottom or about 40% was found. Certainly, having high-value surfaces is important to let shading information play a role. A luminance difference might also help to distinguish the two surfaces, though blending of top and bottom surface colors can complicate the issue. Finally, the opacity coverage of the top surface seemed optimal between 40-50%, and reasonable within 30-60%. A near 50% coverage makes sense since this gives about equal amount of information from each

surface. However, the range of workable opacities was slightly surprising.

This experiment was designed as a pilot study, and so many possible improvements were discovered through the course of running and analyzing the results. In terms of experimental design, the surfaces were not randomized, and the task was very subjective, meaning the results were not necessarily valid for more general surfaces, or real-world tasks. In the next experiments, both of these issues are addressed. In terms of analysis, it was found that the extreme interactions among some variables (some were switches to turn others on and off) made interpretation very difficult, as did having a range of parameter levels from binary to continuous. Also, the need for three levels of marks seemed to be more complicated than was necessary to produce good textures.

To address these problems, a second multi-parameter experiment was designed. It includes randomized surfaces, a less subjective task and a more simplified parameterization. Several other improvements to the specific parameters and display were included and are described in Section 4.

Because of the design methodology used, several questions can be raised about the validity and statistical significance of the results. Naturally, the parameter space search was too sparse to be able to say definitively that all good areas were found. However, this is true of any study, and most do not even make an attempt to explore beyond the variation of a few parameters. This methodology does as good a job as possible, within the limits of time and human ability, of exploring the parameter space in an intelligent manner.

Lastly, the nature of the genetic algorithm used in both this study and the following multi-parameter study tends to bias the values of genes near important parameters on the genome, even when those genes might not be important themselves. This can cause false positives in the analysis. One good way to fix this problem would be to use a random chance for each gene of a parent to be passed on to a child. This may not simulate biology, but it does a better job of not biasing the final dataset. However, this issue was not identified until after both multi-parameter experiments had been run.

# 4. MULTI-PARAMETER EXPERMENT 2 *

## 4.1 Experimental Design

As discussed in the previous section, several conclusions were made during the process of running and analyzing the initial pilot experiment. Among these were a need for randomization, and a task that would not be overly biased by a particular texture's aesthetic appeal. Also, we felt that changes in the texture parameterizations were needed to simplify both analysis and the correlation of parameters and actual perceptual variables.

To address these issues, a follow-up experiment was designed. It used a feature finding task rather than a subjective assessment of shape perception. The task was to find bumps of different sizes on the surfaces. This simulates the tasks a scientist or researcher would perform when looking for important features in surface data. The textures were drastically simplified to use only one set of marks per surface, with a simplified color scheme.

Despite these changes, the human-in-the-loop methodology used was identical to that described in the previous section. It included a parameterization of layered, textured surface rendering, a parameter search technique, and various data analysis techniques. However, there were several significant differences in the specific implementation of the methodology. The overlapping surfaces were again displayed textured and shaded with stereo and motion cues. But instead of polarized stereo glasses, a Wheatstone stereoscope was used, as shown in Figure 49. This device used two mirrors to reflect the images of two high-resolution monitors independently to each eye. Vergence of the eyes makes the left and right monitor images overlap to produce an illusion of stereo. The focal distance was uniformly the distance to the monitor across the whole image. Simulation of different focal planes would require both eye-tracking and interactive

---

* The images and text in this section are reproduced in part with permission from previously published material. See [Bair et al. 2006].

depth-dependent blurring. Because of these difficulties, depth of focus was not simulated. The motion cues given were a gentle rocking of the surface from side to side, with a rotational magnitude of 20° and a period of about 2 seconds. The lighting model used a standard Lambertian shader with a 20% white ambient term. The diffuse shading term was from a white point light above and to the right of the surfaces.

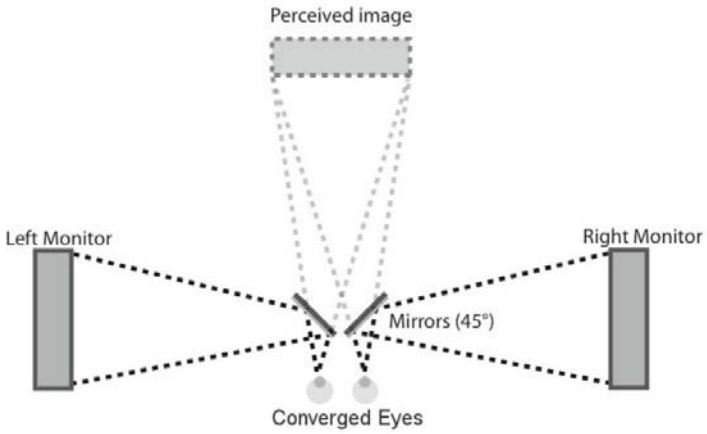



*Figure 49. Wheatstone Stereoscope Diagram. Separate images are presented to each eye to give the illusion of depth.*

The images were displayed on two IBM T221 LCD monitors. The monitor screens were 48 cmwide and 30 cm high (16:10), with a screen resolution of 3840 x 2400, or 9.2 million pixels. The screens were set at a viewing distance of 109 cm, about one and a half arms-length from the eye. This yielded a visual angle per pixel of approx 23 seconds of arc. This is comparable to the size of receptors in the fovea and is sufficient to display the finest grating pattern that can be resolved by the human eye – about 1 cycle per minute of visual angle [Campbell and Green 1965]. According to the manufacturer, the screen contrast ratio is 400:1, and maximum brightness is 235 cd/m$_2$. The stereo images were rendered assuming an eye separation of 6.4 cm. Four computers were used to produce the images, two for each screen. The computers were networked to insure that the motion of the left and right images were in sync. However, machines are

now available that can support hardware-syncing of multiple large monitors so a single machine could perform the same task. An image of a subject viewing a surface through the Wheatstone stereo setup is shown in Figure 50.

*Figure 50. Stereoscope Photograph. A subject using the Wheatstone stereo design.*

The surfaces were designed to mimic shapes that might be found in a real application using layered surfaces. The two critical factors considered were noise and features. Noise could be either an artifact from data acquisition, or simply part of the surface shape that is not important. For example, a doctor looking for a tumor is interested in features that are different from the normal appearance of an organ.
Features are the shapes of interest; a tumor in this example.

Each surface was generated as a height field on an (x, y) Cartesian plane where the height was calculated as the sum of the noise and features. Figure 51 shows example top and bottom surfaces. The noise was computed as the product of the sum of eight sinusoids that vary in the x direction with the sum of eight sinusoids that vary in the y direction. The longest period sinusoid had about 12 degrees of visual angle (two cycles

per display width on our display), and the remaining sinusoids had periods dropping off with the ratio 0.6. Thus, the shortest period sinusoid was about 0.3 degrees of visual angle (72 cycles per display width). Each sinusoid was given a random phase angle between $\pi/2$ and $-\pi/2$, and amplitudes were scaled linearly with the period but with a 20% random variation.

The features used were Gaussian bumps superimposed on the noise. Seven Gaussian bumps of various sizes were arranged in a jittered pattern on this plane, ranging in standard deviation from about 2 degrees of visual angle (1/12th of the full display width), to 1/2 degree of visual angle (1/48th the width). The ratio of bump sizes was 0.8. The largest bump was similar in height to the product of the largest sinusoid amplitudes, with bump heights falling off linearly with standard deviation. This means that the bump height was similar in range to the noise, so the height alone would not give away the location of a bump. Instead, subjects had to rely on seeing the radial symmetry of the bumps against the anisotropic, noisy background. Note that the surfaces shown in Figure 51 have been cropped and not all seven bumps are visible.

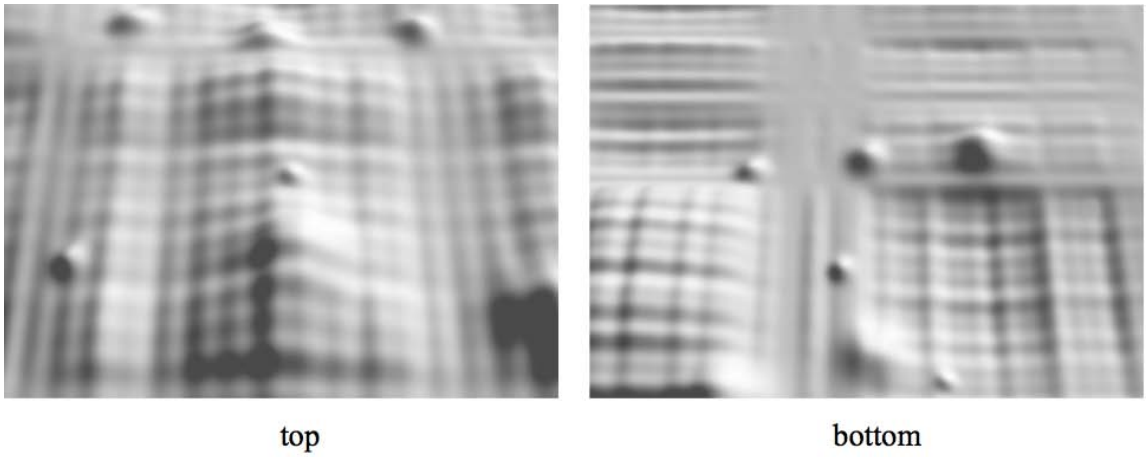



*Figure 51. Top and Bottom Surface Examples. [Bair et al. 2006].*

The texture parameterization had a similar structure to the pilot experiment parameterization, but was made much simpler. The goal was to keep the

parameterization general enough to be able to create a wide space of different patterns, but simple enough to make analysis possible. Again, the textures were tiled, draped textures. But to simplify analysis, all parameters were defined to have seven discrete levels ranging from 0 to 6. Each texture tile was a 512x512 image that covered about 3 degrees of visual angle, tiling 8x8 times across the surface.

The texture synthesis approach was considerably simplified from the previous experiment. First, only one set of marks was used instead of three. This was because the previous experimental results suggested that only one or two sets of marks were truly useful. And using only one set of marks drastically simplifies analysis. Also, the background color was limited to gray values (black through white) because the hue and saturation seemed to matter much less than the value. The overall texture could rotate within the range from 0° to 90° in steps of 12.8°, and the top surface texture background ranges from fully opaque to fully transparent (the bottom surface background is always opaque).

The marks themselves used parameter variables similar to the previous experiment. Marks were drawn within a grid, where the grid had variable number of columns and rows. The marks also had size and aspect ratio parameters that sized them within the grid cell. Figure 52 shows how these four parameters interact to affect mark shape and placement. Grid row and column parameters multiply to produce the number of grid cells. Figure 52 a) shows what a rectangular mark of medium size looks like in grids of 1, $2^2$, $8^2$, and $32^2$ cells. Note that there are actually 7 allowable levels {1, 2, 4, 8, 16, 32, 64} for number of rows and columns, but only 4 are shown. Figure 52 b) shows various grid settings of 2x32, 4x16, 16x4 and 32x2 with a fixed mid-sized mark. Since rows and columns can vary independently, anisotropic grids like this are common.

Mark size ranges from 1/8th of the grid cell size at the lowest level to 7/8th of grid cell size at the highest level. Figure 52 c) shows the effect of the mark size parameter in a grid with a high number of columns and a low number of rows. It behaves as expected, making the mark smaller relative to the overall grid cell shape. Mark aspect ratio ranges from 1:1, conforming to the grid cell shape, to 1:7, which vertically shortens

the mark. Figure 52 d) shows the effect of the extremes of mark aspect ratio in a grid with a high number of columns and a medium number of rows, and Figure 52 e) shows the effect in a grid with a medium number of columns and a high number of rows. For textures that have equal or greater number of rows than columns, aspect ratio works as expected, however, for the case of more columns than rows in figure 52 d), aspect ratio has an opposite to normal effect. Finding a way to modulate aspect ratio independently from the grid shape is difficult. This method does vary the aspect ratio, but not always as expected. However, if the mark size and aspect ratio were to be separated from the grid shape, then overlapping of marks would be a big problem. That is, marks might be far to large or far too small to show up on a given grid well. For this reason, this use of the aspect ratio parameter was thought to be best.

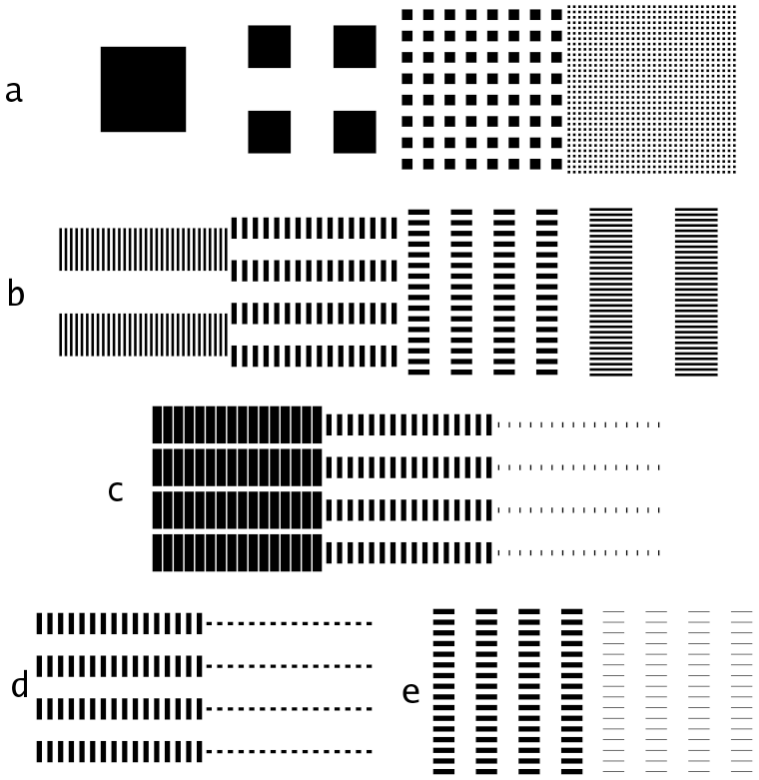



*Figure 52. Grid Mark Shape Parameterization. Effects of shape parameters on mark appearance. a) equal rows, columns b) unequal rows and columns c) mark size d,e) mark aspect ratio. [Bair et al. 2006].*

Marks drawn on the background use the color space shown in the row labeled hue in Figure 53, that interpolates in RGB space from orange through white to cyan at the the highest level. For space reasons, only 5 levels are shown, though 7 were used for the experiment. This limited orange-cyan color space provides the ability for marks across the top and bottom surfaces to be of complimentary colors, but provides a linear, rather than a circular, hue scale. As shown in the row labeled round in Figure 53, marks are superquadric ellipses that range from rectangular to round. Mark opacity ranges from fully opaque to fully transparent, replacing background opacity on the top layer wherever a mark appears, so that top textures may range from being opaque with transparent marks to transparent with opaque marks. A single regularity parameter controls regularity in three ways: jitter with respect to grid cell centers, rotation of marks within a grid cell (from –45° to 45°), and the probability that a mark will be drawn in a grid cell. The net effect of this parameter is shown in the row labeled regular in Figure 53. Lastly, a Gaussian blur filter (not shown) is applied to the texture after it is constructed and a parameter varies the size of the filter kernel from 1 (no blur) to 25 pixels.

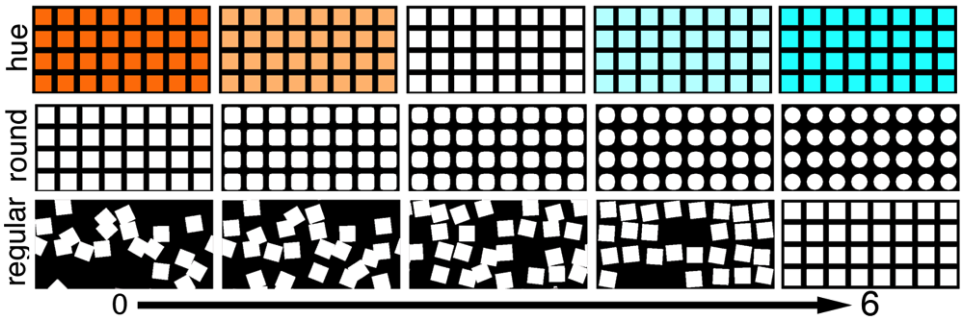



*Figure 53. Color, Roundness and Randomness Parameterization. Effects of hue, roundness and regularity feature parameters. [Bair et al. 2006].*

Instead of using fixed surfaces as in the last experiment, new surfaces were constructed for each presentation so that learning the shape of a specific surface would

not bias perception of texture quality. As in the pilot experiment, the texture parameter space was searched using a genetic algorithm. This meant that each subject started with an initial set of 40 randomly-generated textures, and each successive generation was built from breeding the highly-rated textures from the previous generation using a two-point crossover approach.

One change was made, however, to include 'islanding' with the genetic algorithm. Once the subjects had completed at least two generations, they were told to pick an especially good texture, and indicate it by pressing a key on the keyboard. The islanding algorithm would then create an entire generation of textures whose parameters were each drawn from Gaussian distributions centered around the selected texture. The Gaussians had a standard deviation of 0.7, meaning that most parameters varied by about 1 one a 0-6 scale. Islanding was chosen as an alternative to running the genetic algorithm until convergence. It provided a way for the subject to bring the experimental trial to an end by finely searching an area of texture space known to be good. The pilot multi-parameter experiment took about three hours per trial, whereas with islanding a trial took less than one hour. The islanding was a way for the subjects to use their intelligence in picking an excellent texture. Also, by directly searching the space near a good texture, more fine-grained gradients could be found that show how changing different variables affects good textures.

Finally, the task was made to be less subjective than in the pilot experiment. The subjects were asked to rate their ability to locate all seven bumps first on the bottom surface and then on the top surface. Ratings were from 0 to 9, requiring only a two numeric key presses on a standard keyboard. After the two ratings were entered, the next presentation was made and the process repeated. Subjects were not told how to map what they saw to the numeric values, however some commented they had developed a strategy such as counting the number of bumps they could easily see. A total of six subjects completed 22 total trials, resulting in 4560 rated texture pairs. Subjects were all familiar with computer graphics applications, and were considered experts for the purpose of this study. All subjects had normal or corrected-to-normal visual acuity, as

well as normal color and stereovision (tested using random-dot stereograms). Subjects were given brief training, involving showing them a relatively good set of textures chosen by the experimenters that showed the bumps and noise well enough to be clear.

## 4.2 Analysis of Data

A variety of new statistical techniques used tried for this data set. Again, because of the non-uniform nature of the parameter space search, many standard techniques could not guarantee statistical significance because they assume a uniform sampling. Therefore, several analysis methods were used to triangulate results. These were ANOVA, LDA, decision trees and Parallel Coordinates analysis. Analysis of variance (ANOVA) was a convenient way to look for individual parameters whose settings are critical. Linear Discriminant Analysis (LDA) provided optimal linear classifiers for the data. Decision Trees improved on this, providing non-linear classifiers built as logical expressions. Finally, Parallel Coordinates was a convenient way to visually assess hypotheses about relationships among parameters.

### 4.2.1 ANOVA

Analysis of Variance (ANOVA) is a standard statistical technique to measure if variation in an independent variable can be used to explain variation in a dependent variable. ANOVA compares the variance due to a given parameter with the total variance. This gives a p-value, or probability that each variable is a significant factor in determining the rating. In this case, the independent variables are the texture parameters, and the dependent variable is the measure of texture quality, i.e. the two values given by the subject to rate the top and bottom surfaces. ANOVA is a good way to test for importance of individual parameters that have a strong effect without interacting with other parameters. ANOVA can also test for parameter interactions, but is was not used for that purpose here because better techniques exist for learning that type of information in large parameter spaces.

At the 1% significance level, all parameters were found to be significant except bottom mark chroma and top mark aspect-ratio. The five parameters with the largest

effect sizes were top opacity, bottom luminance, top luminance, top mark luminance and top number of columns. That is, the most important factors seemed to be overall opacity, brightness and grid structure.

To get an idea of optimal levels for the significant parameters, box plots were drawn of the ratings for each level of the variables. In each box plot, the red line is the median and the blue box denotes lower and upper quartiles. Whiskers show the 1.5 inter-quartile range of the data and + indicates outliers beyond the whiskers. Notches in the blue box give the possible range of the median. Since the subjects gave separate ratings for the top and bottom surfaces, ANOVA could be run using either the top or bottom ratings or a combination of them. This allowed a set of guidelines to be build for overall quality, or for maximum visibility of either the top or bottom surface. Below, we show several examples of the box plots to give an idea of how the guidelines were built.

In Figure 54, the top and bottom surface rating distributions are plotted for the different top opacity levels. A clear pattern is visible. The top ratings are bad if the top is transparent, while the bottom ratings are bad if the top is opaque. Note that these graphs do not take into account the opacity of the marks, but the results make sense. One oddity is the extremely low top ratings for some of the fully opaque top surfaces. Although the low ratings could certainly be due to any of the other parameters, the next highest opacity does not have the same spread in distribution. Therefore, we hypothesize that the subjects might not have been able to see the bottom surface through the top, thus mistaking the actual top surface for the bottom. They might then have rated the top very low, as being invisible. Certainly, the bottom quartile of the top ratings distribution overlaps well with the top quartile of the bottom ratings distribution. The top surface was displayed with a noticeable red mark to distinguish the two, but this might not have been sufficient.

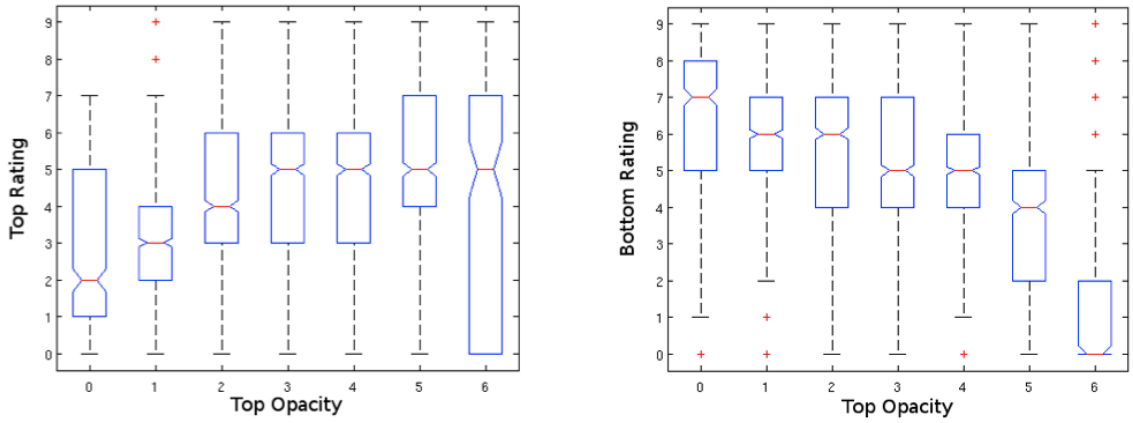



*Figure 54. Top and Bottom Opacity Ratings Box Plots. The top opacity has inverse effects on the top and bottom surface ratings.*

Next, the top opacity is plotted against the average rating, shown in Figure 55. The effect on each surface combines to make 3, 4 and 5 the best opacity levels overall. The effect of opacity on top and bottom is nearly reversed, so middle opacities seem to be best. It can be noted here that the fully opaque (6) top surface opacity distribution is no larger than any of the others. This supports the theory that there might have been some mistaken surface identities, since the average rating variance is consistent.

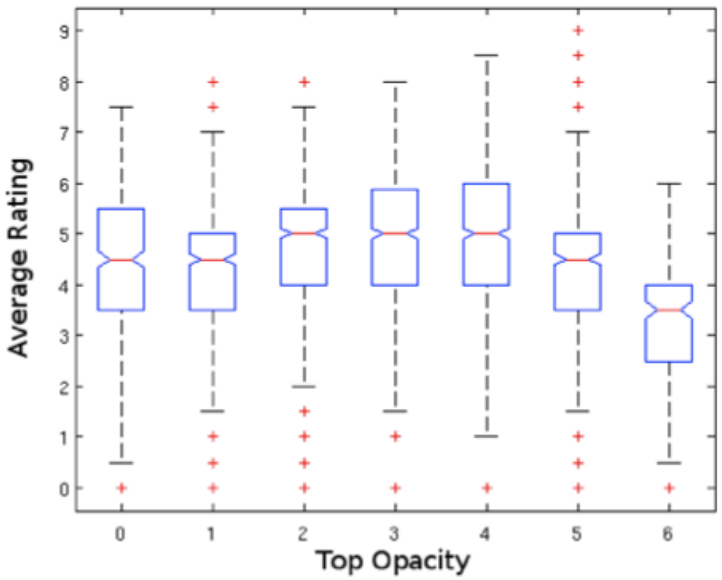



*Figure 55. Average Opacity Box Plot. Average ratings distributions for different levels of opacity. [Bair et al. 2006].*

Another point to be made is that although the intent was to allow opaque backgrounds with transparent marks letting subjects see through the marks to the bottom surface, this was not likely because of the parameterization. Mark size linearly varied the area covered by the mark, but several other parameters reduced the overall mark size. Aspect ratio narrowed the marks, roundness removed the corners, and randomization simply did not draw some of the marks. As a result, textures with marks big enough to see through were quite rare, and it is not surprising that the background opacity had a strong effect.

Top and bottom luminance were the second most important factors in determining surface ratings, so we show box plots for them as well. The top ratings plot in Figure 56 shows that dark texture backgrounds work very poorly, and the trend seems to prefer textures that are as bright as possible. No strong effect was found for the top luminance on the bottom rating, so the average rating increases less dramatically with top luminance.

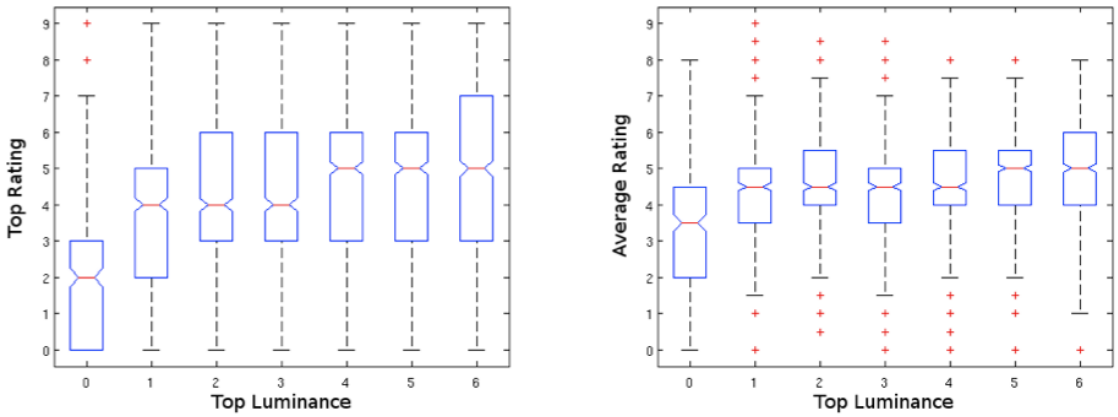



*Figure 56. Top Luminance Box Plots. Dark top textures hurt the top surface ratings.*

Similarly, the bottom luminance affects the bottom surface rating. Figure 57 again shows a trend for poor bottom surface ratings with dark bottom textures. Also, the bottom surface luminance did not seem to strongly affect the top surface ratings, so the average ratings are affected less strongly.

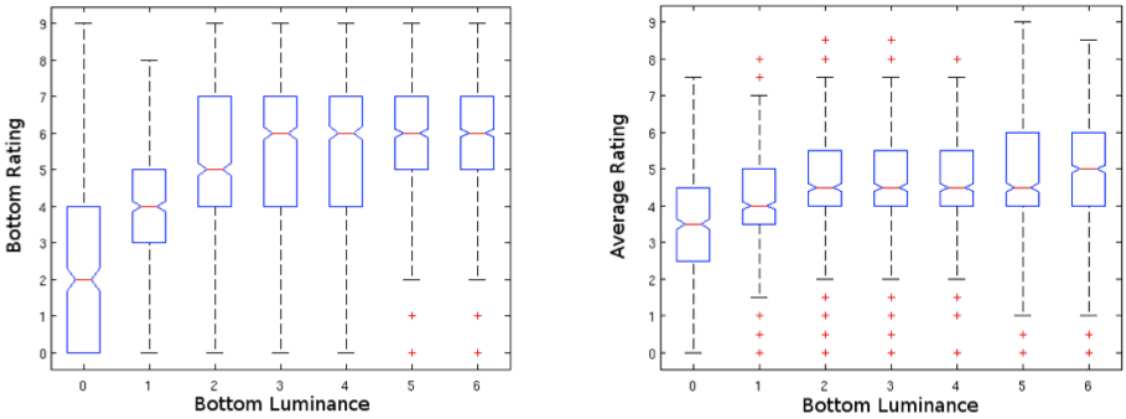



*Figure 57. Bottom Luminance Box Plots. Dark bottom textures hurt bottom surface error.*

The guidelines based on the above analysis are as follows. To see the top surface clearly, the top should have high luminance. To see the bottom surface clearly, the

bottom surface should have high luminance. Since luminance of one surface did not seem to have a strong effect on the rating of the other surface, to see both surfaces clearly, both should have high luminance, with the bottom slightly brighter than the top. Middle-range opacities should be used. Results for luminance of the marks were similar to the surface backgrounds, so marks should be bright overall, except the bottom surface also does well with dark marks. Both surfaces should have relatively opaque marks. The top should have many columns and a medium number of rows, without high blur. The bottom should have many columns, with high aspect ratio marks. Because of the interaction between grid aspect ratio and mark aspect ratio, the bottom marks are likely to be squarer and quite small.

Although the above guidelines seem useful, the down-side to using ANOVA analysis is that it only accounts for the effects of single parameters, ignoring parameter interactions. Interactions could be accounted for by including special interaction terms, but with 26 parameters, accounting for all possible combinations is intractable. For this reason, the next analysis method used was one that deals with combinations of parameters.

### 4.2.2 LDA

Fisher's Linear Discriminant Analysis (LDA) is a method that finds linear vectors that maximally separate classes of data. Data points (textures) are thought of as vectors in a multi-dimensional space, where each dimension corresponds to a parameter. Each data point is classified, for example as good or bad. LDA finds directions in the parameter space that provide the best separation between the classes.

This is done by maximizing the ratio of between-class to within-class variability. For a number of classes C, this can be formulated in terms of an eigenvalue problem, which returns C-1 eigenvectors and eigenvalues. The eigenvectors are direction vectors in parameter space, and projecting along them provides maximum separability. The eigenvalues give a measurement of the amount of separation that is achieved along corresponding eigenvectors, essentially a ratio of between-class to within-class

separability. Components of the separating vector can be analyzed to find importance of parameters and high/low level preferences when the parameters are used together in linear combination. Therefore, LDA finds the importance of parameters and whether parameters should be high or low when linear parameter interactions are included.

In this case the classes are different groups of rated textures. Three classes representing good, medium and poor textures were made by thresholding the top and bottom ratings. The thresholds were chosen to make the number of samples per class equal. Good have both top and bottom surface ratings of 7 or greater. Of the remaining textures, middle have both ratings of 4 or greater, and poor are all the remaining textures. The parameter vectors are directions in the 26-dimensional texture space.

The first LDA eigenvalue was 0.27 and the second only 0.019. Since the effect of the second eigenvector was so low, we only analyzed the first eigenvector. Figure 58 shows the data projected onto the first two eigenvectors. Here, red dots represent good textures, green medium textures, and blue poor textures. It is clear that the data is not fully separable using only LDA, and because of the extent of the overlap it seems likely that linear combinations of the variables cannot solve the layered surface texturing problem. But important information can be learned from the first eigenvector. Comparing the components of the vector with good and bad data projections can tell us what parameters need to be either high or low, or do not matter much. In this case, if the parameter's component is strongly positive than it should have a high level to create good textures, if it is strongly negative then it should have a low level, and if it is near zero then the parameter's level does not strongly affect the rating.

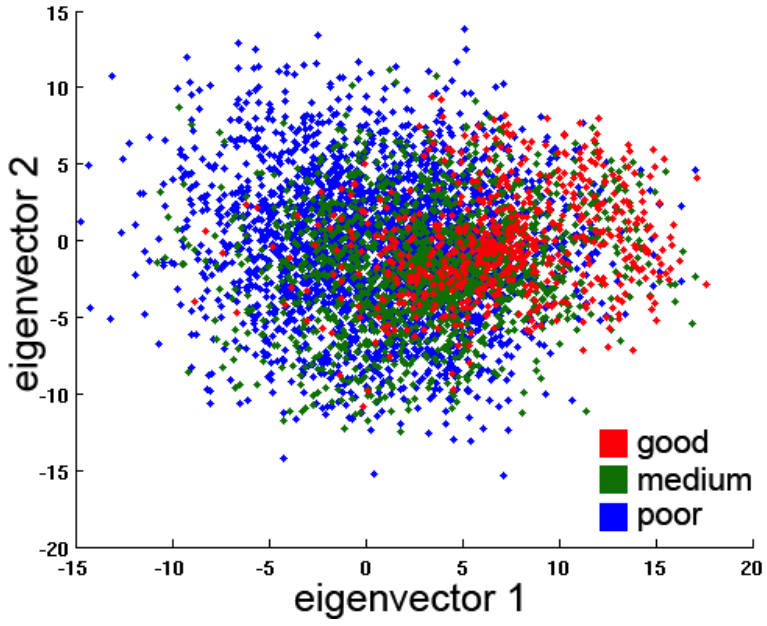



*Figure 58. LDA Separation of Texture Classes. Scatter plot of texture ratings projected along the first two LDA eigenvectors. [Bair et al. 2006].*

The components of the first eigenvector are shown in Table 1. According to the components, to create good visualizations the top should have many columns and moderately high rows, high luminance, relatively high opacity and little blur. The top marks should be medium-sized and irregular, very bright, orange, with a low aspect ratio. The bottom should have many columns and moderately high rows, high luminance, and medium filtering. The bottom marks should be small, with medium opacity and high aspect ratio. Because of the interaction between rows, columns and aspect ratio on top and bottom, the top marks will be long and thin, and the bottom marks will be small and relatively square.

*Table 1. LDA Coefficients. [Bair et al. 2006].*

| **parameter** | **top** | **bottom** |
|---|---|---|
| number of rows | 0.26 | 0.20 |
| number of columns | 1.00 | 0.60 |
| luminance | 0.83 | 0.80 |
| opacity | 0.59 | -0.09 |
| rotation | 0.10 | -0.10 |
| filter-width | -0.60 | -0.20 |
| mark luminance | 0.80 | -0.03 |
| mark chroma | -0.50 | -0.06 |
| mark opacity | 0.21 | 0.37 |
| mark roundness | -0.28 | -0.05 |
| mark aspect ratio | -0.52 | 0.72 |
| mark scale | 0.31 | -0.58 |
| regularity | -0.40 | -0.12 |

Figure 59 shows sample tiles of the top and bottom textures hand-built according to the above criteria. The top texture is shown composited over a checkerboard to show the transparency. The enhanced version of the bottom is shown to clarify that there is a very subtle texture on the bottom, which may prove useful for stereovision. In general the results correspond with those of the ANOVA analysis, though here the top luminance has a slightly higher component than the bottom.

top bottom bottom - enhanced

*Figure 59. LDA Rule-Based Textures. Texture tiles constructed based on LDA analysis. The enhanced version is included to show the more subtle texture on the bottom. [Bair et al. 2006].*

Assuming that the parameter interactions are linear is helpful for getting at basic patterns, but the truth is likely to be much more complicated. As an obvious example, if both the background and the mark color are the same, then the marks cannot be seen at all, and are useless. But the color might be perfectly fine to use on either the background or marks when paired with a different color. Therefore, the next method used for analysis takes into account non-linear combinations of parameters.

### 4.2.3 Decision Trees

Decision trees are a non-linear method for mapping parameter levels to a dependent variable such as rating. Once trained on a set of data, they generate a simple list of rules for what parameter levels give a certain texture quality. This is extremely useful for interpretation and logical analysis. Decision trees are able to model disjunctions of conjunctions, going beyond the limitations of linear models like LDA. For example, Figure 60 shows the statement (A&B) | (C&D) in decision tree form. Every branch can be thought of as an entry in a truth table, and leaves labeled + denote cases that satisfy the expression, while – denotes cases that do not. Each case starts at the top node and travels down the tree where it chooses a branch based on whether it meets the criterion. When it gets to a leaf node it is classified by the value of the leaf node. An example of a simple texture rule that could be put in the form of the decision tree below

is as follows. A texture is good if the top marks are opaque and bright, or if the top background is bright and semi-transparent. Note, this statement is just an example; the actual decision tree results are more complicated.

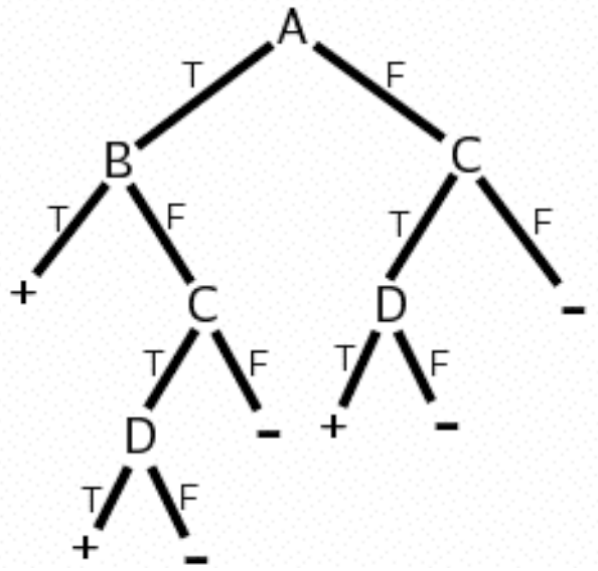



*Figure 60. Decision Tree Example. This tree models the truth statement (A and B) or (C and D). Leaf nodes marked + are in the set, those marked – are not. [Bair et al. 2006].*

Decision trees can also be used for variables with more levels than simply true or false. In our case, each parameter can have levels between 0 and 6, and the ratings can be between 0 and 9. As a result, the leaf nodes have classification values that are between 0 and 9 instead of a boolean true or false. To model parameters with 7 levels but maintain a tree branching factor of 2, the nodes used a cutoff threshold, where the left branch holds data below the threshold, and the right branch holds the data above. For example, one branch might be for top surface opacities less-than-or-equal-to 5, and the other branch would be fully opaque (opacity = 6). The decision thresholds were chosen by maximizing information gain at each node. This method generates a tree that tries to simplify the rules rather than having a separate branch for every parameter level.

The rating of each leaf node is the mean of the combined texture ratings for the data in that leaf. This allow rules to be found for nodes with a given predicted rating by tracing the parameter thresholds from the leaf back to the root of the tree. For each leaf, this gives rules in the form of conjunctions. In theory, if several leaves had the same rating, rules could be formed that were disjunctions of conjuctions ((A&B) | (C&D)). This was not done in this analysis because the decision tree was very complicated, and analyzing combinations of nodes was too difficult.

A large tree can easily over-fit a dataset, modeling noise rather than actual rules. Therefore our tree was pruned to a depth that gave a minimum error when tested using cross-validation. This tree had 153 leaf nodes, and is too big to be effectively shown here. The leaf with the highest predicted rating was 8.06, with a standard deviation of 0.77. Its rules were found by following each decision from the root node to the leaf node; these are shown in Table 2.

*Table 2. Decision Tree Best Leaf. Parameter ranges of best leaf in decision tree. [Bair et al. 2006].*

| **Parameter** | **Range** | **Interpretation** |
|---|---|---|
| Top rows | 0-4 | 1 to 16 rows in texture grid |
| Top columns | 5-6 | 32 to 64 columns in texture grid |
| Top background luminance | 1-2 | 16% to 33% of full luminance |
| Top opacity | 2-5 | 33% to 83% opacity of background |
| Top mark luminance | 3-6 | 50% to 100% of full luminance |
| Bottom columns | 5-6 | 32 to 64 columns in texture grid |
| Bottom background luminance | 2-6 | 33% to 100% of full luminance |
| Bottom mark luminance | 4-6 | 66% to 100% of full luminance |

The range of columns and rows on the top surface does not overlap, so the grid is guaranteed to have at least a 1-2 aspect ratio. This will create a linear structure across the top surface. Background luminance is clearly very important. The bottom surface should

have high luminance, while the top surface has a trade-off between brightness and opacity, giving it a medium apparent brightness. Both top and bottom should have high numbers of columns, tending to give them long, thin marks with a global striping structure. Marks on the top should be prominent - large, bright, opaque, with low aspectratio. Marks on the bottom should be less prominent – small, less opaque, with high aspect-ratio.

By contrast, another leaf that had a rating of 5.47 had the rules shown in Table 3. Although many nodes had similar ratings, it was chosen because it had a short list of rules, and had a relatively large set of textures in it. Therefore it was thought to be more generally applicable. The top rows and columns are different from the best leaf node. There are more rows, and columns are not specified. Also, the top background is much brighter, while the top marks are slightly darker. The bottom luminance is identical to the best leaf. The aspect is set to be high, but it is difficult to know what this means because number of rows and columns are not specified.

*Table 3. Decision Tree Good Leaf. Parameter ranges of a good leaf.*

| **Parameter** | **Range** | **Interpretation** |
|---|---|---|
| Top rows | 3-5 | 8 to 32 rows in texture grid |
| Top background luminance | 3-6 | 50% to 100% of full luminance |
| Top background opacity | 2-5 | 33% to 83% opacity of background |
| Top mark luminance | 0-5 | 0% to 83% opacity of background |
| Bottom background luminance | 2-6 | 33% to 100% of full luminance |
| Bottom mark luminance | 4-6 | 66% to 100% of full luminance |
| Bottom mark aspect | 3-6 | 1/16 to 1/64 aspect ratios |

The decision tree took another step forward in learning about the interactions of parameters. For example, even though the ANOVA and LDA results showed that both top background luminance and top mark luminance should be bright, the results of the

decision tree suggest that perhaps they do not both have to be very bright. The best node had bright marks and a dark background, while the mediumrated node had a bright background and slightly darker marks. Yet both cases had at least one feature with at least 50% brightness. Also, perhaps the relationship of rows to columns seen in the best leaf node might an important interaction as well. Intuitively, the large scale linear features created by an uneven grid seem useful. Still, this connection is tentative and not truly supported by the data so far. Also, no strong guidelines have yet been found yet for several of the parameters, such as mark size. To delve deeper into parameter interaction not yet understood, the last analysis technique chosen was parallel coordinate analysis (PCA).

### 4.2.4 PCA

Parallel coordinates analysis (PCA) is a graphical method for data exploration. This is a good method for finding patterns in data visually rather than looking at significance levels from statistical tests. It works well even with highly non-linear interactions, something that is hard to do with nongraphical methods. Unfortunately even though humans are visual creatures, it can be hard to see patterns when looking at many variables at a time, so parallel coordinates works best when the analyst has some hypotheses about parameters before starting data exploration. This is why it was used after the other techniques had already been used to find both the most important parameters, and probable good levels for them.

Parallel coordinates draws lines for each data point from the value of one parameter to the value of another parameter. Trends in the lines can illuminate patterns in the data. Since our parameter levels are discrete, to keep lines from sitting on top of each other, the parameter values were randomly jittered by (-0.5, 0.5). Choice of parameters and their ordering were chosen based on hypotheses about the data. For example, Figure 61 shows trends for good data, where good data is defined as having both top and bottom ratings of 7 or greater (out of 9). The line colors denote groups based on the opacity of the top surface background. This color grouping illuminates

trends between background opacity and other top surface parameters, like number of top rows and columns, mark scale and aspect ratio.

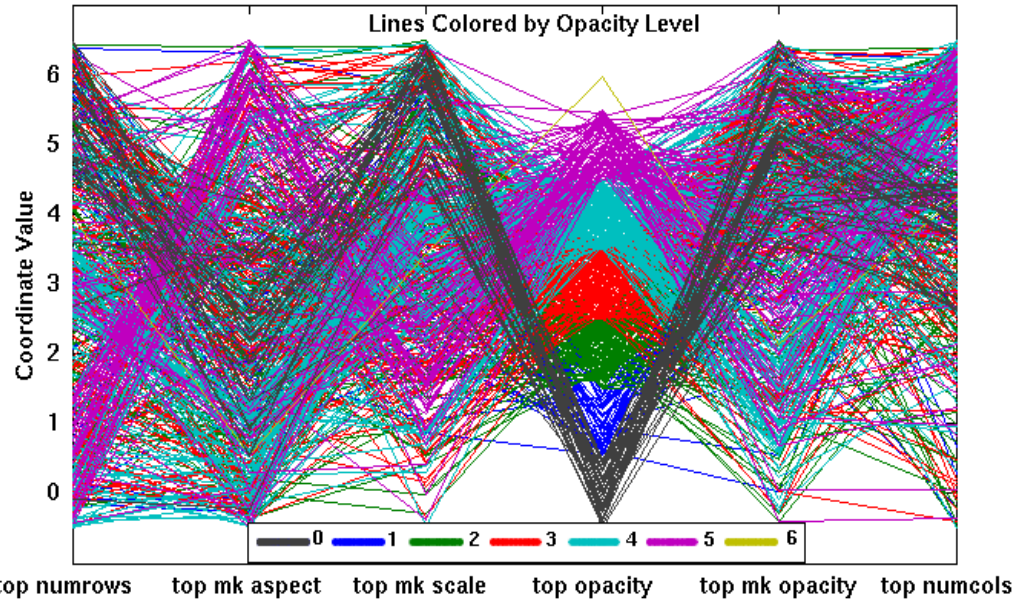



*Figure 61. Parallel Coordinates Plot of Opacity and Shape. Parallel coordinates plot shows strong patterns among five top surface shape and opacity parameters. [Bair et al. 2006].*

Two clear patterns can be seen for textures with very high and very low opacity. The magenta lines represent textures that have almost fully opaque backgrounds, and the gray lines those with fully transparent backgrounds. Since only one fully opaque background even made it into the good texture group, we ignore it here. The high opacity background textures have small, medium-high opacity marks. The grid has an aspect ratio of nearly 1/64, which combined with a high aspect ratio, makes for small square marks grouped in long rows across the surface. In contrast, the low opacity background textures have large opaque marks. The grids are closer to square, having more of a 1/4 aspect ratio, and since there were more rows than columns, the medium aspect ratio makes for long, thin marks.

By contrast, Figure 62 shows little discernible pattern between the mark chroma of top and bottom surfaces. Although two-variable patterns could just as easily be seen with a scatter plot, we wished to show a parallel coordinates plot that lacks a strong pattern. There is a slight trend for blue gray top marks (light blue lines) to correspond to orange bottom marks. However these patterns are not very strong; most of the levels for top chroma have an even spread of values for the bottom chroma levels, indicating that chroma interaction is not important. This is not terribly surprising considering that chroma was not found to be important in any of the other analysis techniques.

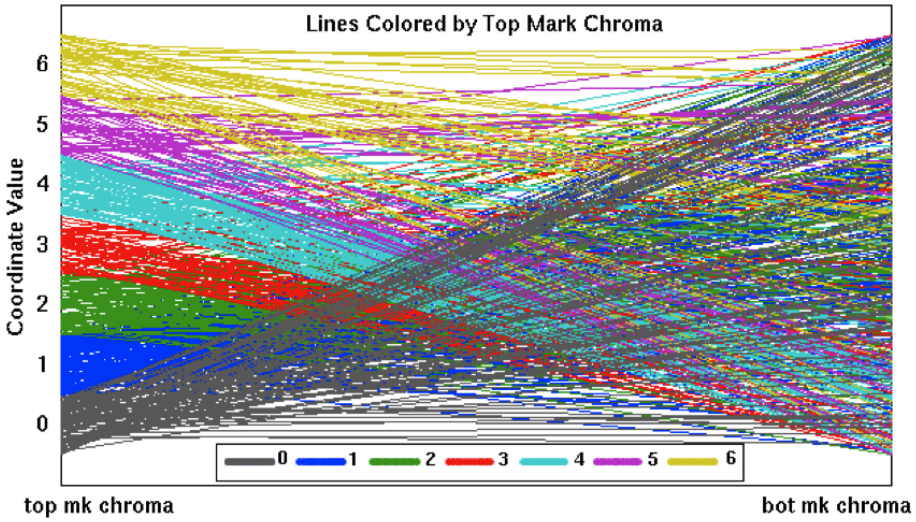



*Figure 62. Parallel Coordinates Plot of Hue. Hardly any patterns are noticeable between the top mark chroma and the bottom mark chroma. [Bair et al. 2006].*

### 4.3 Validation of Results

The guidelines found from the ANOVA, LDA, Decision Trees and PCA analysis seem reasonable. However, they need to be tested more thoroughly for several reasons. Because of the of the genetic algorithm non-uniform search of the parameter space, the results are uncertain. Also, there were no duplicate ratings within or between subjects. It is quite likely that subjects used different subjective ways to map how well they saw a

particular set of surfaces with the 0-9 rating scale. Therefore, a second experiment was designed to test how well different subjects would rate similar textures.

Sixty texture pairs were automatically generated, where the parameters were randomly set to be within the acceptable ranges for the hypotheses. Twenty-nine of these were generated according to decision tree rules. Four were made to be ‘bad’ using the rules for a leaf with a mean rating of 1.15. Five were made to be ‘poor’ using rules for a leaf with a mean rating of 4.57. Ten were made to be ‘fair’ using rules for a leaf with a mean rating of 5.47. Finally, ten were made ‘good’ using the rules for the best leaf node with a mean rating of 8.06. The other thirty-one textures were created by narrowing the decision tree rules using the good variable ranges found from the other analysis methods. Twenty (enhanced-A) were created to have a medium opacity top surface with medium-sized marks, and eleven (enhanced-B) were created to have a low opacity top surface with larger marks. The number of textures created for each set of rules was chosen in proportion to the expected quality. The limit of sixty visualizations was chosen as the total number of visualizations that subjects could rate within about 1/2 hour. Six subjects were asked to rate the sixty randomly ordered visualizations. The subjects were chosen based on the same criterion as the initial experiment, their experience with visualization. Two of the subjects had also participated in the previous experiment, but four were new. Figure 63 shows a plot of the combined ratings given by each subject, colored by subject identity. The subject ratings were scaled to use the full 0-9 range because some subjects did not use the entire range. Specifically, the subject colored dark blue tended to rate very low relative to all the other subjects. There is a clear agreement in general trends; correlations between subjects were greater than 0.57 for all subject pairings, which has a significance p-value of less than .0001.

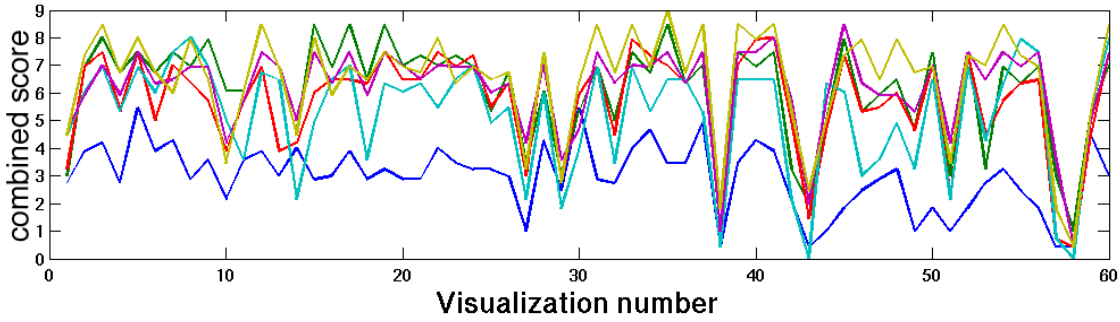



*Figure 63. Subject Ratings Plot. Ratings for the different visualization by subject. Ratings have been scaled to a 0-9 scale because some subjects did not use the full range. [Bair et al. 2006].*

Figure 64 is a box plot of the rating distributions for textures with expected ratings of bad, poor, fair, good, and the two enhanced categories. Both bad and poor textures performed almost exactly as expected, though the fair textures did better than expected, with median better than the textures based on the good rules. Also, the enhanced-B category, with low opacity top surface and larger marks, performed little better than the ‘good’ textures. However, the best performers were the enhanced-A group having mid-range opacity and medium top marks.

Since the 'fair' textures performed better than expected, we look back at the difference in the parameter ranges from the decisions tree nodes in Table Good Leaf. One of the marked differences was that the best leaf had a dark top background and bright marks, while the good leaf had a bright top background and medium bright marks. Is seems possible that the brighter background might actually be better than the darker, especially considering that the background luminance was found to be so important in both the ANOVA and LDA analyses. Overall, the Enhanced-A worked best. These were ones that had a medium opacity top surface with medium-sized marks.

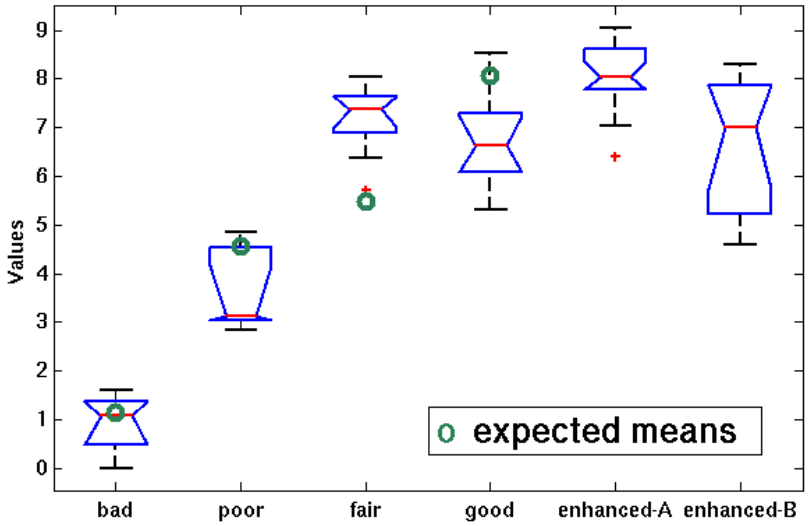



*Figure 64. Expected and Measured Ratings Box Plot. Measured ratings were similar to expected. [Bair et al. 2006].*

Figure 65 shows the texture pairs with the eight best mean ratings. Six of the eight best texture pairs had medium top opacities, high luminance, with relatively small marks, while two others, texture numbers (40) and (39), had low opacities and large marks. Various color combinations existed, and all had prominent marks either on the top or the bottom, more commonly the top. In (2), the only case with visible marks on both layers, the bottom marks are strongly blurred. (34) has no visible marks on the top, but prominent ones on the bottom. All have medium to high randomness, but most of the top textures have a visible global line structure.

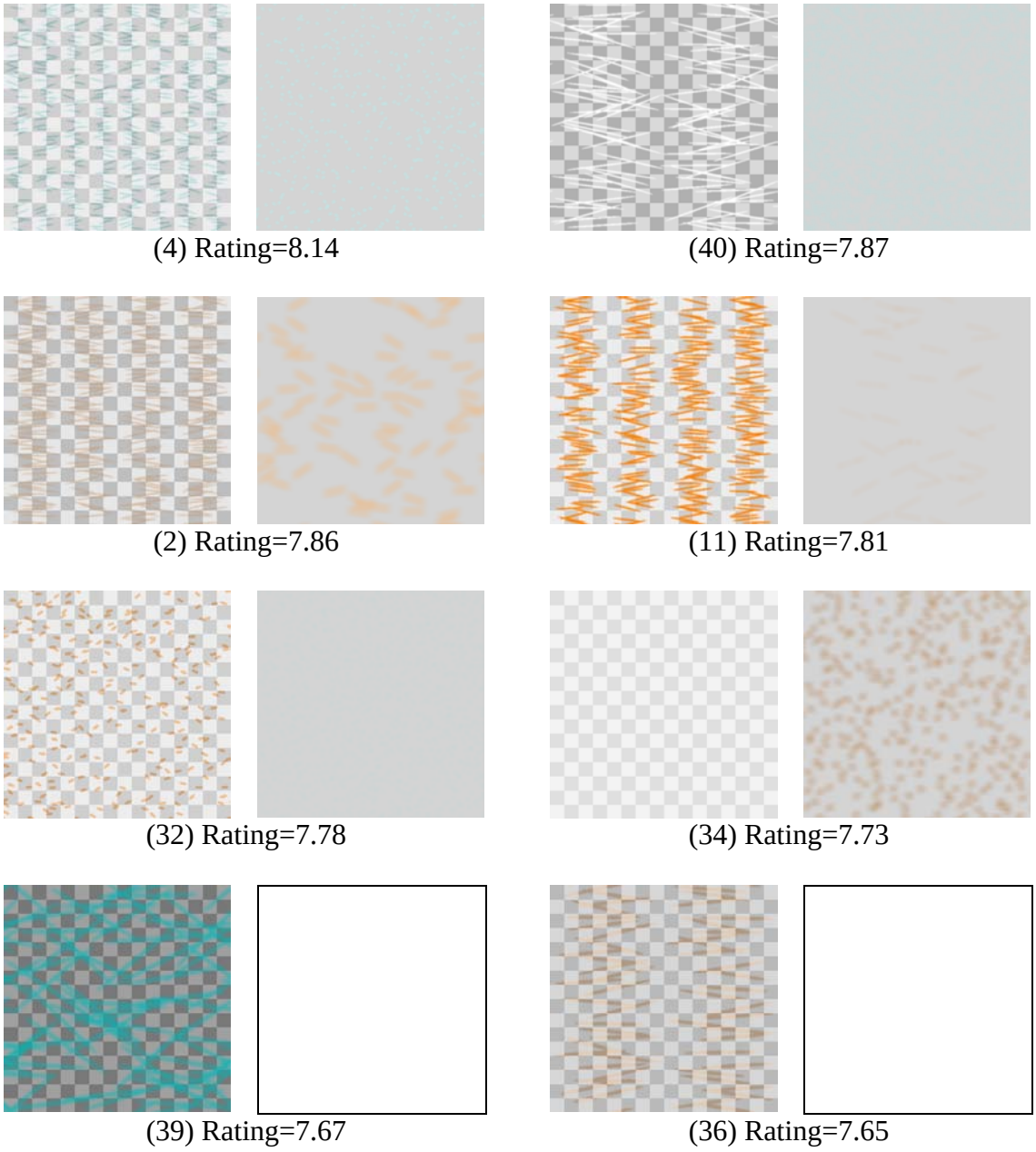



*Figure 65. Best Eight Textures. Highly rated example textures from the evaluation experiment. Top textures are shown on the left, bottom textures on the right. [Bair et al. 2006].*

Figure 66 shows screen captures of the textures in (4), (40), (2) and (11). The case of (40) is rather striking in print because the lines on the top surface clearly show

bumps through their shading, while the bottom bumps are clear from shading on the bottom gray background. It should be understood that when viewed in motion and stereo, (4) is equally convincing, and merits its high rating. However, it is harder to tell which bumps are on the top and bottom from a still image.

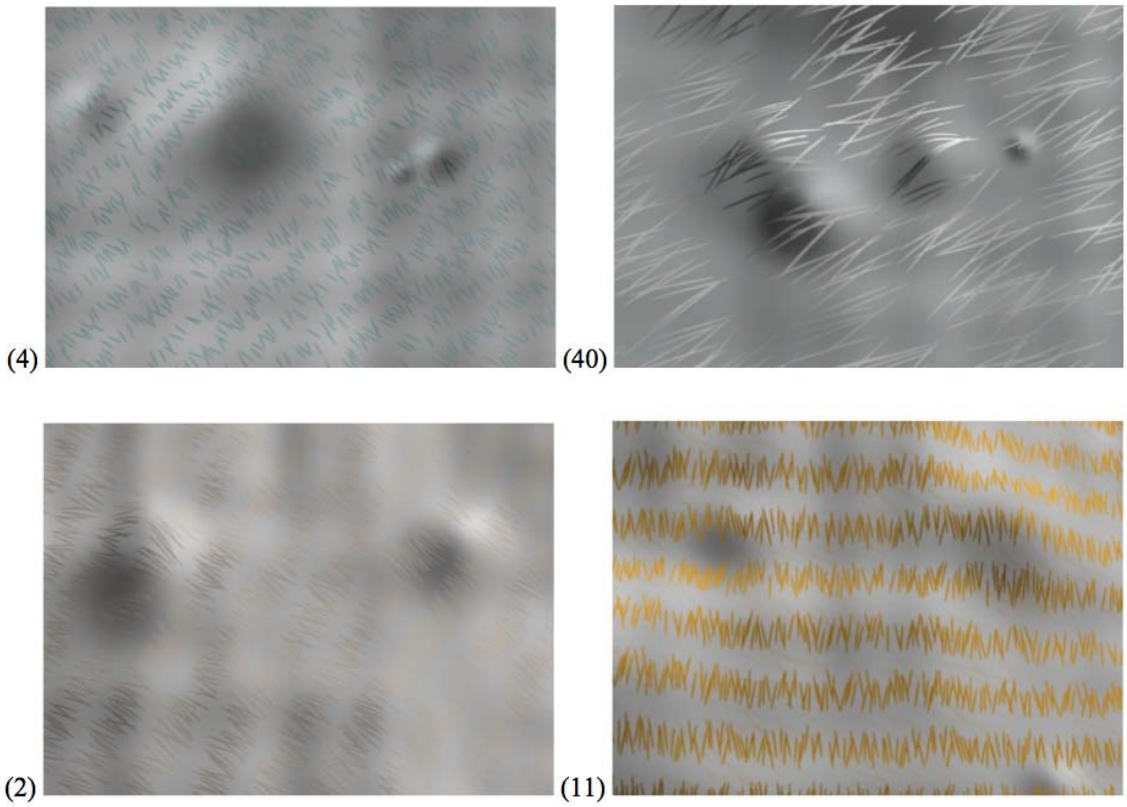



*Figure 66. Best Four Textures on Surfaces. The four best textures shown in close-ups on the surfaces.*

### 4.4 Discussion

Overall, the second multi-parameter experiment was successful in a number of ways. First, a better understanding of parameter importance was found. The grid rows and columns, mark size and aspect ratio, luminance of the background and marks, and top opacity were the most important parameters. This agrees with the results of the previous experiment that overall structure as well as brightness are critical for texture quality.

Results for randomness and filtering were mixed. LDA seemed to suggest that randomness should be high and filtering low on top, and mattered less on the bottom. Although neither parameter showed up in the decision tree rules, the results of the validation experiment seem to support this. The good textures had lots of randomness and low blur on top. The bottom textures seem to only have high blur when the marks are large enough to be noticeable. Lastly, chroma, rotation and mark roundness did not seem to matter at all. This is good information because chroma and roundness could both be used to encode other information about a surface.

Optimal values for the importance parameters were also analyzed. Marks on the top should be long and thin, and on the bottom small dots. Both surfaces should be high in value**.** Middle-range opacities work best, though all but completely opaque tops could work as long as other parameters were set correctly. Two broad types of top textures seemed to work well. One had medium opacity background with small marks, while the other had a low opacity background with longer marks. We hypothesize that in the first case, bumps are mostly seen through shading of the background, while in the second case the marks are large enough to provide shading information. The following set of guidelines is set:

-Bright top, and bright bottom surfaces
-Medium to high grid aspect ratio to provide global structure
-Long, thin lines on top
-Medium to high randomness overall
-Prominent (large, bright, opaque) marks on top
-Subtle (small, low contrast) marks on bottom
Either:
-Medium top background opacity with short, thin top marks
or
-Low top background opacity with long, thin top marks
-Color freely chosen
-Little blur on top, higher blur on bottom

One important factor to discuss is how the task affected the outcome of this experiment. In the first experiment, subjects were simply asked to rate how well they

could see the shape of two known surfaces. It was expected that subjects would consciously or unconsciously include aesthetics in their rating to some extent. This was considered a feature rather than a flaw because beauty has an important place in visualization. Ugly visualizations could be detrimental to task-based performance simply because the subject does not want to look at them. At the very least, an ugly visualization could distract from the purpose of showing surface shape.

Similarly, the task of feature detection causes its own bias in texture ratings simply because of the nature of the task. The task was not truly a measure of subjects' perception of surface shape, but their ability to spot significant changes in surface shape. When using a diffuse lighting model, one of the strongest cues to changes in surface direction is the surface shading. Changes in brightness or darkness of the surface clearly indicate a change in direction. It is quite possible that strong textures are not particularly necessary for subjects to see these changes in shading. This is probably why most of the good textures had at least one surface with almost no noticeable texture. However, as is very clear looking at still images, determining which surface the bumps are on can be difficult. Subjects had the advantages of stereo and motion cues, but most good textures did have noticeable top textures. We believe that this helped distinguish the top and bottom marks without interfering too much with the shading cues.

To build on these conclusions the next step seemed to be to test the hypotheses with controlled experiments. The next two sections describe experiments that were run to follow up the multiparameter experiment. They use only a few parameters so that the results can be tested for statistical significance. All the experiments use a new task as well. Since the feature finding seemed to bias the perceptual cues to favor shading, we looked for a task that would rely more heavily on texture information. Change in surface direction can be identified solely from shading, but actual direction needs more information. Therefore, the task was a measure of surface normal perception. Section 5 compares surfaces without texture to surfaces textured using different sizes and opacities.

# 5. SIZE-OPACITY EXPERIMENTS *

## 5.1 Introduction

To follow up the multi-parameter experiments described in the last two sections, two sets of experiments were designed and run in parallel. These experiments were more typical statistical studies that only varied a limited parameter set. These studies have the advantage of being able to gather sufficient uniformly sampled data on a parameter's levels to measure statistically significance. The disadvantage to these studies is that they do not account for covariance of all parameters. The studies described in this section vary the parameters of texture vs. no texture, size, and opacity. The study in the following section 6 co-varies the parameters of texture structure and size.

Besides varying only a few variables, these experiments use an objective task to measure shape perception. The first multi-parameter experiment used an extremely subjective measure of shape perception, and results might have been strongly colored by the aesthetics of the visualizations as opposed to the actual performance. The second experiment used a feature-finding task. While less subjective, subjects were still given freedom to rate the visualizations, and there was evidence that all subjects were not consistent in this. Also, the feature finding task probably used surface shading as an important cue to simply locate the bumps. This did not mean that subjects could accurately judge the shape at all locations. Some metrics that have been used to measure shape perception are surface normal estimation [Koenderink et al. 1992; Mamassan and Kersten 1996], shape categorization [De Vries et al. 1993; Kim et al. 2004], and curvature estimation [Todd and Mingolla 1983]. Here, we chose to use surface normal estimation as an objective task, because it is a task that can be used with complicated surface shapes that are comparable to real world problems.

---

* The images and text in this section are reproduced in part with permission from previously published material. See [Bair et al. 2007].

The earlier multi-parameter experiments answered a few questions relatively conclusively. Brightness of the both surfaces was important, and higher brightness was generally better. Setting the average opacity of the top surface seemed to require a tradeoff between seeing the top or the bottom surface, with mid-opacities working best overall. Finally, both studies had a bias for top marks being larger and more prominent. Because brightness did not seem to be a tradeoff in any way, it was not investigated further. In all of the further studies, brightness of marks and background are set relatively bright, with exact brightness and colors chosen at the discretion of the experimenters. Since average top opacity was found to be important in both previous experiments, and setting it leads to a tradeoff between top and bottom surface error, it was thought to be a good candidate for further study. In particular, we wished to determine optimal opacity values using an objective task. Also, the top surface bias toward larger and more prominent textures was considered worthy of further investigation. In the first multi-parameter experiment, the larger top textures might only have been optimal because of the large top surface features that were used. In the second multi-parameter experiment, the subtle bottom textures might have been found useful only because of the feature-finding task that emphasized shading over texture. Therefore, large vs. small and obvious vs. subtle top to bottom texture differences were deemed worthy of investigation using an objective shape estimation task.

### 5.2 Experimental Design of Texture Style Experiment

The surfaces were designed as height fields with the intent to simulate terrain or other natural features. The surface construction algorithm drew 100 Gabor bumps [Gabor Filter; Feichtinger and Strohmer] on each surface. The Gabor bumps are formed by multiplying a cosine function with a Gaussian function. When the period of the cosine function is similar in magnitude to the Gaussian falloff parameter, this produces hills with small valleys on either side. Since hills and valleys were the desired shapes to mimic terrain, the Gabor functions were ideal to use. The position, amplitude, orientation, cosine period and the Gaussian falloff parameters for each bump were

randomized. The cosine period varied between 7.5-20% of the surface width, and σ varied from 40-70% of the period. This made features with about 1.3-6° of visual angle, larger on average than the bumps from the feature finding experiment, but smaller than the surfaces in the first multi-parameter experiment. Figure 67 shows an example of a single surface made using these parameters, along with the probe used for the task of estimating surface normals. The surfaces were displayed full screen at a distance of 85cm (or slightly more than arm's length) from the subject, on monitors 48 cm wide and 30 cm high, with a screen resolution of 3840 x 2400. Surfaces were displayed in stereo using a Wheatstone stereoscope setup as described in Section 4. Surfaces were rocked with an amplitude of 8º and a period of 2 seconds to provide motion cues to depth. The motion was less than in previous experiments so as not to interfere with the task of manipulating a surface normal probe. The camera positions and field of view were adjusted to make the leading edge of the top surface appear to have the same depth as the physical monitor. At these settings, the upper and lower surfaces would appear to have a vertical separation of about 5cm, with maximum height variations of around 4cm.

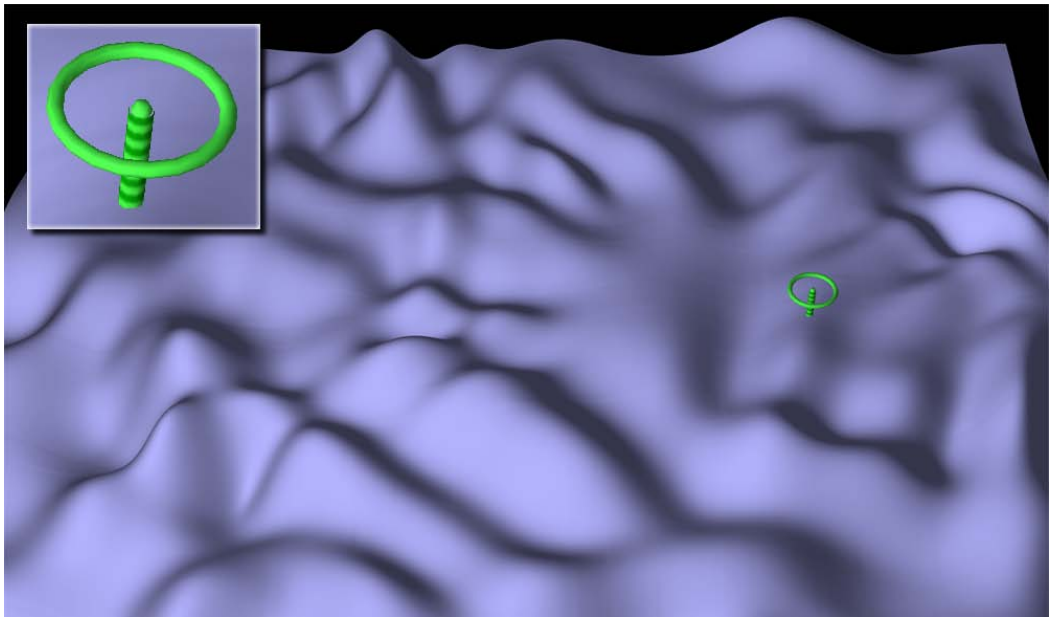


*Figure 67. Gabor Surface and Probe. The Gabor surface simulates a natural terrain, shown with a surface normal probe. [Bair et al. 2007].*

Subjects were given the task of using the mouse to manipulate a direction probe in order to align it with the perceived surface normal at a given point on either the top or bottom surface. The task was intended to be an objective measure of shape perception by measuring the error in surface normal estimation. Figure 68 shows examples of probes aligned either with or far from the correct surface normal. The probe is of an unpublished design by Colin Ware. It consists of a fully three-dimensional cylindrical pole with an applied stripe texture, topped by a torus that is at right angles to the pole. The ellipse of the projected torus helps show the probe direction with minimal occlusion of both the cylinder and underlying surface. The torus is at the top of the cylinder because if it were at the bottom, occlusion with the surface would give additional shape information. Finally, the stripes mirror the torus direction as well as providing high-frequency texture to aid stereo depth perception. The probe is rendered with the same lighting and shading as the surface. Thus, probe visual orientation cues come from the shape, shading and foreshortening of the texture stripes. This design makes the probe direction much clearer than the simple line probes used in previous research [Interrante and Kim 2001; Koenderink et al. 1992], and does not require a separate enlarged view of the probe [Sweet and Ware 2004]. Measured errors should be primarily due to errors in surface direction estimation rather than probe direction estimation, however, we have not done a controlled evaluation of this design verify this.

Subjects were first trained to manipulate the probes while being shown the correct normal, and a color-coding system indicating their current angular error. These are the right-hand probes shown in Figure 68. Next, subjects were given a series of 'practice' probes in which they aligned the probe to the best of their ability, and then only later were shown the correct normal and their angular error. During the actual experiment, subjects were not shown their errors. All training was done on a single surface to avoid bias in training.

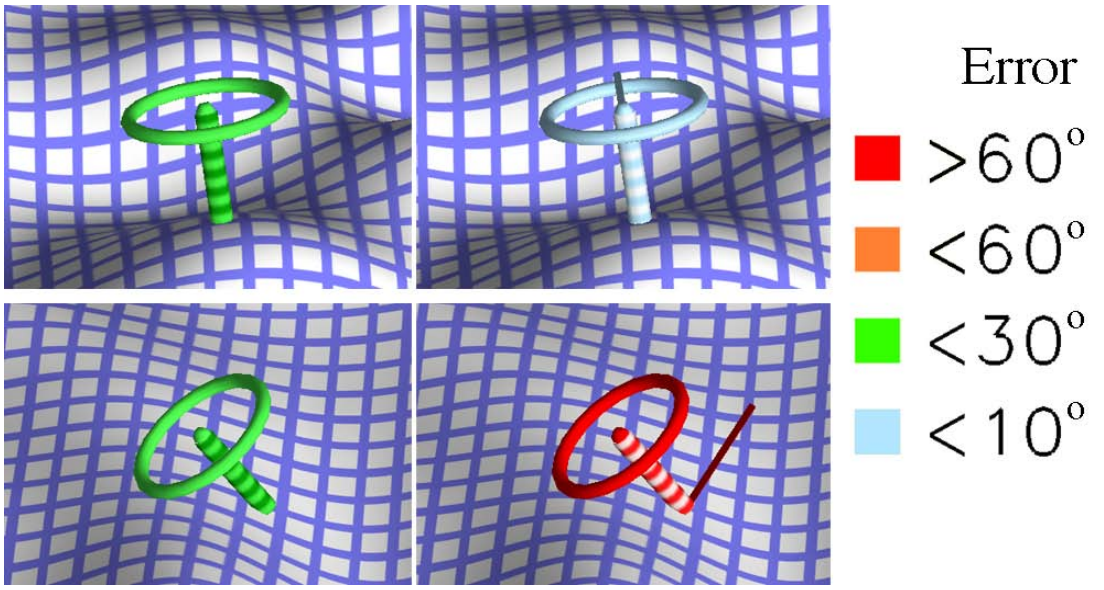



*Figure 68. Probe Examples. Examples of well and poorly aligned probes. [Bair et al. 2007].*

The overall texture structure was chosen to be uniform grids. Both previous multi-parameter experiments suggested that structure was useful, particularly on the top surface, and independent studies [Sweet and Ware 2004] and [Kim et al. 2003] have shown that bi-directional textures work particularly well for single surfaces. Grids textures are a simple solution that gives bi-directional texture information, parallel lines and line junctions as surface shape cues. Some variations on texture structure are investigated in Section 6.

Although the multi-parameter experiments suggested that subtle bottom textures worked best, pilot studies showed that these subtle textures did not work well for the task of orienting normal probes. Presumably, subtle bottom surface textures are better for a task that simply requires recognizing the presence or absence of bumps because shading information is sufficient to distinguish a symmetric bump from a noisy background. Also, a subtle texture on the bottom might be more aesthetically pleasing because it does not visually interfere with the top texture. However, in the case of a probe-alignment task, subtle textures on the bottom did not prove adequate. The subtle textures tested had either very thin lines or line colors nearly identical to the background.

As a result, in the textures picked for the actual experiment, brightness and saturation were made equal on both top and bottom surfaces, and the grid line coverage area was identical on top and bottom. Neither surface used either blur or randomization. Top surface texture backgrounds were given 50% opacity so that approximately the same pixel intensity contribution was made by each surface. The grid lines were made opaque to give high texture contrast between the lines and holes.

Grid spacing was chosen to be smaller than the smallest surface feature size, but big enough to be easily recognized as a grid. The three grid spacings chosen had 60, 100 or 140 grid lines along a surface edge, which roughly correspond to a grid spacing of 1.0 cm, 0.5 cm and 0.3 cm seen at 85 cm. The smallest features had a period of about 2 cm, so the grid lines had periods roughly two to six times the feature periods. The top surface texture was rotated 45° relative to the bottom texture to minimize Moire patterns from near-parallel lines, and because the first multi-parameter experiment had some evidence that a difference in rotation was important. Finally, hues were chosen giving a red top surface, a blue bottom surface and a green probe. This was solely for aesthetic reasons, since previous results suggest that hue is unlikely to affect surface shape perception. Color cues are probably ignored when the much stronger cues of shading, stereo and motion are present.

Six different texture cases were used. They are shown in Figure 69, with larger versions in Figure 81 at the end of this section. Case 0 is a single surface with a grid texture. This gives a baseline for minimum subject error, because we can safely assume that a task is easier for a single surface than for two, and as stated above, a grid has been shown to be a good single surface texture. Case 1 is the no-texture case. Although preliminary work showed that subtle textures worked poorly, and by extension no texture will work even more poorly, it seemed important to include this case to give data for an upper limit on worst-case error. Case 2 has medium grids on both top and bottom surfaces. This was intended to serve as a baseline for the grid textures. By comparison, case 3 has a small top grid with a large bottom grid, and case 4 has a large top grid and a small bottom grid. Since our pilot study showed that very subtle bottom textures were

not helpful, difference in texture size was another possible useful cue. Finally, the last texture shown in case 5 was constructed from equally spaced grids on top and bottom, but with thinner lines on top and a translucent background between lines. The width of the thin lines was set to provide an average of 25% opacity, and the background transparency to give another 25%, making the total average opacity again 50%. Two types of textures were found to be useful in the second multi-parameter experiment. The first had a semi-transparent background and small marks while the second had a fully transparent background and larger marks. By using thinner lines and a translucent background in (5), we can compare it with the base texture in (2).

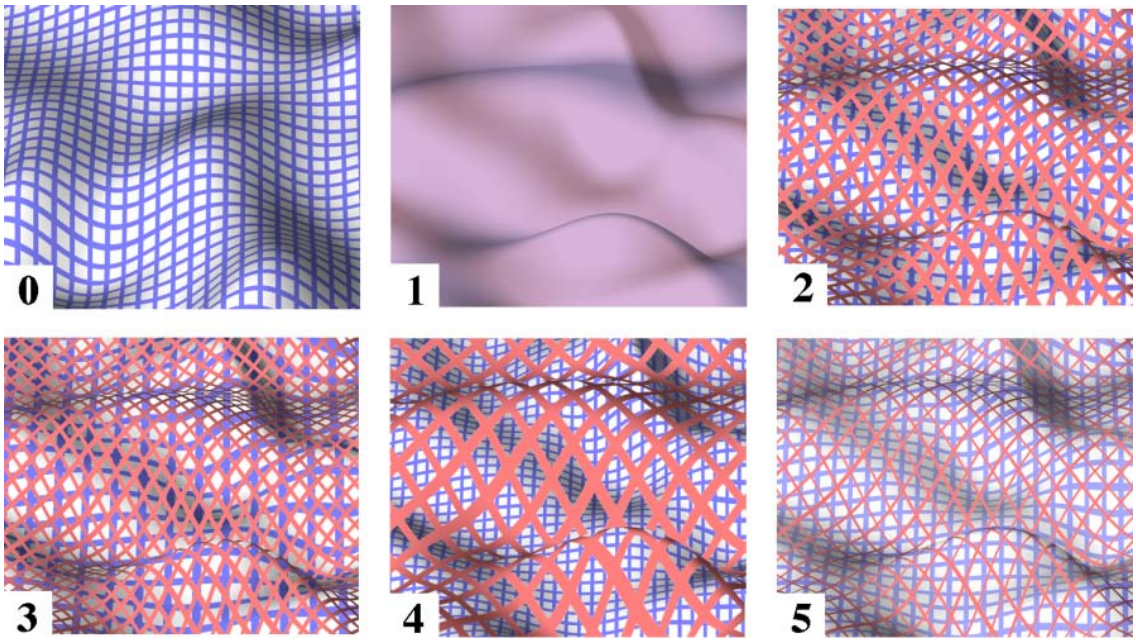



*Figure 69. Texture Variations. The six textures styles used for our first experiment. [Bair et al. 2007].*

14 subjects (11 students and three faculty, six female and eight male) participated in the first experiment. Subjects had normal or corrected-to-normal vision, and were tested for stereovision and color-blindness. All subjects had significant experience in visualization, and were familiar with the concept of surface normals. However, most

were naïve to the purpose of the experiment.

Each subject was presented with a total of 240 probes; 40 total probes for each of the six texture types, with half of the probes on the top surface, and half on the bottom. The experiment was divided into five sets, with a random ordering of textures and top/bottom probes within each set to minimize learning and fatigue effects. The surfaces were generated randomly, and each probe was placed in a random position on the selected surface.

After the experiment, subjects were given a questionnaire designed to evaluate how the subjects had perceived the different textures, and to determine if any unpredicted factors might be affecting the results. The subjects were asked to rank order the five layered texture cases according to how attractive they found each visualization. This was asked because a visualization that subjects find dreadfully ugly may not be useful even if by perceptual measures it works fine. Subjects were also asked to rank order their estimated performance on the five texture cases. Some subjects admitted they had trouble separating these two qualities, but most answered differently for 'attractiveness' and 'performance'. They were also asked to report any problems they might have had, or comments and suggestions.

### 5.3 Results of Texture Style Experiment

Probe error rates were analyzed using a two-way analysis of variance measuring the effects of texture type and subject. The following three figures show multi-comparison plots with a controlled family-wise error rate of $\alpha=0.05$ showing each texture type and corresponding errors. Top and bottom surface errors are analyzed separately, as are the combined errors. Error is measured as the angle in degrees between the probe direction and the true surface normal direction. Line lengths show the 95% confidence intervals for each texture style, so lines that do not overlap can be considered significantly different, according to Tukey's honestly significant difference (hsd) test. Color variation and the dashed lines are included for clarity. Red lines denote all texture styles that are significantly different from the selected blue line. From Figure 70 it is

clear that there is a significant difference [df=5, F=180, p=0] in error between the single-surface case 0, the un-textured layered case 1, and the grid-textured layered cases 2-5. Looking at a single surface alone gives a relatively low average error of 9.33º. Although the textured layered surfaces have higher errors (12.4º, 11.7º, 11.2º and 11.1º), the best was the thin line case 5, with an average error only 1.8º worse, or about 20% more error than for a single surface. It is also clear that the no-texture case 1 is by far the worst, with an average error of 22.1º. Because of the way that the height fields used for this experiment were constructed, simply guessing straight up for every probe would give an average error of 25.6º.

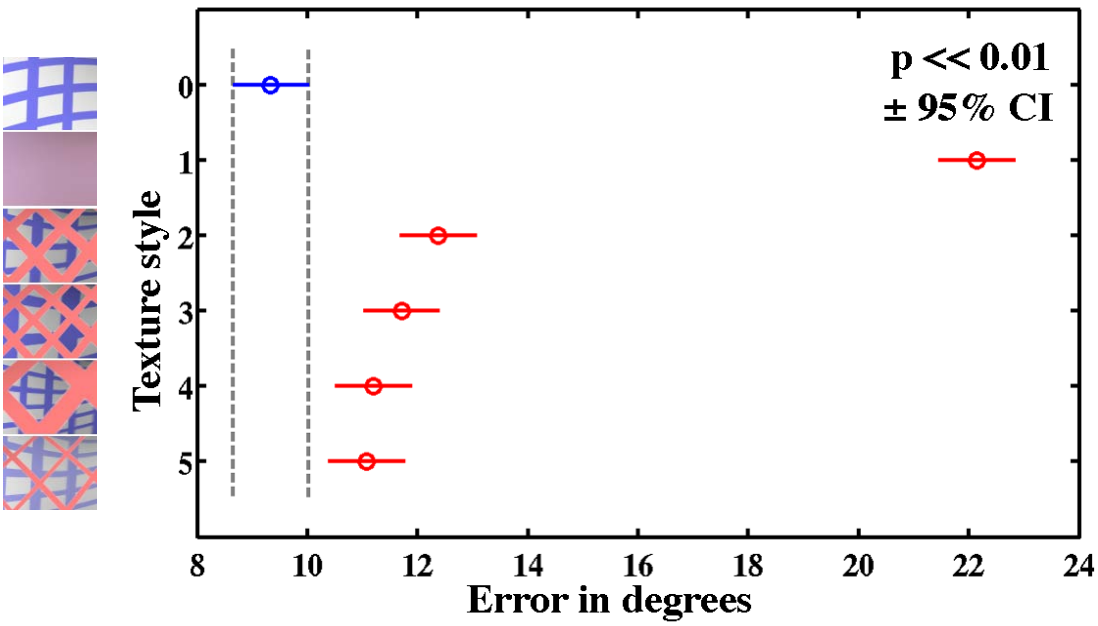



*Figure 70. Average Texture Errors. Average error rates for all textures. [Bair et al. 2007].*

Next, we compared error rates for the four layered, grid-textured cases on top and bottom surfaces separately. Figure 71 shows no significant differences [df=3, F=1.1, p=0.34] between the grid styles for top surface error. However, Figure 72 shows that case 2 is significantly worse than cases 4 and 5, while case 5 is significantly better than

cases 2 and 3 [df=3, F=5.5. p=0.001]. The texture-subject interaction term was significant; several subjects performed better under the equal-grids or small-top grid conditions, but overall cases 4 and 5 are better. It seems that equally spaced grids with equal line widths make it most difficult to see the bottom surface.

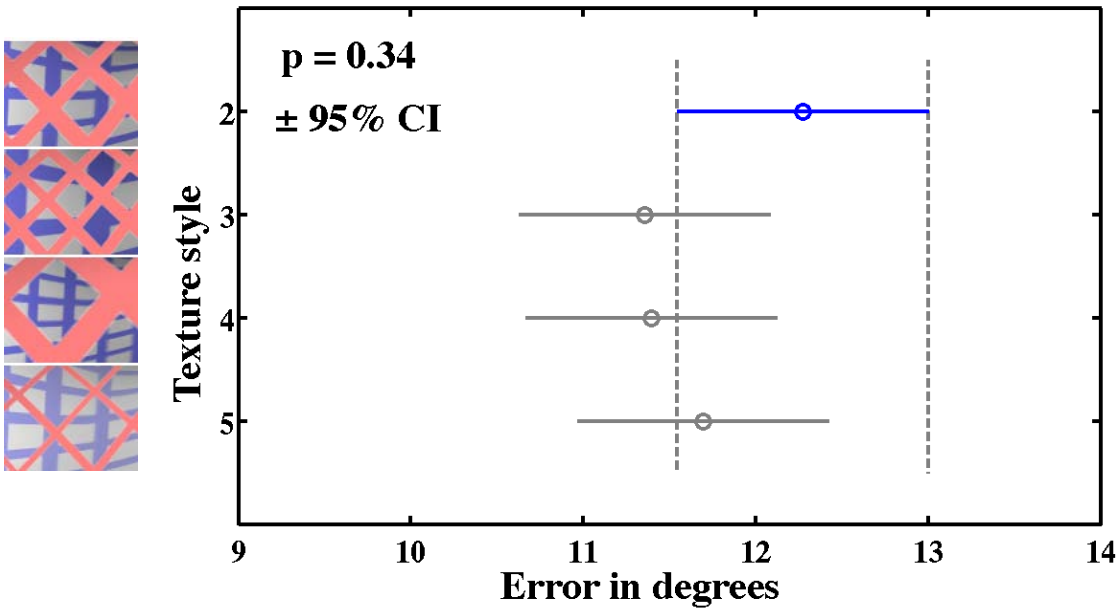



*Figure 71. Top Surface Errors. Top surface error rates for grid textures. [Bair et al. 2007].*

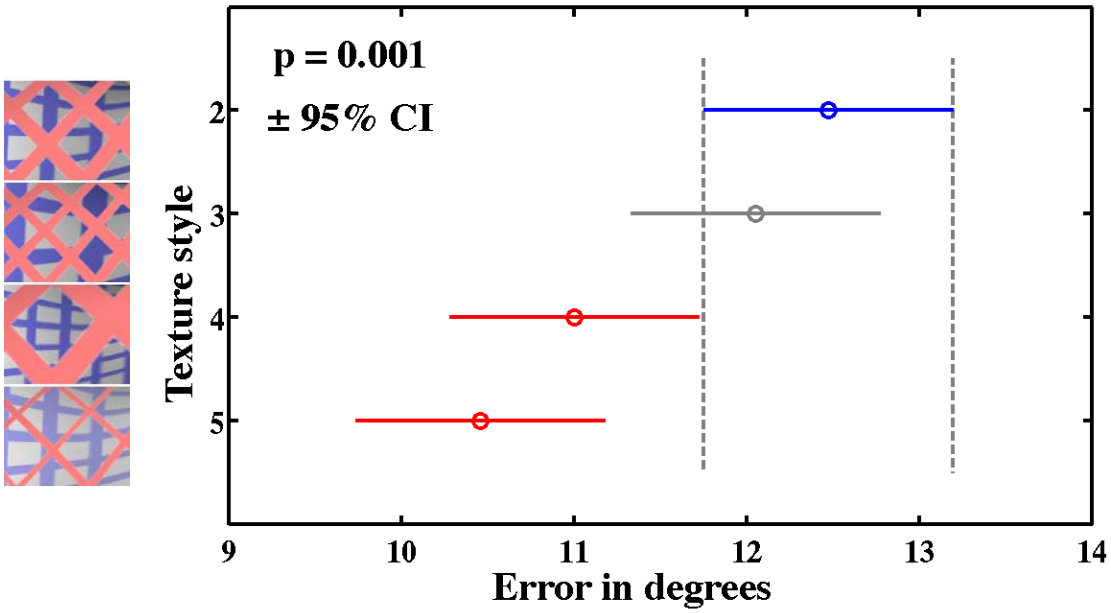



*Figure 72. Bottom Surface Errors. Bottom surface error rates for grid textures. [Bair et al. 2007].*

We have two possible interpretations for these results. The first is a global argument, and the second a more local one. The global argument is that having grids with different spatial frequencies makes it easier for the human visual system to separate the visual signals across the two surfaces. The top and bottom frequency spectra in case 5 are also different because the thin lines on the top surface add more high frequencies.

The second, local argument involves line junctions. Extreme close-up views of the textures are shown in Figure 73 so that the line junctions can be clearly seen. As discussed in detail in the Background section, humans use line junctions for reconstructing 3D shape from line drawings. But with equal grid spacing, as in case 2, most of the bottom surface line junctions are at least partially occluded by the top grid. The same is true in case 3, but the thicker bottom lines allow easier interpolation of the grid lines, so that junctions can be inferred. On the other hand, both cases 4 and 5 have larger openings in the top grid because of grid spacing and line width respectively. Consequently, at least one line junction is visible through every opening, improving the ability to extract 3D shape.

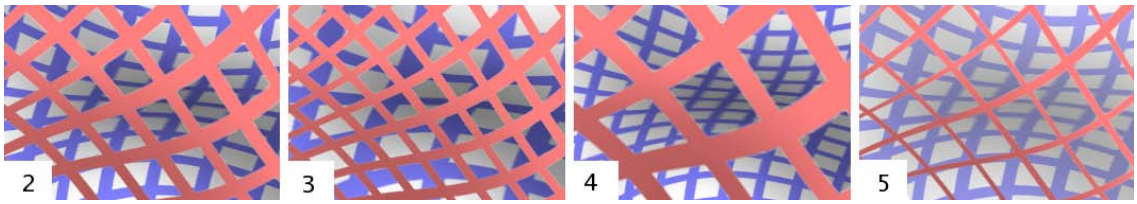



*Figure 73. Textures Zoomed. A zoomed view of the layered grid textures.*

Interestingly enough, although subjects were told to try to keep a consistent pace while aligning the probes, time spent on probes mirrors the error rates. This effect is apparent in Figure 74 [df=5, F=5.9, p=0.00002]. Although the effects were not as extreme, this does support the findings based on error-rates; it seems that subjects tended to both spend more time and perform worse on some of the visualizations. Unlike error, times for top and bottom surfaces were both similar and not significant, and so are not shown separately.

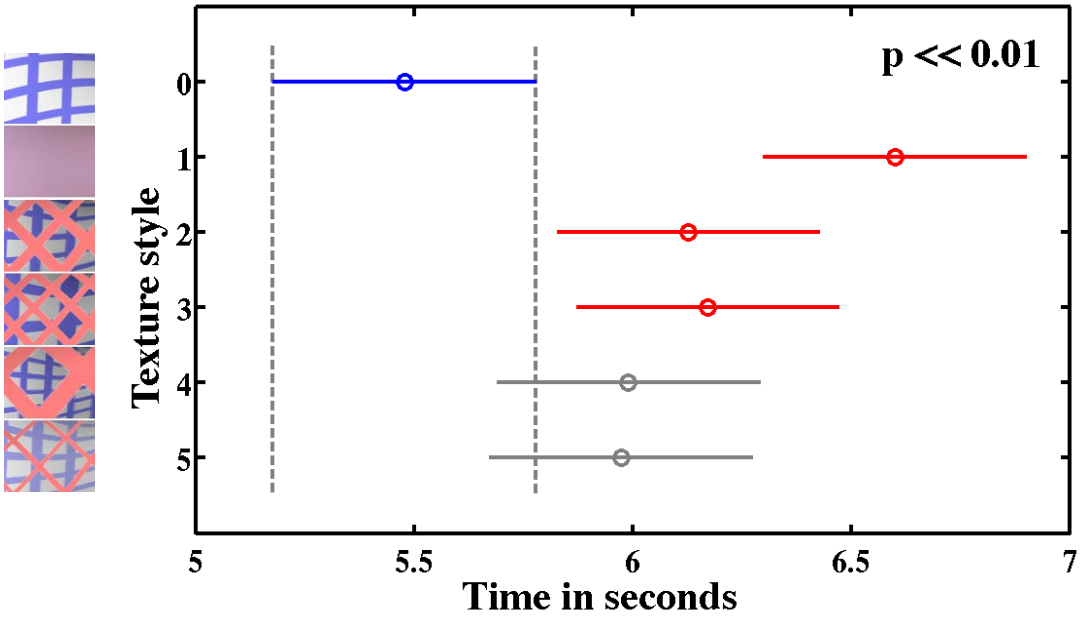



*Figure 74. Average Times. Average time for all textures. [Bair et al. 2007].*

The results for the post-experiment questionnaire are shown in Figure 75. Results for all 14 subjects were summed with a ‘Best’ answer being +2, ‘Better’ being +1, ‘Average’ being 0, ‘Worse’ being –1 and ‘Worst’ begin –2. This makes +28 and -28 the best and worst possible scores. Clearly the subjects’ beliefs about how well they could perform the tasks are very similar to their actual results. The left half of Figure 75 shows performance ratings. Every subject rated the un-textured case as ‘Worst’, and most thought that cases 4 or 5 were best. It seems worthwhile to note that when asked about any problems with the experiment, several subjects complained that they had trouble telling which surface the probe was on in the un-textured case. Clearly, use of texture is extremely important even when many other cues are available, like shading, stereo and motion. The subjective ‘beauty’ results shown in the right half of Figure 75 strengthen the argument for using texture types of 4 or 5 as well. One goal of visualization is to be attractive; subjects are more likely to use a visualization technique if they find it attractive to look at.

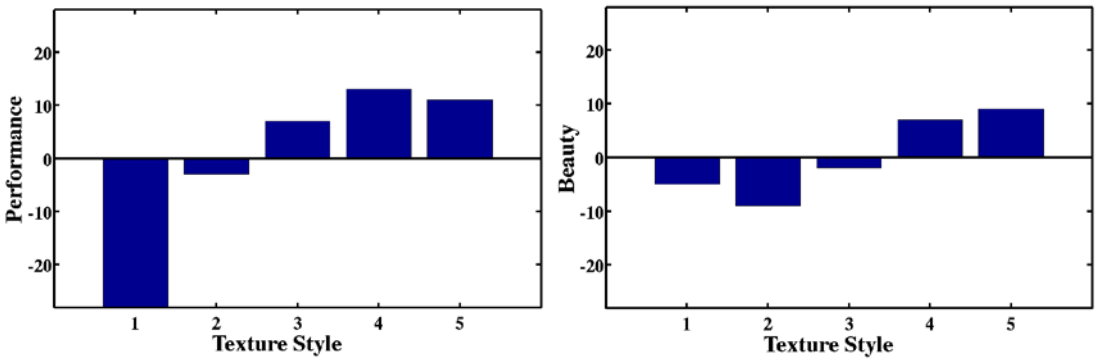



*Figure 75. Questionnaire Results. How subjects rated texture beauty and their expected performance. [Bair et al. 2007].*

## 5.4 Experimental Design of Opacity Experiment

In the next experiment we investigated how variations in the opacity of the top surface might affect error rates on the top and bottom surfaces. Opacity was varied

between 30% and 70% on the top surface, with five levels. The previous studies suggested that mid-range opacities worked best, and the goal here was to test this guideline with an objective task. Grids were drawn with a combination of the cases 4 and 5 from the previous experiment. We hypothesized, based on the results from the first experiment, that thin lines and larger grid spacing on the top surface made it easier to see and interpolate line junctions on the bottom. Therefore the grids for this experiment were drawn with both larger spacing and with thin lines and a translucent background. Both line width and background opacity were varied to create the 30%, 40%, 50%, 60% and 70% opacities shown in Figure 76. Full-surface versions of these are shown at the end of the section.

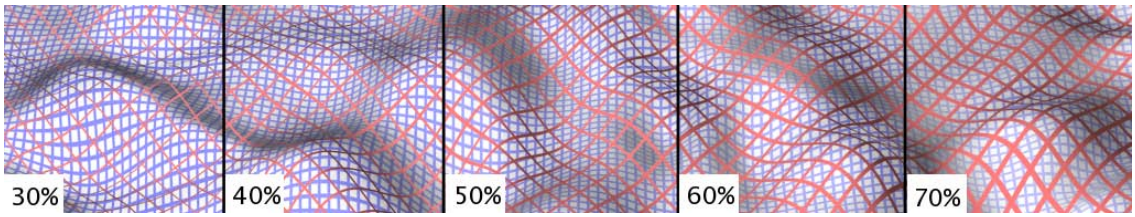



*Figure 76. Opacity Levels. Layered surfaces with top opacity level varying from 30-70%. [Bair et al. 2007].*

Seven subjects ran this experiment (3 female, 4 male, 1 professor, 6 students). Three of these subjects had participated in the previous experiment. Conditions and training were like those of the first experiment. Each subject ran 200 total probes giving 20 probes each for both top and bottom surfaces for each of the five opacity levels.

### 5.5 Results of Opacity Experiment

Surprisingly, our results show almost no significant differences among the various opacities. Figure 77 shows average error for both top and bottom surfaces, with no significant effect [df=4, F=0.9, p=0.46]. Figure 78 shows a general trend toward higher bottom error with higher top surface opacity [df=4, F=2.7, p=0.03]. This we

would expect. On the other hand, accuracy on the top surface, in Figure 79, did not vary significantly with opacity [df=4, F=1.3, p=0.28]. This is surprising, but it fits with our results for the grid size experiment, in which top surface errors had no significant variation with texture sizes. We think that perhaps seeing the top surface is generally the easier task, as it does not require estimating surface shape from behind occluding objects. Also, the grid seems to work very well for showing the top surface shape even with very little opacity.

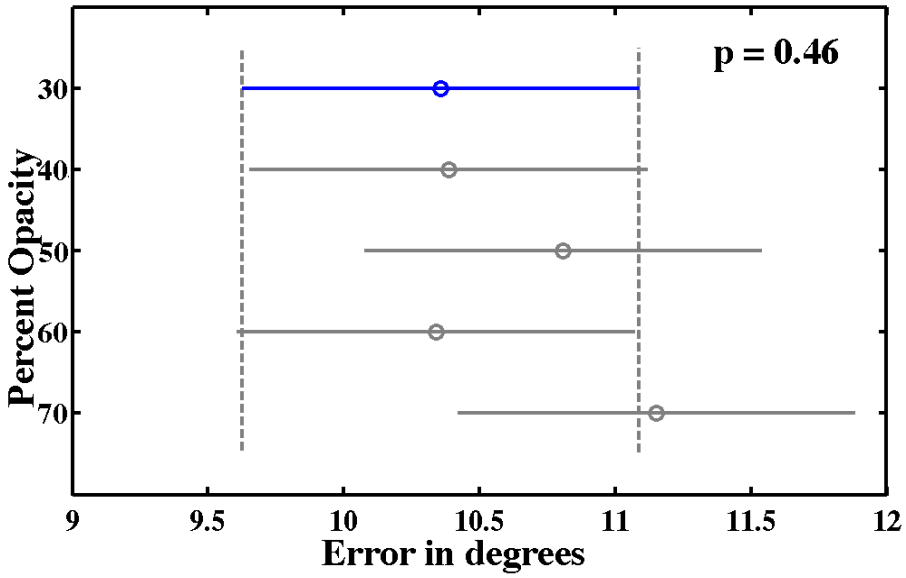



*Figure 77. Average Opacity Errors. Average errors for different top opacities. The p-value was not significant. [Bair et al. 2007].*

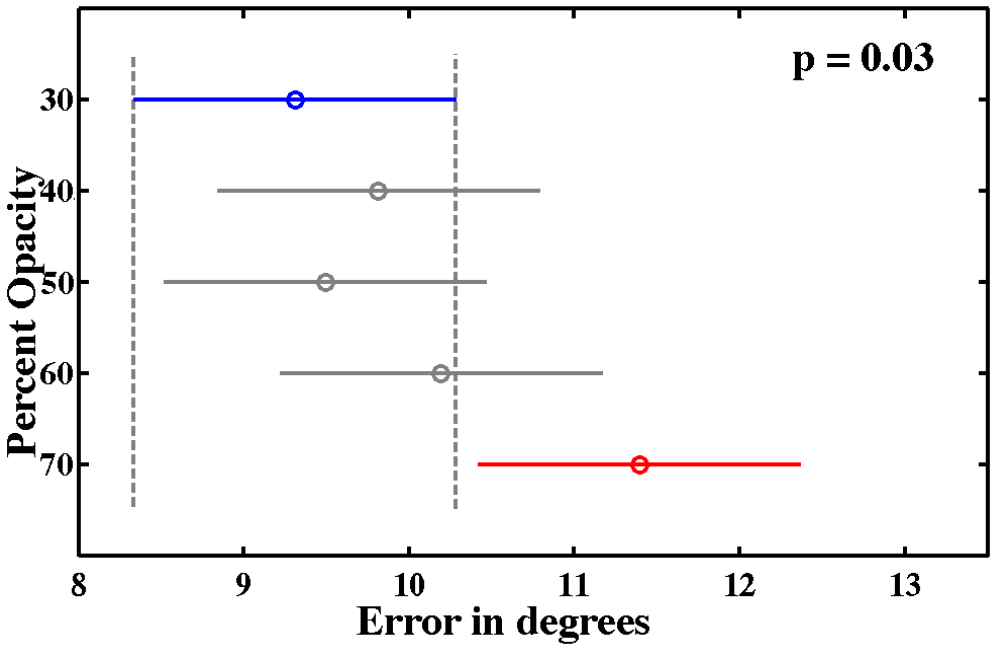



*Figure 78. Bottom Surface Errors. Bottom surface errors for different top opacities. [Bair et al. 2007].*

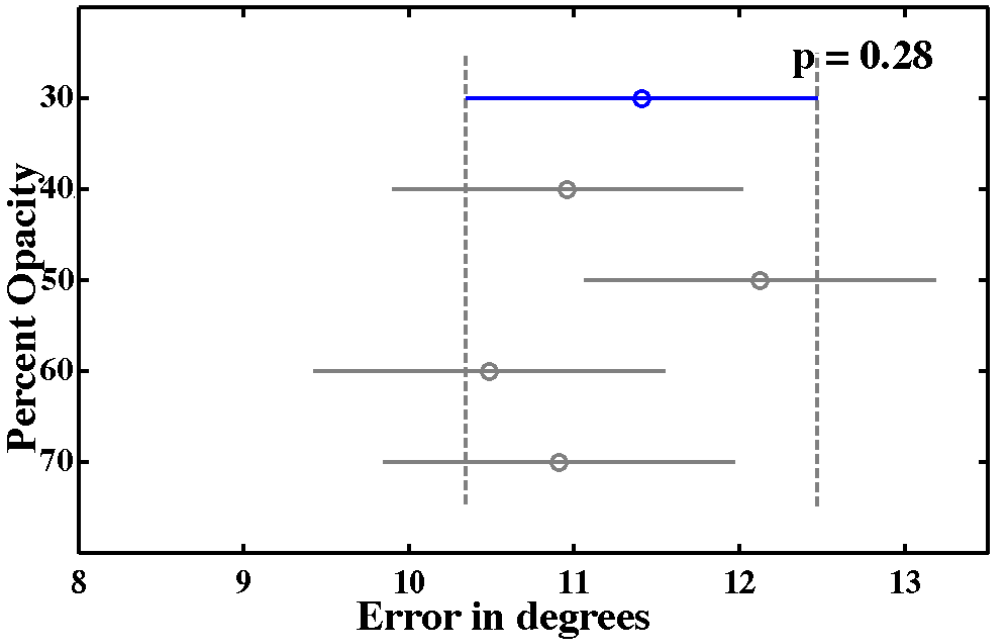



*Figure 79. Top Surface Errors. Top surface errors for different top opacities. The p-value was not significant. [Bair et al. 2007].*

### 5.6 Follow-up Experiment

We were curious to see if our hypothesis that a combination of larger top grid and thinner top lines would work better than changing grid spacing and line width individually. A small experiment was run, identical in setup to the first two experiments, comparing cases 4 and 5 from the first experiment with the 50% opacity case from the second experiment. Figure 80 shows these three textures. The first has a coarse top grid and a fine bottom grid. The second has a medium top grid with thin lines and a medium bottom grid. The third has a coarse top grid with thin lines and a fine bottom grid, combining the aspects of the first two that make bottom surface visibility clearer.

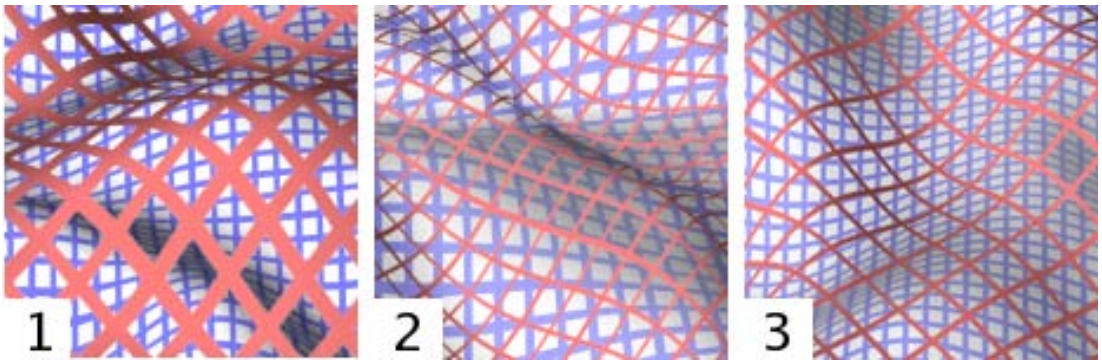



*Figure 80. Follow-Up Cases. The three texture types compared in the follow-up experiment.*

Four subjects ran the experiment (three male, one female, one professor, three students), though three had participated in both previous experiments, so only one subject was truly naïve to the purpose of the study. 20 probes were set for both the top and bottom surfaces for each of the three textures conditions totaling 60 probes per subject. Possibly due to the small number of subjects, no significant differences were found in either error or time differences between the three textures styles on either top or bottom surfaces. However, as suspected, the trend on the bottom surface was for the thin lines (case 2) to perform better than the larger grid spacing (case 1), and the combination of the two (case 3) performed the best. Also, although still not significant, the trend on

the top surface was the opposite. The combined thin lines and large spacing grid (case 3) performed worst on average, while the coarse grid (case 1) performed the best. Again, time spent correlated strongly with error rate. Although no significant results were found, this seems to demonstrate the balancing act required to show both surfaces clearly, with case 2 (case 5 in the first experiment) being arguably the best solution.

### 5.7 Discussion

The most simple, yet most dramatic result from these experiments is that applying well-chosen textures to the surfaces in a layered display dramatically improves the ability to understand surface shape. This is true even when other viewing conditions are optimal for seeing shape, such as shading, stereo and motion. Other researchers have shown this result, but it is important that it be emphasized again because simple shading is still the most common method used in layered surface visualization. For example, when viewing the simple shaded case, several subjects said they had difficulty telling whether the probe was on the top or bottom surface, or even which surface was on top or bottom. Many also said that the probe appeared to be floating in midair, not *attached* to any surface. Without fine-grained texture, even the combination of stereo, motion and shading were apparently not enough to allow subjects much better than a random guess. To be fair, common layered surface visualizations often use objects with boundaries and/or differing spatial-frequencies, allowing shading to act like a texture. These shading and boundary cues may lessen the confusion and decrease the likelihood that subjects will perceive a feature on one surface when it is actually on the other surface. Still, our results are a compelling argument for including more use of texture in scientific visualization when understanding surface shape is an important goal.

Also, our results give an estimate of the error increase when surfaces are layered. In the first experiment, the best layered-texture case had only 20% more error while showing twice the information of the single-surface case. Certainly, for some applications a decrease in accuracy would be a fair trade-off given the extra information shown. Even visualizations with more than two layers might benefit from showing two

of those surfaces simultaneously, letting the others be occluded or made mostly transparent. This begs the question of how layering more than two surfaces might perform. An experiment to estimate how error scales with number of layers is described later in Section 7.

Our results also suggest the hypothesis that for the task of estimating surface shape, larger top surface texture patterns or thinner top lines enhance perception of the bottom surface. That is, grids with larger transparent areas make it easier to see the bottom surface because more line junctions are visible, without compromising top surface visibility. Examples of these are the rightmost textures in Figure 81. The follow-up experiment suggested that making the top lines too thin and widely spaced might hurt the top surface errors. This indicates that the translucent background that was shown in the previous section to be helpful for feature detection on the top surface might not be as useful as the grid lines when the task is estimating the surface normal.

We were surprised to find how little top surface opacity affected surface shape estimation accuracy over a large range of different opacities. Like the texture spacing, opacity only seemed to affect bottom surface accuracy, with low and middle opacities all working quite well. We believe that the small effect opacity appears to have is due to the initial quality of the visualization, including lighting, viewing angle, texture, stereo and motion. By setting these parameters carefully, a poor choice of opacity simply has less of an effect. Also, it should be noted that for high top opacities there was a significant drop in accuracy, so for applications where the bottom surface is more important, such as in viewing the brain inside the skull, the top opacity should not be too high, like the 70% opacity in Figure 82. On the other hand, it is intriguing to see that the top surface accuracy can be very high even with very low opacities, like the 30% opacity in Figure 82. A low opacity top surface distracts little from the bottom surface, but apparently conveys a great deal about its shape. However, we would hesitate to recommend using extremely low top opacities for tasks that require picking out features quickly, since in this case shading may be a more important factor than texture, and low opacities reduce shading information.

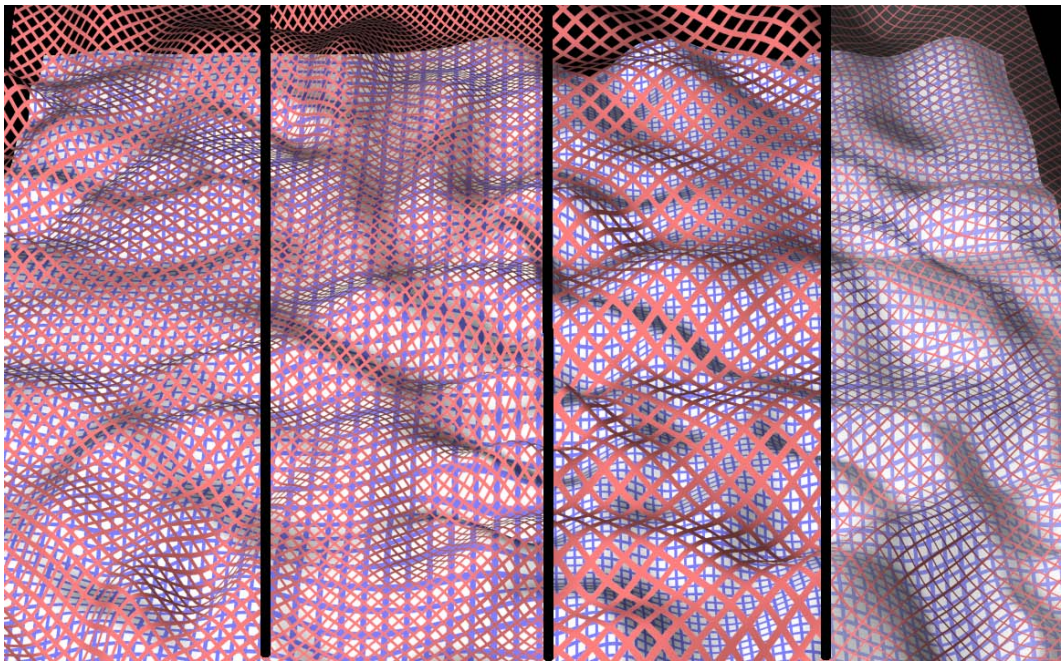


*Figure 81. Texture Variations Full-View. The four texture size combinations used in the first experiment.*

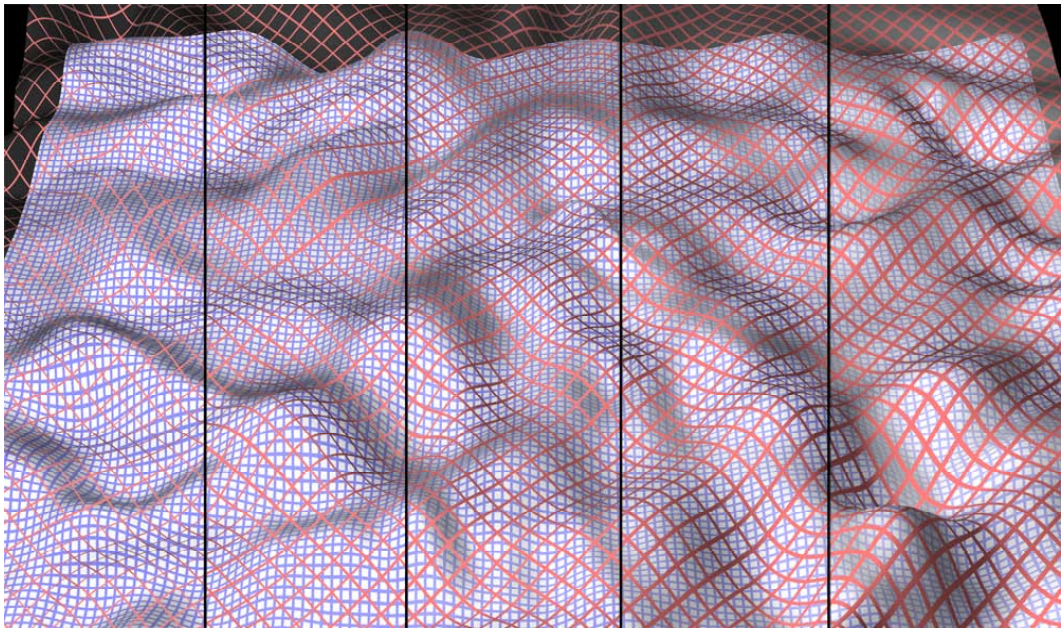


*Figure 82. Opacity Levels Full-View. A full-surface view of the different opacities used in the second experiment. [Bair et al. 2007].*

The experiments discussed in this section certainly do not cover all of the possible interactions of texture size. Therefore the next experiment, described in Section 6, includes top and bottom texture size as parameters. This new experiment also investigates texture structure looking at grids, hatches and dots.

# 6. SIZE-STRUCTURE EXPERIMENT

## 6.1 Introduction

The results for the texture size experiments suggested that large top grid spacing and thin top lines made it easier to see the bottom surface. Although the results were not significant, the trend in both the main and follow-up experiment was for large top grid spacing and thin top lines to lead to relatively worse top surface error. However, the experiments did not fully test all the top and bottom grid size possibilities, and so more combinations were tested and are reported here. Also, texture structure was found to be important for top surfaces in both the multi-parameter experiments, and the structure of a texture has already been shown to be important for single surfaces. This experiment is designed to confirm whether structure is important for both surfaces in a layered visualization.

## 6.2 Experimental Design

This experiment seeks to explore both main effects and interactions between texture structure and size on the top and bottom surfaces. As shown in Figure 83, three types of textures are used: dots, hatched-lines and grids, and two sizes, coarse and fine. Each surface could have any combination of texture size and structure type, giving 6 possibilities per surface, and a total of 36 possible combinations. Texture was used in all cases because both the previous experiment in Section 5 and other research has shown that using texture is very important to performance for a task like surface normal estimation.

Grids were chosen because they have been shown several times to be a useful texture structure. They offer parallel and evenly spaced lines that can be used to mentally measure distance, and junctions that are a powerful shape cue. The hatched lines showed promising results in both of the multi-parameter experiments, and were included as an alternative to grids. Hatched lines also offer global structure, although it is not evenly spaced. Since the hatches are arranged in lines perpendicular to the actual hatch marks,

they offer information in two directions, although the second direction must be mentally interpolated. Randomized dots were included to test whether texture structure is important, or if the texture cues that are available from a randomized texture are enough. The Background Section explains the perceptual cues, such as compression or size and density gradients, which can be used with any uniform texture. The dots were made to be relatively small because a fine texture is helpful to using stereo as a depth cue.

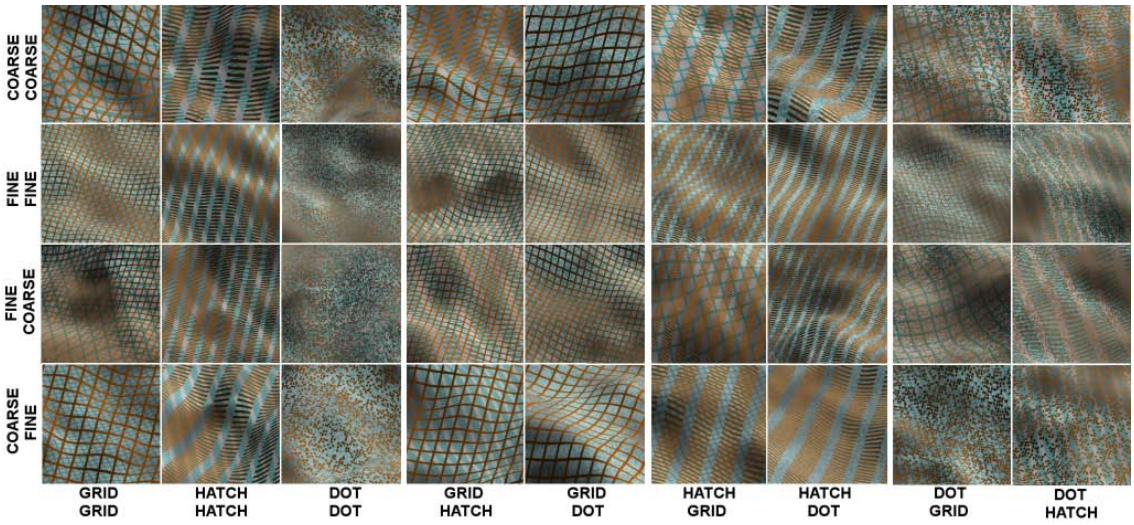



*Figure 83. Structure Size Combinations. Three texture structure and two size combinations on two surfaces.*

The surfaces and viewing conditions were similar to the experiments in the previous section. The surfaces were created using sums of Gabor bumps, and were viewed on a high-resolution stereoscope and with a rocking motion to provide motion parallax cues. One change to the surface shape was that the Gabor bump sizes were chosen from a fractal distribution instead of a flat distribution because many natural phenomena are fractal [Mandelbrot 1982]. The fractal function had three layers. The first layer had 40 bumps whose amplitudes and periods were about 20% of the surface length. The second layer had 160 bumps sized at 10% of the surface length, and the last layer had 640 bumps at 5% of the surface length. The bumps were all randomly placed. A full-

surface view is shown in Figure 84. Texture colors were chosen to be bright, saturated, and with hues opposite each other on the color wheel. The probe was given a yellow hue to contrast with the two surfaces. The coarse grid cells and hatch marks repeated across the surface 64 times, making them about 0.7 cm in size at a distance of 85 cm, and the fine repeated 128 times, making them about 0.4 cm. These sizes correspond to about 0.5º and 0.25º of arc, and meant that the grids had frequencies of about 3 or 6 times the frequency of the smallest surface features. The dot textures were made much smaller. The dot marks were placed on a grid five times smaller than the grid and hatch textures, and then jittered to randomize the texture. The top texture was rotated 15 degrees clockwise, and the bottom texture was rotated 15 degrees counterclockwise to give a relative rotation of 30 degrees.

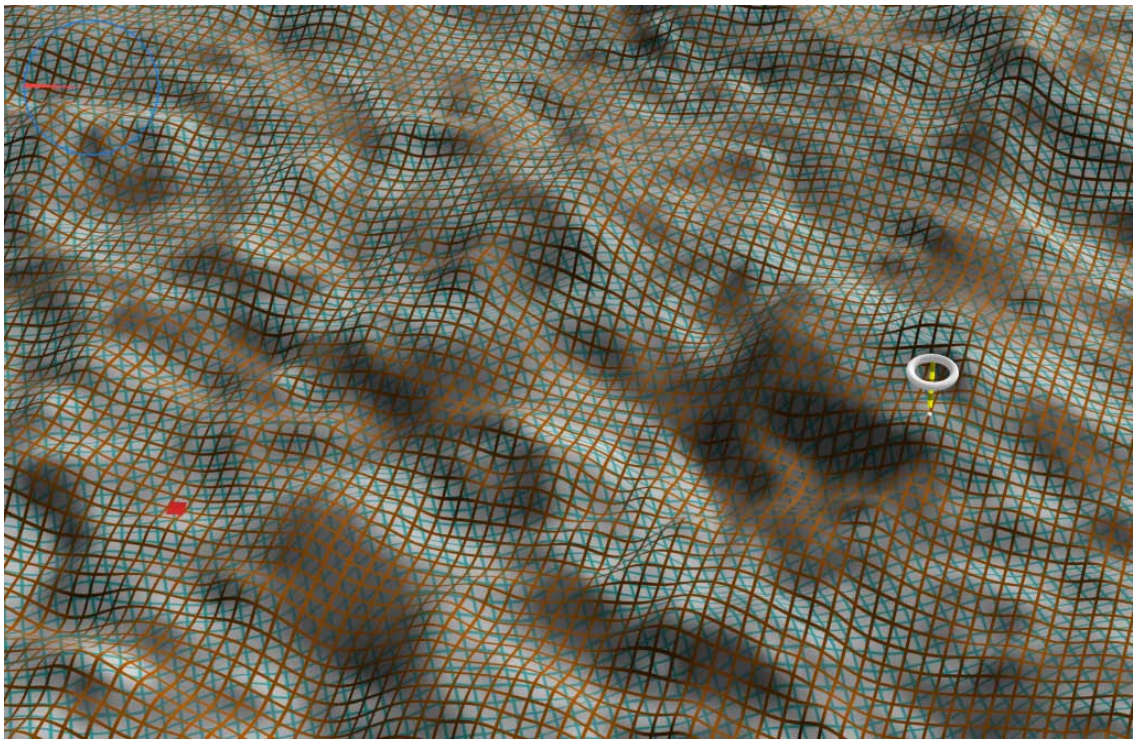


*Figure 84. Surfaces and Probe Example. An example of the surfaces used, showing the probe and an error measure.*

Ten subjects completed all four sessions of the entire experiment. Because of the length of the experiment, three others did not complete all the sessions, and were not included in the results. The subject pool included undergraduate and graduate students as well as two professors. The design used 32 probes for each of the 36 combinations, equaling a total of 1152 probes. The experiment was split into up to four sessions to reduce fatigue, because the total number of probes took about 1.5-2 hours to complete. The task was to align a probe to the surface normal, as in the last section. The probe design was also similar to previous experiments. One improvement that was made was putting a small sphere at the base of the probe so that subjects could not use occlusion of the probe column with the surface as a direction cue. Subjects were trained before running actual trials. During training, they were asked to align the probes to the surface normal, while a diagram on the screen showed their angular error. Figure 85 shows an example of two probes and the accompanying angular error diagrams. After training, the angular error diagram was only displayed after the subject had set the probe, so it could not be used as an aid in setting the probe.

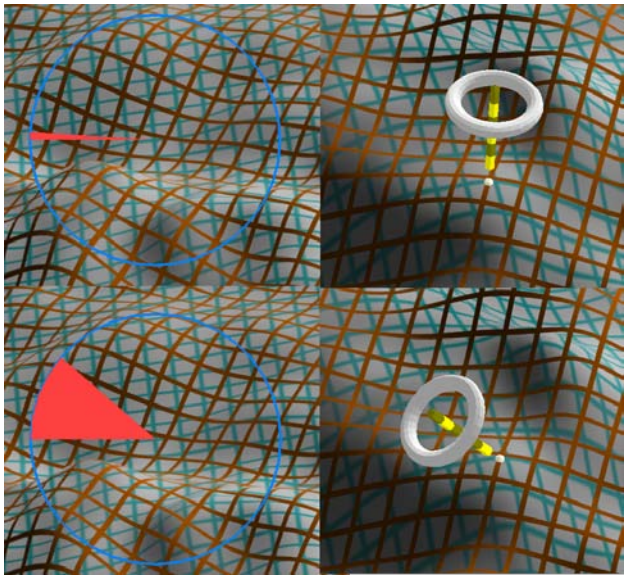


*Figure 85. Probes and Errors. Shows the probe direction and the angular error for an accurate probe (top) and an inaccurate probe (bottom).*

### 6.3 Results

The data from each subject was collected, including the angular error for each probe, the position and normal of the surface at the probe location, and the texture structures and sizes. Before running ANOVA or other more complicated analysis techniques, the data was sorted by the mean angular probe error for each combination of top type and size with bottom type and size. The results are shown in Table 4, and give a feel for simple trends.

Simply counting the number of grids, dots and hatches in the different ranges of error, we find that the top quartile has 11 grids, 7 hatches and 0 dots. Meanwhile, the bottom quartile of highest errors has only 3 grids, 3 hatches, and 12 dots. Thus it is clear that dots are correlated with higher error rates. Counting the texture sizes in the top and bottom quartiles shows more even distributions: 9 fine and 9 coarse in the best textures, and 12 fine and 6 coarse in the worst.

*Table 4. Textures Ordered by Average Error. Textures ordered by average error.*

| **Texture parameters** | **Average error (degrees)** |
|---|---|
| top course hatch , bot fine grid | 10.86 |
| top course grid , bot course hatch | 11.63 |
| top fine grid , bot course grid | 11.67 |
| top course hatch , bot course grid | 11.84 |
| top course grid , bot fine grid | 11.93 |
| top fine hatch , bot course grid | 12.00 |
| top fine grid , bot fine grid | 12.09 |
| top fine hatch , bot course hatch | 12.16 |
| top fine hatch , bot fine grid | 12.19 |
| top fine dots , bot fine grid | 12.25 |
| top course dots , bot course grid | 12.26 |
| top course grid , bot course grid | 12.26 |
| top course grid , bot course dots | 12.53 |
| top fine hatch , bot course dots | 12.54 |

*Table 4. Continued.*

| Texture parameters | Average error (degrees) |
|---|---|
| top course hatch , bot fine hatch | 12.58 |
| top fine grid , bot course dots | 12.62 |
| top fine grid , bot fine hatch | 12.73 |
| top fine grid , bot course hatch | 12.80 |
| top course hatch , bot course dots | 12.81 |
| top course dots , bot fine grid | 12.85 |
| top fine hatch , bot fine hatch | 12.93 |
| top course hatch , bot fine dots | 12.99 |
| top course dots , bot course hatch | 13.04 |
| top course grid , bot fine hatch | 13.13 |
| top fine hatch , bot fine dots | 13.14 |
| top course dots , bot course dots | 13.20 |
| top course hatch , bot course hatch | 13.21 |
| top fine dots , bot fine hatch | 13.22 |
| top course grid , bot fine dots | 13.31 |
| top course dots , bot fine hatch | 13.32 |
| top fine dots , bot course grid | 13.62 |
| top fine grid , bot fine dots | 13.76 |
| top fine dots , bot course dots | 13.91 |
| top fine dots , bot fine dots | 13.97 |
| top fine dots , bot course hatch | 14.27 |
| top course dots , bot fine dots | 14.83 |

The recorded probe angular errors were also split into two groups, depending on whether they were from a probe on the sop or bottom surface. It was found that these errors were correlated with $r = 0.5$. This is interesting, because it suggests that texture conditions affect both top and bottom error. That is to say that textures that are good for one surface are likely to be good for the other surface. This is explored more in the following analysis.

To find out more fine-grained rules for the textures, ANOVA analysis was run on the dataset. The five factors used were top texture type, bottom texture type, top texture

size, bottom texture size and subject. The dependent variable was the angular error measurement for each probe. The ANOVA analysis was run three times on three data sets, where the data sets were overall error, top probe errors and bottom probe errors. Separating the top probe errors from the bottom probe errors allows the analysis to take into account when changes in textures affects one surface but not the other. The results are described below, and in a series of figures. Except for a few cases, results are only shown for significance values below 5%. Where this is not the case, the results are indicated to be not significant, and the p value is shown.

Figure 86 shows the errors for different top grid types. A clear trend is visible for the entire dataset, and the top and bottom surface probe datasets, as shown from left to right. Dots are clearly worse than both grids and hatches. Hatches have a slightly lower mean error than grids, but the difference is not significant for any of the data sets. The result that top dots perform poorly for probes on the top surface is not a very surprising result. The lack of several important texture cues in dots makes them likely to work poorly compared to textures with structure. However, an exciting and interesting result is that dots on the top surface actually appear to cause worse error for bottom probes. This signifies that placing random dots on the top surface makes this surface difficult to see through. Possibly the lack of structure make it more difficult to visually segment the parts of the image that apply to the top surface from those that apply to the bottom surface. The reader may perhaps see this for herself looking at the columns in Figure 83 that have dots on the top.

This result also explains the correlation, noted above, between bottom and top average errors. Top grid and hatch textures are good for both top and bottom probe errors. Another trend made clear from the graphs is that the average bottom error is higher than the average top error. This is similar to previous experiments that found higher bottom errors. It seems that seeing the bottom surface is a harder problem, probably because the brain must reconstruct the surface shape from fragmented information due to occlusion by the top surface.

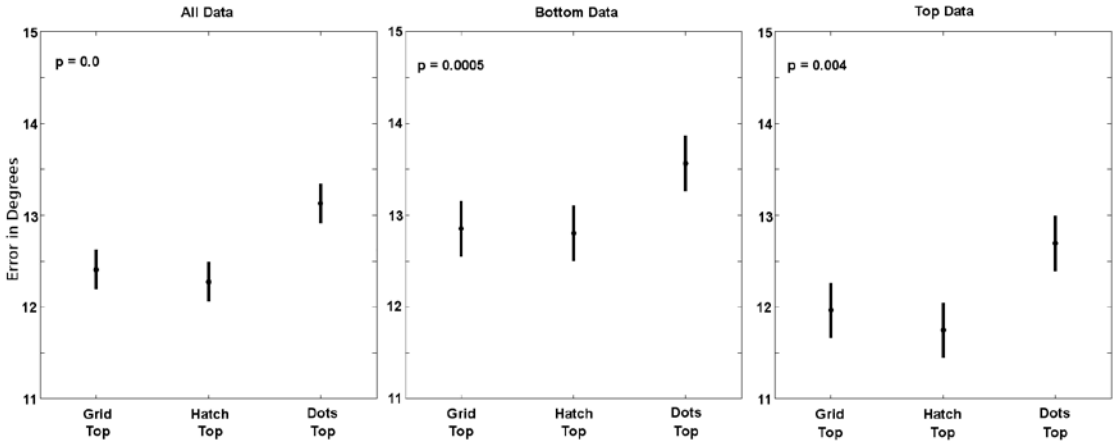



*Figure 86. Top Texture Type Effects. Effect of top texture type on different surface errors.*

In the middle graph of Figure 87, a clear pattern shows that using a grid texture on the bottom surface leads to lower error in estimating surface normal on the bottom surface. The combined dataset on the left also shows a significant effect of bottom type, but based on the other two graphs this is due almost entirely to the bottom data. The top data on the right does not show a significant effect for bottom texture type. This implies that bottom texture type does not interfere with top probe errors the way that top textures interfere with bottom probe errors. In this case, hatch and dot textures do not have significantly different error rates. Apparently, although both grids and hatches are good for the top, only grids should be used on the bottom. These results definitively show that the subtle, randomized bottom textures shown to be good in the multi-parameter experiments, where the tasks were either subjective or feature recognition, do not work well with an objective surface normal task. Here, although the same average difference between top and bottom average errors is seen, bottom error rates for bottom grids are as good as the best-case top error rates. Apparently using the right texture (grids) on the bottom makes all the difference for bottom error rates, though the top texture type has some effect as discussed above.

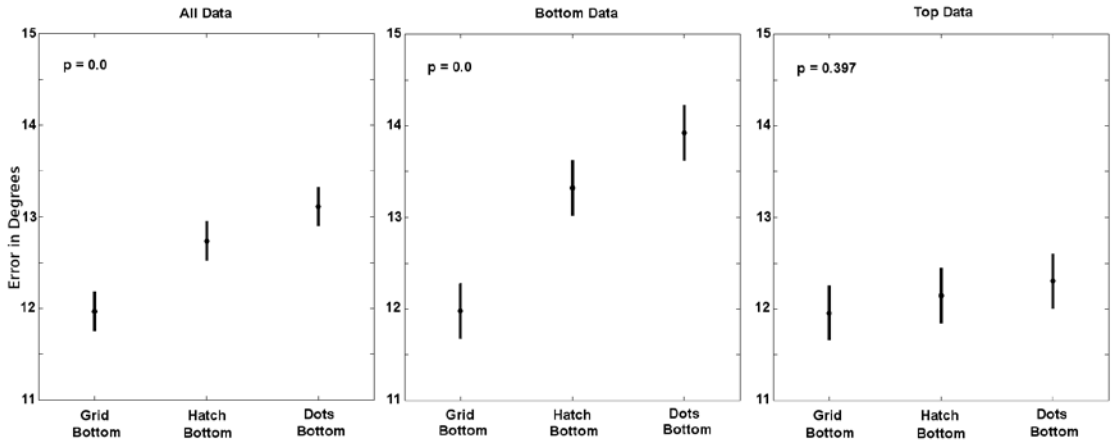



*Figure 87. Bottom Texture Type Effects. Effect of bottom texture type on different surface errors.*

In another example of top texture parameters affecting the bottom probe errors, the left-most graph in Figure 88 shows that coarse top textures gave significantly lower errors overall for bottom data. However, there was also a significant interaction between top type, and bottom and top sizes for the bottom surface data. The middle graph shows that for top grids and hatches (already shown to be optimal), both coarse and fine tops work rather well, as long as both the top and bottom are not fine. However, the two best grid textures had coarse tops, agreeing with the results from the last section that larger top textures are helpful for seeing the bottom surface. Although the interaction was not significant at the 5% level (p=0.051), the rightmost graph shows that fine top textures seem to be better for top error. Clearly, looking at the middle interaction graph, fine textures should not be paired for both the top and bottom, but other combinations seem to involve trade-offs. Coarse top grids and hatches are the best cases for the bottom error, but fine top grids and hatches apparently work best for the top error.

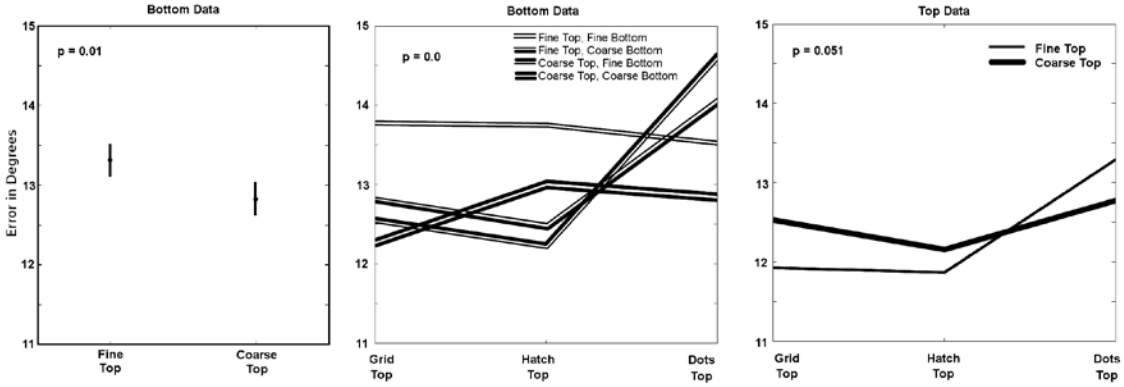



*Figure 88. Top Texture Size Effects. Top texture size main effect and interactions.*

The left-most graph in Figure Bottom Size shows that a coarse bottom texture works significantly better for bottom data than a fine texture. However, a significant interaction between bottom type and size must be taken into account here. From the middle graph, it is clear that the poor performance of fine bottom textures is almost entirely due to the inclusion of dot textures in the data set. We have already learned that dots are a very poor choice for bottom texture, and therefore texture size decisions should not be based on performance for dots. The effect of size on bottom error for grids is at best minimal; though coarse bottom hatches seem to do slightly better than fine ones. However, since hatches were not found to be useful bottom textures, we can say that bottom size does not appear to affect the bottom probe errors for the optimal case of a bottom grid. Interestingly, there is a possible interaction between bottom texture size and top texture type for top probe data. Although the interaction is not significant at the 5% level (p=0.063), the right-most graph shows that top grids have about 0.5 degrees less error with a coarse bottom texture. Top hatches perform slightly worse for coarse bottoms, but the effect is so small as to be likely not important. So for optimal texture types of either grid or hatch on top and grid on the bottom, a coarse bottom texture is unlikely to hurt, and may help top error.

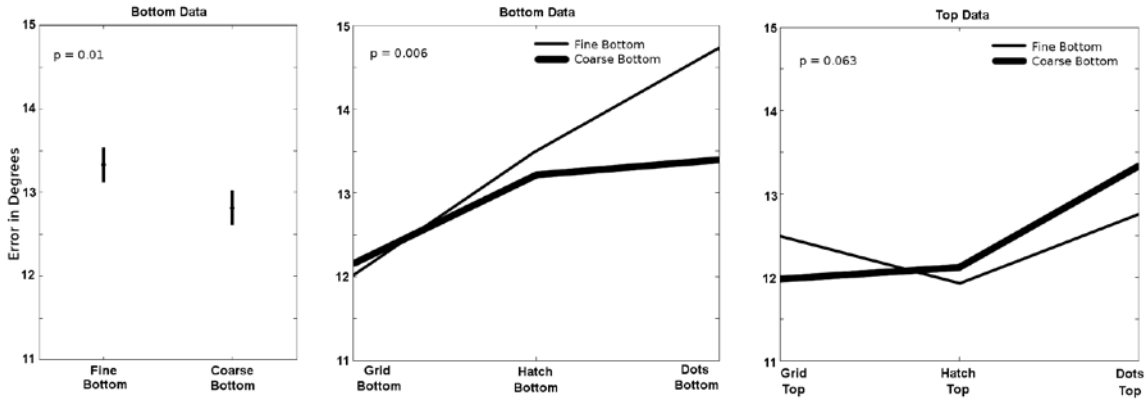



*Figure 89. Bottom Texture Size Effects. Bottom texture size main effect and interactions.*

## 6.4 Discussion

For the purpose of comparison, we present the best and worst-case texture type combinations. Figure 90 shows all size combinations of Grid-Grid and Hatch-Grid surfaces. Note that the fine-fine column is not recommended according to the optimal size results. For contrast, the worst-case Dot-Dot surfaces are also shown.

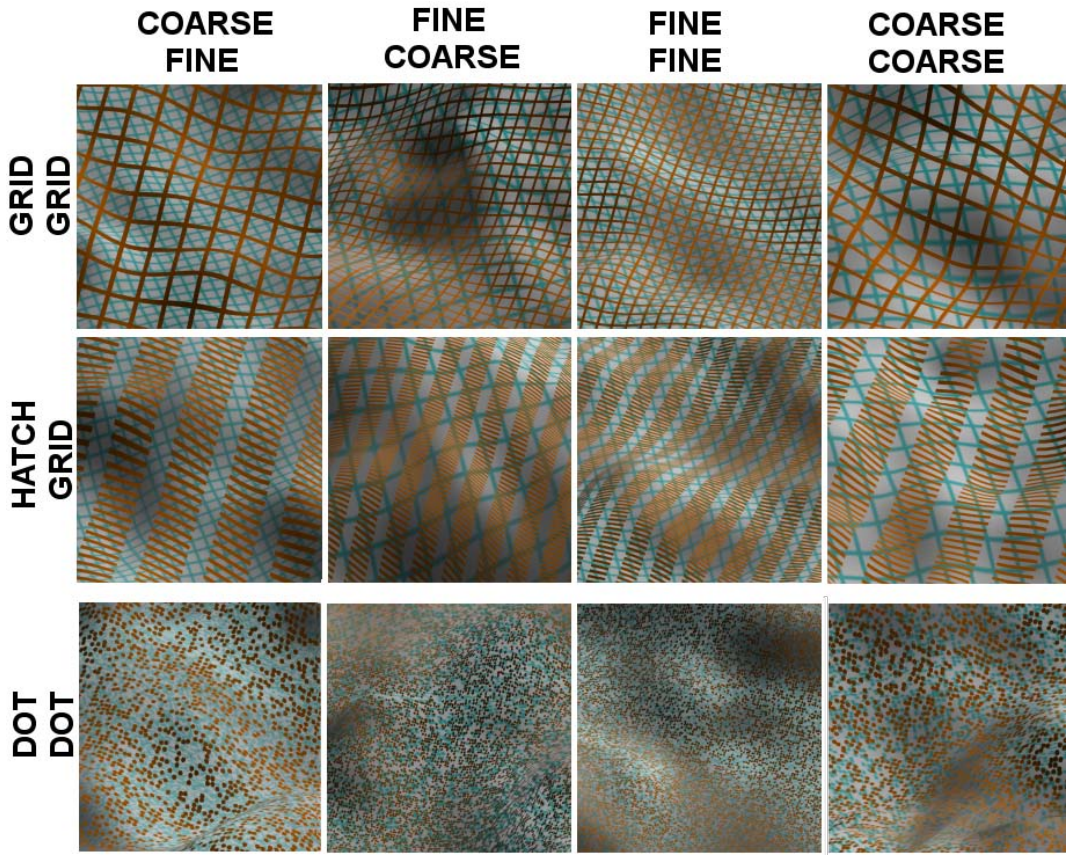



*Figure 90. Rule-Based Best and Worst Cases. There is a clear difference in visibility of the two surfaces between the 8 best-case examples (two top rows) and the 4 worst-case examples (bottom row).*

Lastly, we compare the textures with the overall best mean error in Figure 91. The combination with the best mean error has a coarse top hatch with a fine bottom grid. Interestingly, the second best case has a coarse top grid and a coarse bottom hatch, even though overall hatches on the bottom did not work as well as grids. The third best has a fine grid over a coarse grid, and the fourth has a coarse hatch over a coarse grid.

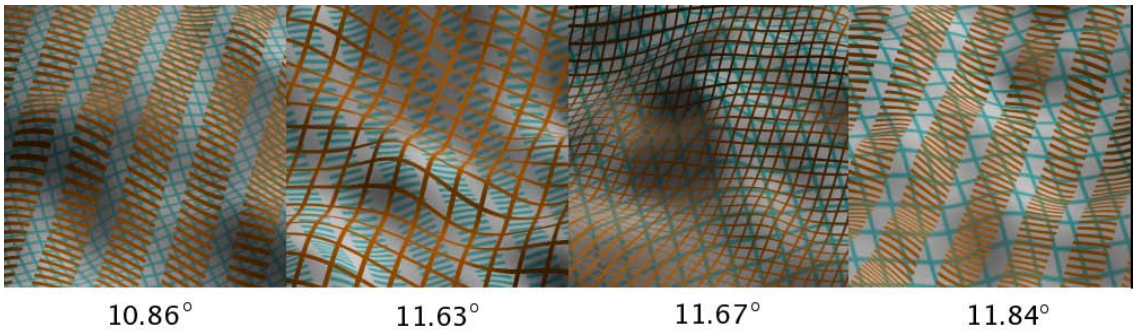



*Figure 91. Average Best Cases. Coarse Hatches over fine grid performed the best.*

The worst-case average errors are shown in Figure 92. They all involve dots; in fact the only texture type not a dot is the bottom hatch in the second-worst texture. Since in print this texture actually shows the bottom surface rather well, we assume that the hatch marks distracted from the very fine top dots making the top surface more difficult to see. Also, the case of large top and bottom dots is not here, ranking as the 11$^{th}$ worst texture. It would seem that if dots absolutely must be used, it works better to make them large.

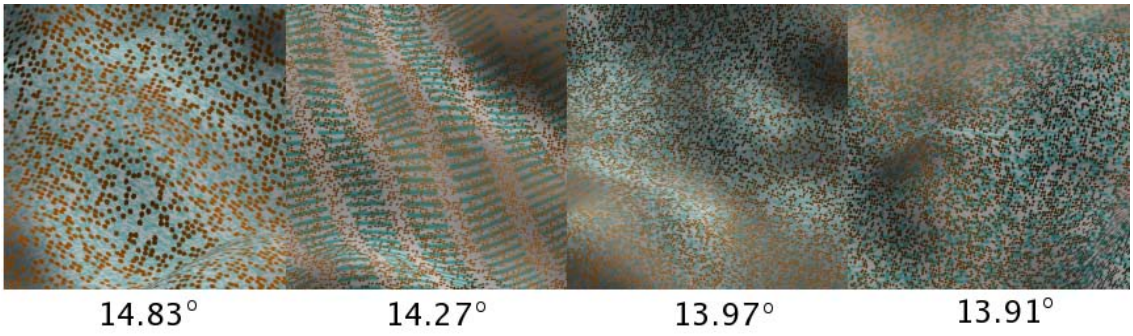



*Figure 92. Average Worst Cases. The worst average error cases ordered from left to right.*

In conclusion, for estimating surface normals with layered textured surfaces, of the three texture types tested, the overall optimal bottom texture is a grid, while the top should be either a grid or hatch. Results for optimal size of grids were not as clear, but it

does seem that two fine textures should not be combined. As found in the previous study, coarse top textures worked best for showing bottom data, and a non-significant effect suggested that fine top textures were better for showing the top data. Bottom texture size had no significant effect for bottom grids, but a coarse bottom texture might be useful in reducing shape error on top grids. These results are consistent with those found in the last section.

Almost as interesting as the effects and interactions that were found, were ones that did not come across as significant. For example, there was no interaction at all found between texture types. It seems that choice of type for one surface does not affect the proper choice for the other surface. If this is true in general, it simplifies the problem of layered texturing.

In this experiment, since the dots are randomized, it is not clear if the shape of the dots or their organization causes the poor error rate. If dots were organized in a grid pattern for instance, they might well form illusory contours that would work similar to the grid or hatched lines for showing surface shape. However we highly doubt that arranging dots could ever do better than the lines.

From here, two more experiments were designed. The first, in Section 7, tries to measure how useful layered surfaces might be with more than two layers. The second, in Section 8, looks at texture direction on single surfaces.

# 7. MULTIPLE LAYERS FEASIBILITY EXPERIMENT

## 7.1 Experimental Design

The previous four sections described experiments designed to find optimal texture configurations for the visualization of layered surfaces. Several metrics were used to measure how well the layered surfaces could be perceived. For example, it was found certain textures that worked well with a feature finding task did not work well with a surface normal estimation task. However, all of the experiments used only two layered surfaces. Certainly, in real applications there might be cases where more than two layers are required. For example, in a geological visualization, several sediment layers might need to be shown simultaneously. But the are no results indicating the scalability of the layered surface problem. Intuition suggests that adding more layers will make the visual problem progressively more difficult, if only because more information will be forced into a limited number of pixels. To begin to understand the effect of multiple layers, we conducted a feasibility study.

The feasibility study was relatively simple. The design compares performance in surface normal estimation for a single surface, the top and bottom for two surfaces, and all three layers of a three-surface visualization. These six cases were compared using standard analysis of variance tests. The surfaces were height fields made from sums of Gabor bumps as introduced in Section 5. The task for the subjects was to estimate the surface normal using a probe introduced in the same section. Error was measured as the angle between the probe direction and the actual surface normal.

Texturing the surfaces for this experiment naturally required some extrapolation of the guidelines found in the previous experiments. The top and bottom surfaces could be textured using the guidelines, but was not obvious how to texture the middle surface so as to maximize its visibility while minimizing conflict with the surfaces above and below. Certainly, the middle surface must have some of the qualities of the top surface that make it easy to see through, combined with the qualities that make the bottom surface clear through distracting upper surfaces.

All textures were grids because grids were shown to work best on both top and bottom in the texture structure experiment that compared grids, hatches and dots. The texture grids were rotated 0º on the bottom surface, 22.5º on the middle surface, and 45º on the top surface. This gave a maximum rotational difference between pair of surfaces, since rotating a grid an angle θ than 45º is equivalent to a rotation of 90º - θ. The images were displayed on a high-resolution stereoscope as described first in Section 4. The surfaces were rocked to provide motion cues, with a rocking amplitude of 8º and a period of 2 seconds.

Although the previous experiments suggested that larger openings on the top surface are preferable for seeing the bottom surface, to remove any bias toward a particular surface, grid sizes were varied in this experiment. Since certain combinations of grid sizes seemed to involve tradeoffs, we did not want to choose a specific combination for three surfaces. Similarly, although hue has not been shown to be important in any of the studies, the texture hues were varied with each presentation. Figure 93 shows three variations in color and size, with nine total texture combinations. Colors were chosen to be bright and with noticeably distinct hues. The grid spacings had 40, 80 and 140 lines along the edge of a surface, corresponding to roughly 1.5 cm, 0.7 cm and 0.3 cm seen at a distance of 85 cm. Since the smallest surface shape features were roughly 2 cm in size, the coarsest grid may be pushing the limit of how widely spaced the lines can be before information is lost.

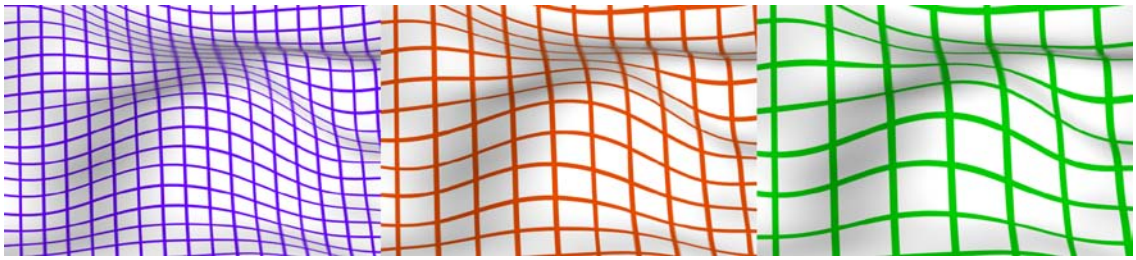


*Figure 93. Grid Colors and Sizes. Three colors (blue-violet, orange and green) and three grid sizes (small, medium and large) were chosen for the grids on each surface.*

The surfaces were shown in single, double or triple layers. Colors and grid spacings of different layers were constrained to be different, giving 36 possible combinations of color and spacing for the three-layer case. Figure 94 shows close-ups of example color and size combinations on one, two and three surfaces. Full surface images are shown at the end of this section, in Figure 100. The textures were calibrated so that the overall opacity for the top layer was 50% in the two-layer case, and both the middle and top were calibrated so that their overall opacities were 33% in the three-layer case.

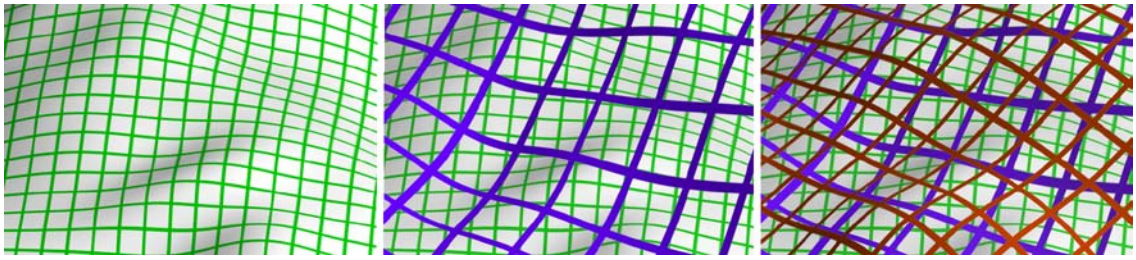


*Figure 94. Single, Double and Triple Layers. A closeup example of the textures used on one, two and three layered surfaces.*

36 probes were measured for each of the six cases: single surface, the bottom and top of two surfaces, and the bottom, middle and top of three surfaces. The surface normal probes, shown in Figure 95, were very similar to the probes introduced in Section 6. One small difference was the color of the probe base. The white base used in the size-structure experiment was hard to see against the white texture background of the bottom surface. Therefore, the base used here was black so that there was high contrast. Before the actual experiment, subjects were trained to use the probe with error-based color-coding and the correct surface normal (seen in the figure as a thin green cylinder). After training, subjects aligned a total of 216 probes (36 for each of the six cases) to the perceived surface normals, which was done in a single session of about half an hour. Five students were subjects in the experiment, one female and four male. All had normal or adjusted to normal vision and were tested for stereo and color blindness.

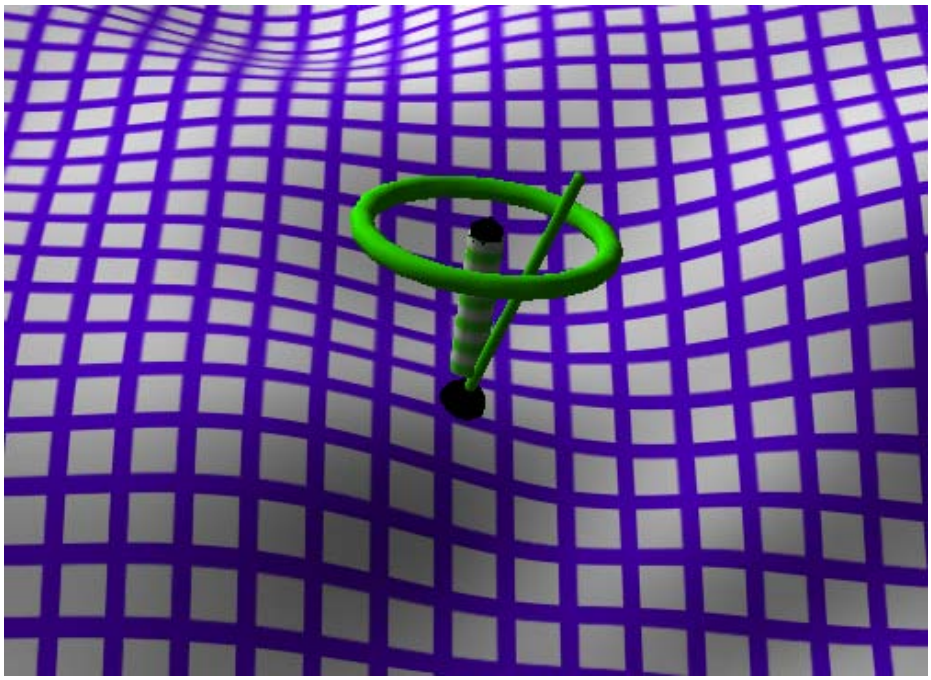


*Figure 95. Probe Design. The probe used by subjects to estimate surface normal shown with the correct surface normal.*

### 7.2 Results

As expected, the experimental errors were worse on average for two surfaces than for a single surface, and for three surfaces than for two. The average errors with 95% confidence intervals are shown in Figure 96. The results are encouraging. The single surface case has an average error of about 8.5º, the best case. However, it is only significantly better than the two worst cases, the bottom and top surfaces of the three-surface case. It is interesting to note that the middle surface actually had lower errors than the bottom and top in the three-surface case. The error rates for two layers fell between the single and three-surface values, though the only significant difference was that the bottom of the two-surface case had lower error than the top of the three-surface case. The time spent on each surface was similar to the error rates, except much less significant. This means that as more layers were added, subjects were taking slightly longer as well as getting higher errors.

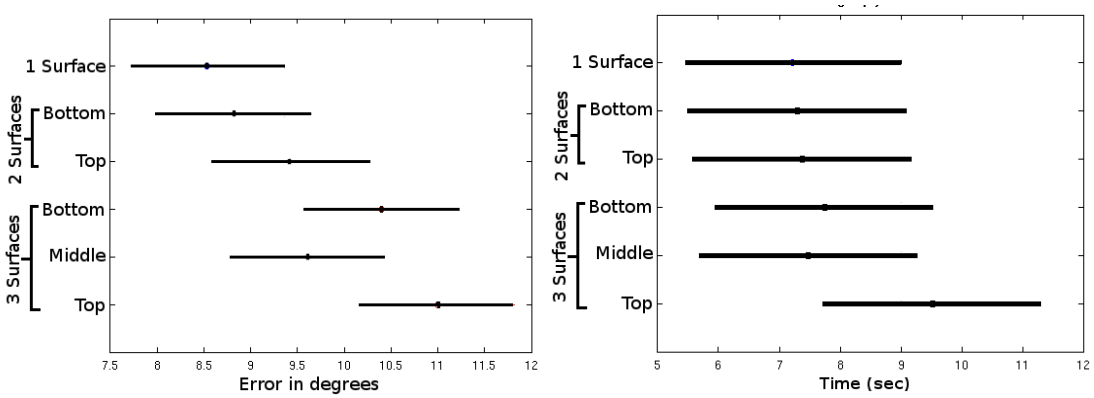



*Figure 96. Errors and Times for Multi-Layered Surfaces. A multicomparison plot of the different errors and times for each surfaces under the case of 1, 2 or 3 surfaces.*

The measurements for one, two and three surfaces give us three data points to extrapolate performance on more layers using quadratic approximation. Figure 97 shows several extrapolation curves. Linear, exponential, quadratic and cubic curves were fitted using least squares. The zero information line is the error rate that would occur for the surfaces in this experiment if all probes were simply aligned with the 'up' direction. To minimize intra-surface penetration for the three-surface case, surfaces were made slightly flatter than in previous experiments. Therefore, average errors for subjects were slightly better, and the zero information line occurs at about 20º instead of 24º, as in previous experiments.

Naturally any function could be used to fit three points, since three points is such a small sample. Based on the figure, the higher order polynomial fits suggest that six layered surfaces is very unlikely to provide any useful surface shape information. On the other hand, the optimistic linear fit allows for 13 surfaces before the visualization becomes useless. The truth is likely to be somewhere in between. There were two reasons for only comparing single, double and triple surfaces, thereby giving only three points to fit. The first was evaluation time. The experiment was designed to include all combinations of color and grid size to eliminate biases. Including a fourth surface would have increased the number of combinations from 6x6 = 36 to 24*24 = 576. This was an unacceptable amount of combinations to test in a short experiment, and we wanted to

keep the unbiased experimental design so that relative errors on different levels could be compared. The second reason was interactivity. It was very important that the rocking motion and surface normal probe be of reasonable speeds. On our hardware, adding a fourth surface would reduce interactivity to a questionable level. The only fix would be reducing the size or resolution of the rendered images, neither of which were desirable in this case.

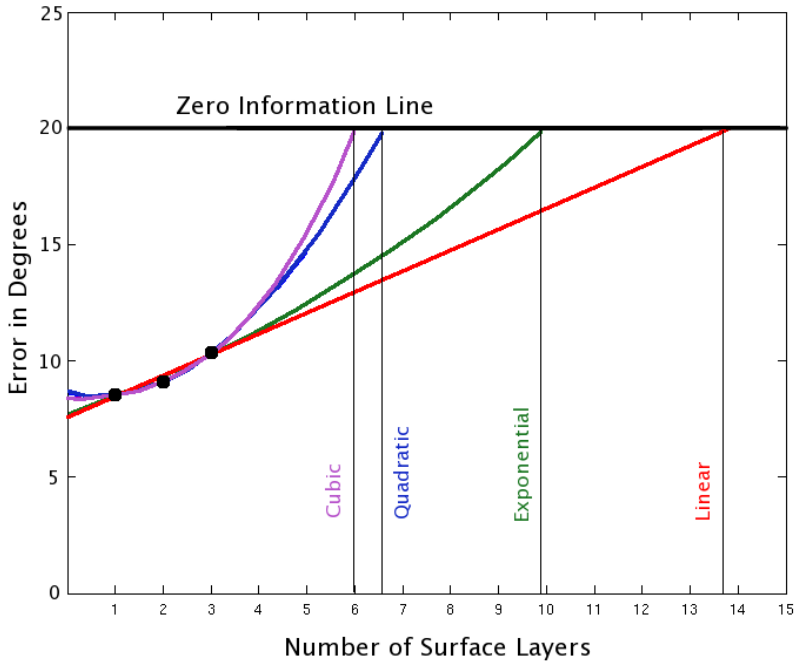



*Figure 97. Least Square Fits and Extrapolations. Different functional least-square fits to the data.*

Since six surfaces seemed to be a possible limit to layered surface perception, a six-surface visualization was made, and is shown in Figure 98. Certainly it is a rather confusing image at first glance, but the author was pleasantly surprised to be able to see many of the surfaces as coherent shapes. For some reason, the green surface is rather difficult to see in this case. However, the success of this six-layer image was not measured using any user studies. It merely suggests that the increase in error may not be as fast as the quadratic and cubic curve fits.

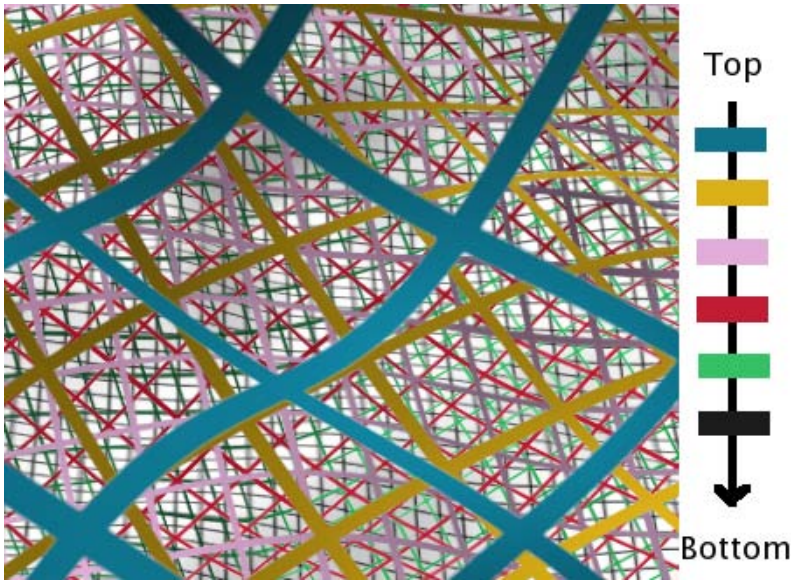



*Figure 98. Six Surface Layers Example. Each surface was given a unique color and grid size.*

The feasibility experiment was not designed to provide enough data to do a thorough exploration of effects of grid size and color on error rates. However, for the case of single surfaces, each of the nine size and color combinations had four identical repetitions. As a result, there was enough data to check for significant results for size or color effects. Considering results from previous research and knowledge about human perception, we were not surprised to find that color had no significant effect on error, with p=0.45. However, as seen in Figure 99, the mid-sized grids performed significantly worse than both the large and small grids. The effect was rather strong, with p = 0.0036. One theory for why the mid-sized textures did not work as well is that subjects could be using different techniques when looking at the small and large grids. As can be seen in Figure 94, the small grid cells are nearly planar, whereas the large grid cells are clearly not planar, so subjects must interpolate normals for probes not on a line crossing. It is possible that the middle-size grid cells work poorly because the cells are not as obviously non-planar, and subjects might tend not to interpolate surface normal

information when necessary. More research must be done to fully understand how these results fit with human perceptual theory.

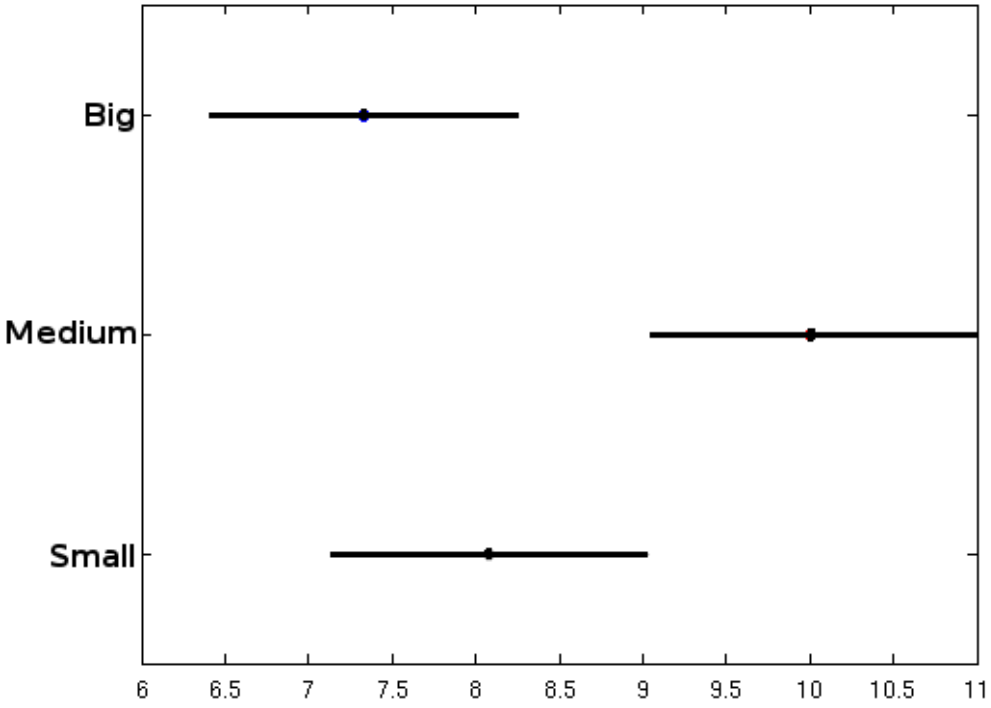



*Figure 99. Single Surface Texture Size Effect. The medium grids are significantly worse than both the big and small grids. (p=0.0036).*

### 7.3 Discussion

Since this is a feasibility experiment, one technical problem should be noted. In order to maintain an interactive frame rate with three layered surfaces, as shown in Figure 100, on our hardware, the number of polygons in the surfaces had to be reduced to 200x200 instead of 400x400. Although hardware speeds at the time of this writing are still increasing without any hard limits in sight, rendering speeds could certainly be an issue for multiple layered visualizations. Although 200x200 polygons were more than sufficient to display the features on our surfaces, many scientific datasets are larger by

orders of magnitude. The compositing operations required for rendering transparent surfaces significantly increases the rendering time. Therefore, rendering speed should at least be considered as a factor along with the perceptual limits on layered surface perception.

Based on this experiment, it seems that several layers, at least as many as five, could feasibly be used in a layered surface visualization without an extreme loss in clarity of the surface shape. Because of the difficulty in understanding some of the results from the previous experiment, we wanted to run a final experiment that tried to analyze more carefully the mechanics of how texture helps with surface perception. Although we have used projected textures in all of the studies, it has been argued that textures that follow the principal curvature of the surface are superior for showing surface shape [Interrante and Kim 2001]. Therefore, the final study compares textures with marks in the principal curvature directions with parallel line textures like the grids we have been using.

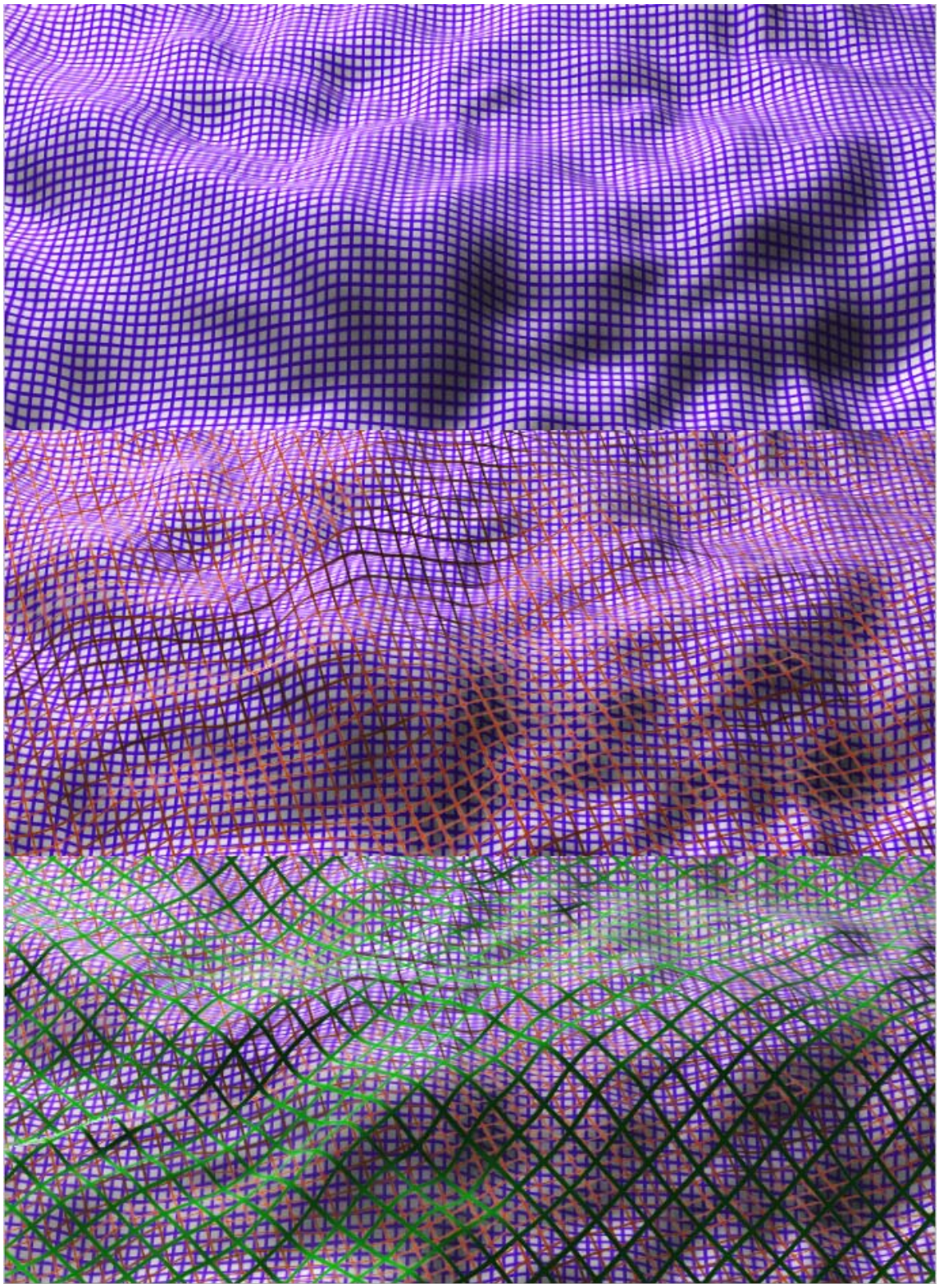


*Figure 100. Single, Double and Triple Layers Full View.*

# 8. TEXTURE DIRECTION ON SINGLE SURFACES

## 8.1 Introduction

The final experiment in this line of research is designed to focus on how textures are actually used by the human visual system to derive surface information. The experiment uses single surfaces with simple shapes. The metric used to measure surface shape perception is surface normal estimation. This metric was used successfully in several of the previous experiments, and appears to be a good measure of shape perception. The surfaces are sampled with normal probes densely spaced relative to the surface shapes and the texture frequency to get fine-grained error information. The main variable considered is a comparison of projected grid textures with principal direction textures. The goal is not simply to compare the overall performance, but investigate if and how different features of the two texture types work to show surface shape. Textures that have a directionality or 'flow' have benefits over isotropic textures. The argument has been made several times that principal direction textures are the 'gold standard' for texture direction. However, our experience with projected grid textures leads us to believe that they may perform just as well in many cases.

In the literature, lines in the direction of maximum curvature have been shown to be better for shape perception than lines in the direction of minimum curvature or intermediate directions for single-layered, developable surfaces [Li and Zaidi 2000; Li and Zaidi 2001]. Interestingly, a model described in Knill [2001] showed that strongly oriented textures with marks along parallel geodesics allow shape from contour information to be used on developable surfaces even for geodesics not parallel to the direction of maximum curvature. Also, Todd and Oomes [2002] showed that the principal direction restriction was only valid under a limited set of viewing conditions. Mamassian et al. [1996] compared observer shape choices of small surface patches with a Bayesian model and found a bias toward interpreting the curves as lines in the direction of principal curvature for doubly curved surfaces. However, this analysis only looked at local patches with six curves to show the surface shape, and therefore no

global information was available, such as parallel correspondence information. Both principal direction and uniform direction textures can give information through line junctions, but projected textures have a global organization that does not exist in principal-direction textures.

No experiments so far have directly compared principal direction textures with projected textures, but a few studies have compared similar kinds of textures. Interrante and Kim [2001] compared line-integral-convolution textures using either the first principal curvature direction, a random direction, a uniform direction and coherently-varying vectors that are not associated with the surface shape. The results showed that the principal curvature and uniform directions worked better than the other two, with principal curvature working slightly better with monocular viewing and uniform working slightly better with stereo viewing. Interrante et al. [2002] showed that principal directions worked better than either swirly or uniform directions under a wide variety of lighting conditions using a feature identification task. However, half of the presentations used a viewing angle directly above the surfaces meaning that very little contour information was available from the uniform direction lines. Due to the small subject pool (3 subjects), the lack of training provided to the subjects, and the fact that the unidirectional textures were not truly projected textures, we felt that further investigation comparing projected grids and principal component textures was warranted.

One possible issue with projected textures is that since they stretch with the surface, they are not truly homogeneous. This means that the grid shapes will be elongated on surface patches orientated away from the projection direction. These can be thought of as the sides of hills and valleys. Also, line junctions will not be perpendicular if both lines have significant components down-slope. The effect is most exaggerated when the surface normal is perpendicular to the projection direction.

### 8.2 Experimental Design

Despite several studies, little is definitively known about how texture direction affects perception of doubly-curved surfaces. It is known that having directional textures is better than not, but that textures with a random direction do not help and may hurt. Beyond this, it is not clear exactly what shape cues can be gotten from projected or principle direction textures. Because we wished to focus on the performance of textures in perception of surface shape, his experiment is limited to a single surface.

Three parameters were varied in this experiment, namely texture type, texture rotation and camera rotation. Figure 101 shows all eight of the texture possibilities. The primary variable was texture type: either principal direction textures or projected grid textures. Projected grid textures have been used in several of the previous experiments. However, principal direction textures are more complicated because they use lines that follow the principal curvature directions of the surface. Aside from texture type, two other variables were introduced to keep the experiment from being biased by a particular viewpoint or texture rotation. Texture rotation and camera viewpoint were varied, each with two levels. Since the surfaces were relatively constrained to always have two hills and two valleys in a particular relationship, it seemed important to vary both the texture orientation relative to the surface and the viewpoint relative to the surface independently. Each had two rotation cases: 0° and 45°. For the texture rotation, the projected grid texture was rotated about the surface center. Since the principal direction lines are defined by the surface curvature, they do not change with texture rotation, but the *location* of line junctions changes to match the rotated grid texture's line junctions. That is, the seeding grid for the principal direction textures was rotated by either 0° or 45°. For the camera viewpoint, the camera was rotated counter-clockwise around the surface center. This is equivalent to rotating the surface itself clockwise. We choose to think of the rotation as a camera motion so that both the texture and camera motions can be thought of relative to the surface.

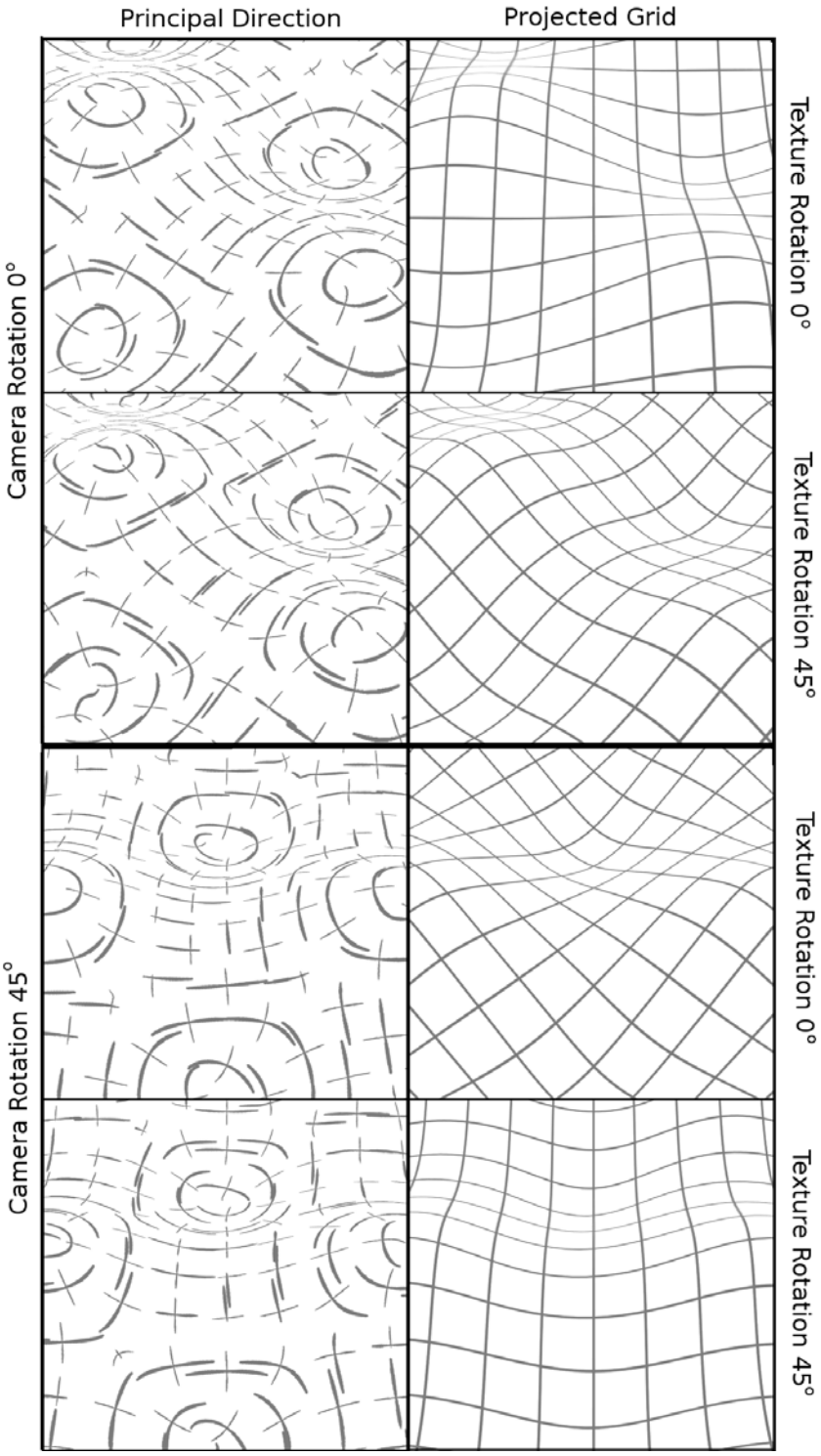



*Figure 101. Texture Type and Rotation Combinations. The set of eight textures used in the experiment.*

The surface shape cues were limited as much as possible to be only texture line cues for this experiment. The surface was tiled and the bumps were shallow enough that surface edge and occluding contour cues were not available. Shading was not used. Subjects viewed the visualizations on the high-resolution stereoscope first used in Section 4, so stereo cues were used. Depth information from stereo is actually duplicated to some extent in compression of the grid size in the case of projected grids, and line length in the case of principal direction textures. Motion was not used because motion of the surface of any kind would change the angle the subject was viewing the surface. We wanted to be able to compare errors with the exact surface orientation relative to the viewpoint. Perspective projection was set to be correct according to the distance between the subject's eye and the monitor. Therefore, the perspective cues were consistent with the stereo cues for depth.

After a training period, each subject was presented with a surface in a random order. For each surface the subject rated 81 probes (9x9 grid). Probes were presented in a random order, and only one probe was shown at a time so that previous probes would have a minimal effect in subsequent probes. With each probe, the camera was translated relative to the tiled surface so that the probe is in the center of the camera's field of view. The grid of probes had a higher frequency than the projected texture, to allow comparison of probe location relative to the texture lines. The total number of probes run for each subject was 81 surface probe locations by 2 camera rotations by 2 texture rotations by 2 texture types = 648. In previous experiments, subjects took about 5-7 seconds to align each probe. Subjects were chosen based on their previous experience with computer graphics as well as their known dependability in order to get good quality data. The experiment took about 1.5-2 hours for each subject to complete, with time included for training. Subjects completed the experiment in either two or three sessions over the course of a few days.

### 8.3 Surface Construction and Tiling

Several of the previous studies used randomized surfaces composed of many superimposed bumps. Here, since the goal was to understand how the textures work under controlled shape presentation, the surfaces were highly simplified. Locally, surface shape at any point can be described using only the maximum and minimum principle curvatures. Depending on these curvature values, the surface might be planar, cylindrical, hill/valley or saddle. Texturing of surfaces in the case of planar and cylindrical shapes is relatively well understood [Knill, 2001; Knill, 1998; Saunders and Backus 2006], so we concerned ourselves only with the hill, valley and saddle cases.

Hills and valleys were created using Gaussian functions. Figure 102 shows the parameters used to create a Gaussian bump, and a shaded example. The two-dimensional function has a center position in x and y, falloff parameters $\sigma_x$ and $\sigma_y$, a rotation $\theta$, and an overall height.

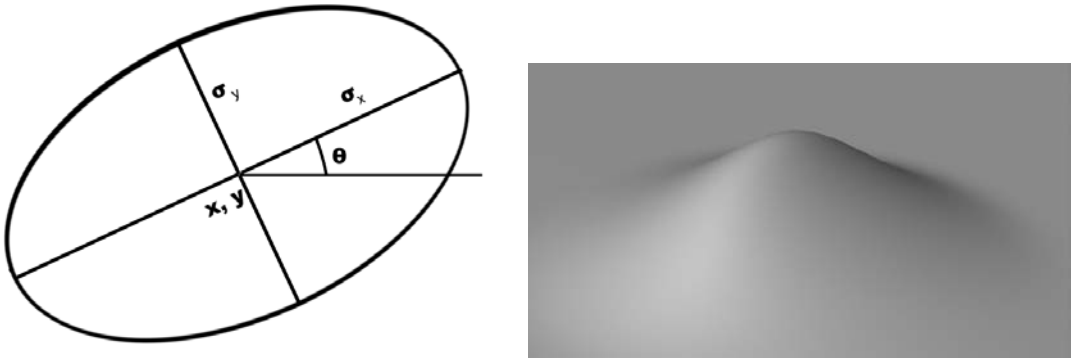



*Figure 102. Gaussian Bump Diagram and Example. Parameters that make the Gaussian bumps used for hills and valleys, and example shaded bump.*

The surface was designed to be as simple as possible, but still contain hills, valleys and saddles. Figure 103 shows a surface designed using superimposed Gaussians that has two hills and two valleys, with the center being a saddle. Note that the design is tiled beyond the extent of the central surface so that surface edges are not visible. This

was to keep subjects from using edge cues to estimate surface normals. Each hill or valley uses the six Gaussian parameters described above with randomization. The range of parameters was chosen so that the hills and valleys overlap slightly, and the height to width ratios are small enough that occluding contours are not seen from the camera viewpoint. The bump centers (x, y) were allowed to vary by $1/8^{th}$ of the surface width and depth. Bump angles were chosen randomly. The Gaussian falloff $\sigma_x$ was chosen to be between $1/8^{th}$ and $3/16^{th}$ of the surface width to ensure that the bumps and valleys overlapped, and $\sigma_y$ was chosen based on an aspect ratio between 1.0 and 1.5. The hill and valley heights varied between 0.16 and 0.26 of the surface width, making the maximum slope about 35°.

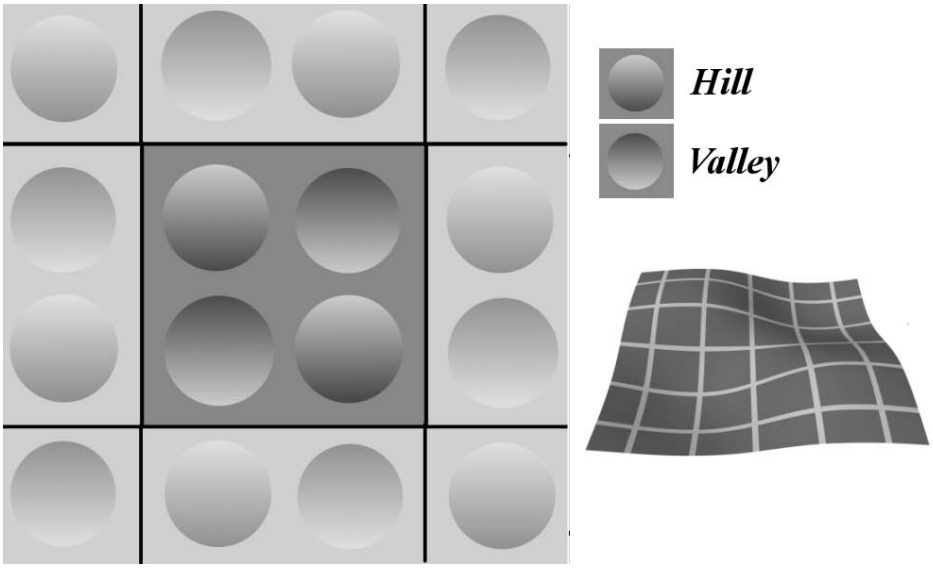



*Figure 103. Hill, Valley and Saddle Surface Tiling. Diagram of the surface from above. Bumps represent hills, dips represent valleys, and the spaces between are saddles. A single tile is shown to the right as it would be rendered with perspective projection and a grid texture.*

The simplification of the surface allows for better analysis of where errors occur for the different texture types. Finding where the errors occur, and any trends in the errors could provide significant clues toward what assumptions people make for each

texture type to reconstruct the surface. Also, the low frequency surface shape allows the lines to be spaced more widely than in previous experiments. This allows testing of how placement of the probe relative to lines and line junctions affects performance, and could give clues as to how humans interpolate the information they get from the lines.

### 8.4 Probe Design and Sampling

Surface normal probes were used as in the previous three studies to estimate perception of surface orientation. To get a good understanding of how humans use individual lines to judge surface shape, it is necessary to sample a large number of probes relative to the number of lines on a surface. Figure 104 shows an example of the placement and density of probes used. In the actual experiment, the probes were presented to the subject one-by-one to minimize the tendency for nearby probes to bias the local surface normal estimation. The probe design was nearly identical to that used in the previous experiments. The one change that was made was that the location of the probe was shown on the surface as a dot rather than a sphere. Using a sphere could possibly let subjects use the sphere-surface intersection to estimate the surface normal rather than using surface and texture cues. Use of a dot eliminated this possible ‘cheat’.

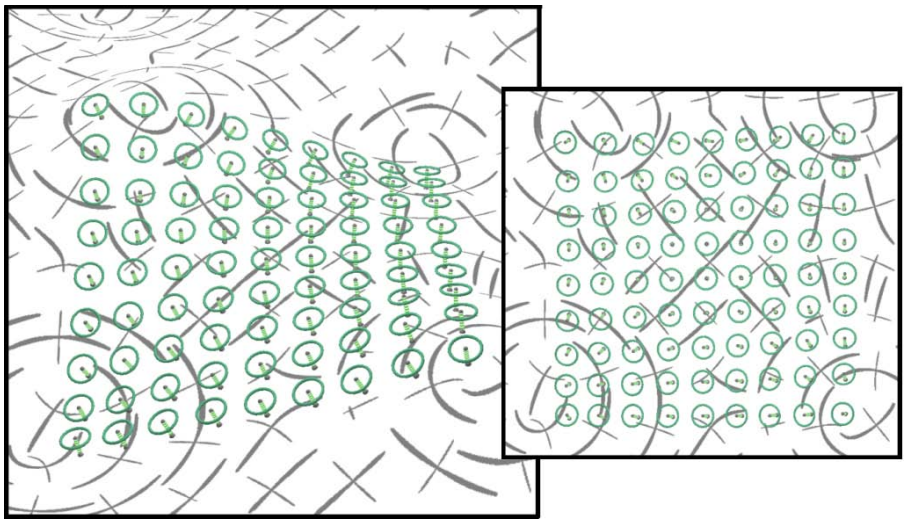


*Figure 104. Surface Normal Probe Area. Area to be covered by probes shown from a 45° and top view.*

### 8.5 Principal Direction Texture Design

Two main types of textures were compared: projected grids and principal direction textures. In order to truly compare the different texture types, they were made to be as similar as possible to avoid bias from unexpected variables. The grid lines were spaced 10 across each surface tile, so there were about five grid lines along each hill or valley giving what should be more than sufficient information about surface shape. The lines were thin, covering 5% of the texture area, to leave plenty of space between line junctions. This was done to permit examination of whether proximity to a line junction made a difference in error. The colors were chosen to be monochromatic, with medium grey lines and a white background, because hue was not expected to be of any importance.

Several methods were tried for designing the principal direction textures. Figure 105 shows our final solution. Some of the factors that were considered were: using one or two principal direction lines, scaling principal direction line lengths by the local curvature magnitude, and placement of the principal direction marks to best cover the surface. These three factors are actually related, and our motivation for the final texture design is as follows. Several studies have shown that bi-directional textures are better than unidirectional textures for many surface shapes. Since the grid textures were bi-directional, it was decided that the both principal curvature directions should be used as well. Another decision that was made was to use individual marks to show the principal curvature vectors rather than a texture-stitching algorithm such as used in Gorla et al. [2003]. The reason for this was that we wanted to compare errors with individual marks. Texture synthesis methods can create textures with ambiguous marks and intersections. For example, marks may be truncated because they run into other marks, or change directions sharply where texture seaming boundaries exist. Quality seaming algorithms will generally guarantee that marks continue across boundaries, but the mark direction may not be as continuous as we would like. By limiting our textures to individual marks, we make it so that probes have clear relationships to individual marks and intersections.

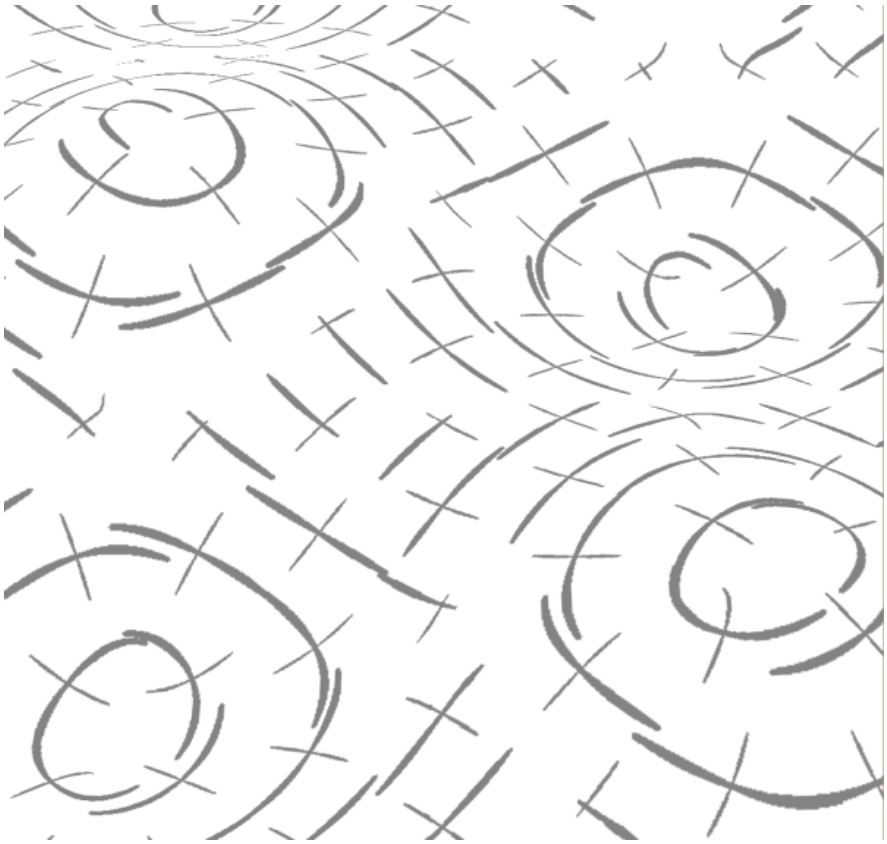


*Figure 105. Principal Direction Texture Example. An example of a principal direction texture created using our design.*

The choice for the scaling and placement of the marks was not as straightforward. Initially, based on Interrante et al. [1997], we tried scaling the principal direction line length by the curvature magnitude to provide information about both the curvature direction and magnitude. However the scale of the curvature magnitudes was very small over a large area of the figure relative to the maximum curvatures in the peaks and valleys. This created a problem because the lines in some areas were too small to see easily. Several functions of the curvature magnitude were tried, such as logarithm, and square root, but none looked very good. It was decided that using a function of the magnitude would negate the purpose of using the length to map the curvature amount.

Another option that was tried was simply adding curvature lines where the curvature was small to increase the coverage. In this method, lines were initially placed

randomly, and the positions were adjusted according to the size and proximity of neighboring lines, using an algorithm similar to Kindlmann and Westin [2006]. However, it quickly became apparent that the number of marks required to fill the relatively flat areas effectively would far outnumber the number of lines in the projected grid textures. Since line junctions are such an important cue, we did not want to bias the results by having many more line junctions in the principal direction textures than in the projected grid textures. Therefore, the decision was made to scale the principal direction lines uniformly across the surface. The maximum curvature was made longer and slightly thicker than the minimum curvature to distinguish the two vectors. With this line length, it was found that placing principal direction vector junctions at the same points as the projected grid junctions produced a pleasing texture. This simple solution removed the location of line junctions as a variable so that only line direction varied between the texture types.

Using two sets of marks, one for the maximum curvature direction and a second for the minimum curvature direction, was also tried. The thought was that allowing the two sets to be independent would give an effect similar to cross-hatching in artistic drawings. However, it was found that for the small number of lines used here, the actual effect was simply confusing. The lines were sparse enough that they tended not to overlap, so it was decided that using a single set of marks with two lines for each junction was a superior method for our purposes.

The principal direction lines were created by integrating backward and forward along both the principal direction vectors at each point. This meant that they curved as the surface curvature changed. One issue came up due to the length normalization of the principal direction vectors. There are certain areas of a doubly curved surface where the directions of the maximum and minimum curvature are discontinuous and flip. These areas are also where the surface is locally nearly planar, so that the principal direction vectors would ordinarily have near-zero length and the discontinuities would not be noticeable. However, when the vectors are normalized, integrating along them gives strange jumps and it is not always clear which direction to follow for the best results.

These abrupt changes in line direction were deemed unaesthetic and potentially confusing for subjects. Therefore, in areas with sudden discontinuities in principal curvature directions, integration of the lines was stopped. This left areas with shorter lines, but since these areas were nearly planar anyway, the overall image was much more pleasing. Finally, the width of the principal direction lines was varied slightly along the length to emphasize the line junctions and make the ends taper off rather than stopping abruptly.

### 8.6 Information from Texture Lines

Stevens, [1981] gives a good review of possible ways to reconstruct surface shape from the 2D line information available in a monocular image. Locally, surfaces can be categorized as either planar or curved. Given that a surface is curved, it can either be singly curved (a developable surface) or doubly curved. Doubly curved surfaces can be convex, concave or hyperbolic depending on their Gaussian curvature measurements. Finally, surface shape can be represented by the surface orientation at each point, though this is dependent on the viewpoint.

Determining surface normal from lines of curvature is an ill-constrained problem, however it can be solved when additional constraints are introduced. We start with the assumption that we have identified two lines that cross in a local region of the image. Given that the lines are texture elements, they can be assumed to lie in the local plane of the surface, and the side-to-side and up vector components are given from the line projections in the image. This leaves two unknowns: the depth component of each of the lines. Under the assumption that the lines are perpendicular, there is only a single unknown left. With this single degree of freedom, the surface normal tilt is constrained to a range of degrees as shown in Figure 106. The possible normals are bounded by the two vectors perpendicular to the line projections. If $\beta$ is the larger of the two angles formed by the lines, the surface normal tilt is constrained to a range of $180° - \beta$. The figure shows examples of $\beta_1 = 125°$ and $\beta_2 = 160°$, where the surface normal tilts are constrained to 55° and 20° respectively.

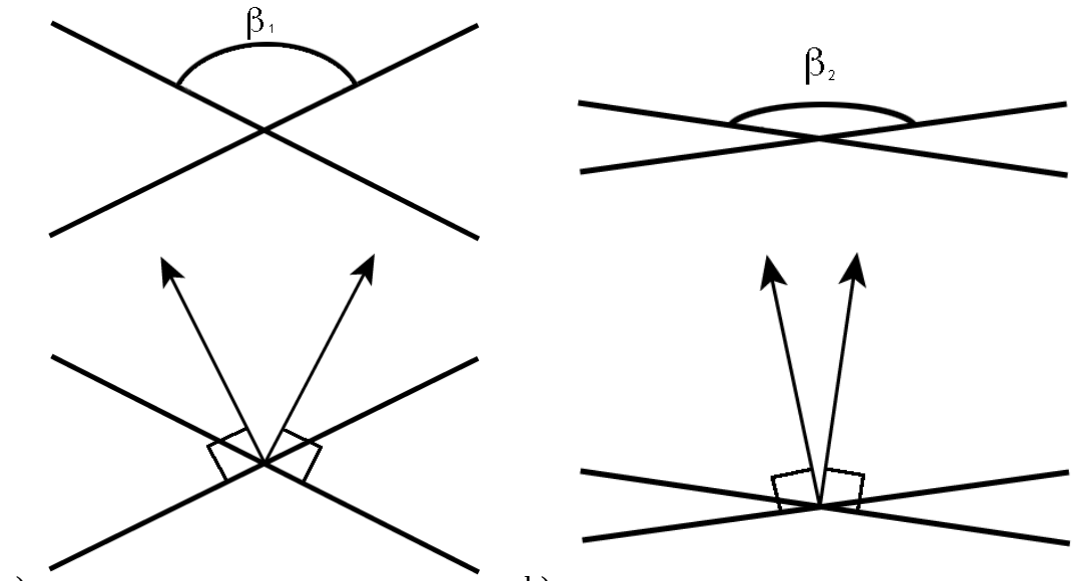



*Figure 106. Surface Normal Constraints. Junctions with significant foreshortening significantly restrict the possible surface normal tilt. The set of possible angles is bounded by the arrows formed by perpendiculars to the crossed lines, as shown.*

In the case of a surface shown with multiple contours, as in Figure 107, if those contours can be assumed to be planar a further constraint is added, and given one surface normal, others can be computed. This method is called parallel correspondence, and can be used to propagate well-constrained normals in some areas to other areas where the texture information does not fully constrain the surface normal. The figure shows how this would be done using one set of parallel lines on a simple surface. V and W are simply tangents to one of the curves at two places in the projected image. The U vectors, which are parallel, are found by connecting points on adjacent curves where the curve tangents are parallel.

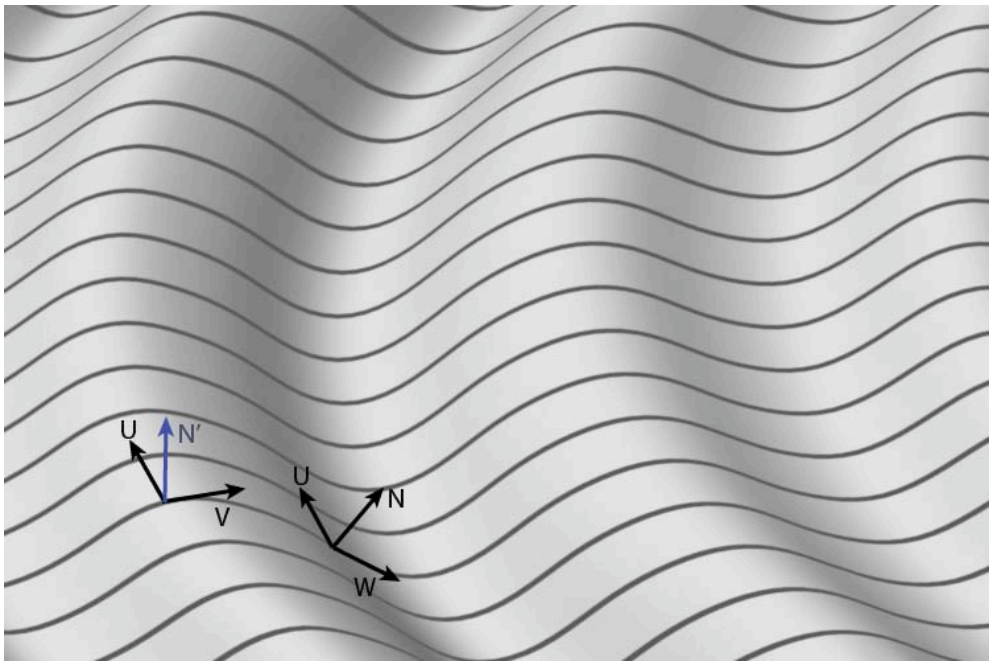



*Figure 107. Parallel Correspondence. Using the constraint that U is equal at both points, an estimate of the normal at the point with more foreshortening can be used to solve for the normal at another point.*

This could in theory be applied to doubly curved surfaces as well as singly curved surfaces. Although Figure 107 is technically a doubly curved surface, in general the U vectors will not be parallel in 3D across the length of a line. Figure 108 shows an example of two U vectors that are parallel in the image plane, but would not be parallel in 3D. However, the difference in depth of the two vectors can be estimated using the changing distance between the parallel lines in the image plane. That is, shorter vectors imply a surface facing farther away from the viewer (foreshortened). Therefore, parallel correspondence can be a useful method for determining the shape of an arbitrary surface.

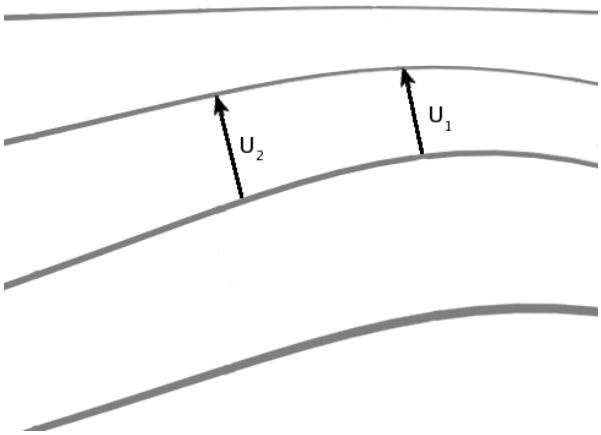



*Figure 108. Doubly Curved Surfaces. $U_1$ and $U_2$ are parallel in the image plane, but $U_1$ has a larger depth component, and so they are not parallel in 3D.*

### 8.7 Data Analysis

The data from the five subjects and eight texture conditions was first analyzed using a multi-variable Analysis of Variance (ANOVA) on camera viewpoint, texture rotation and texture type. Two of the three variables were found to be significant: camera viewpoint and texture rotation, as shown in Figure 109. Remember that for the case of principal direction textures, the rotation refers to a rotation of the mark seeding locations. For both camera angle and texture angle, the 0° case worked better. The implications of this are explored further later. Interestingly, the comparison of texture types showed no significant difference between projected grid errors and principal direction errors. This supports the idea that both projected grid textures and principal direction textures offer sufficient texture information for surface reconstruction.

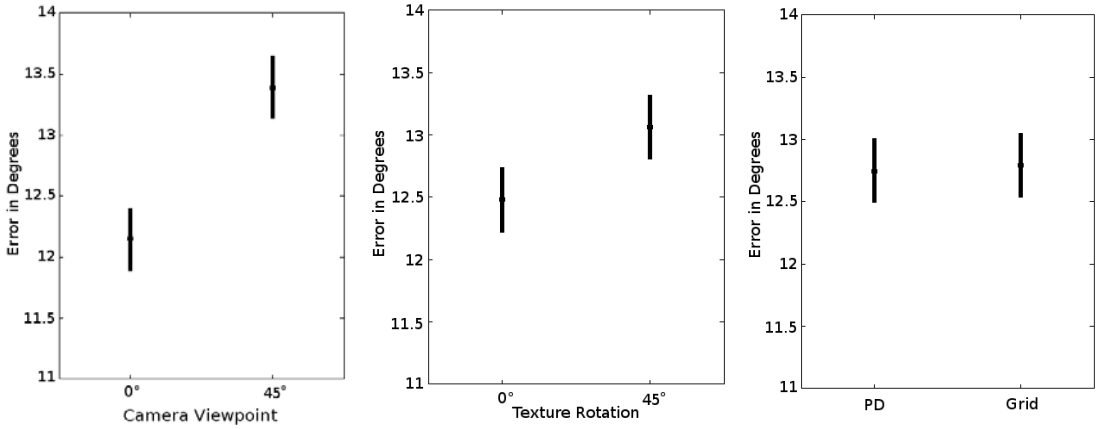



*Figure 109. ANOVA Results. Both camera and texture rotation worked better for the 0° cases. There was no significant difference between PD textures and projected grids.*

As shown in Figure 110, one significant interaction was found: between the camera and texture angles. It appears that when the camera was facing the surface at the default 0° angle, the texture direction did not matter, but when the camera was rotated, the texture rotation had a significant effect. This interesting result is investigated further below, but we will see that it is due to several effects. Although the 45° texture rotation is worse than the 0° rotation with a 45° camera angle for both projected grids and principal direction textures, we expect that it is actually for different reasons.

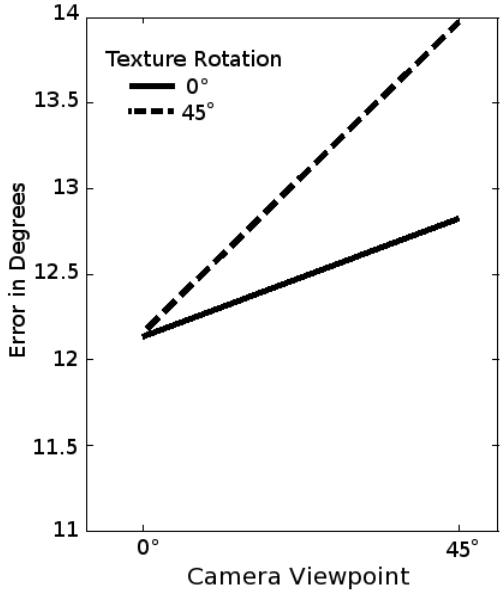



*Figure 110. Texture Rotation and Camera Viewpoint Interaction. Texture rotation only has an effect for a camera viewpoint of 45°.*

To learn more about the specifics of the error directions and what factors generate higher error, the probe data was grouped into five canonical directions that we call left, right, up, middle and front. Because the surfaces were constructed as height fields with slopes no greater than 45°, the set of possible normals lie within a cone facing up, that has a height no greater than the base radius. Since the surface was tipped 45° toward the viewer, the set of possible normals is also tipped toward the viewer. The five canonical angles that were chosen for analysis are shown in Figure 111. Left and right correspond to the surface normal directions with the smallest and largest x components. Up denotes surface areas that are seen nearly edge-on from the observer's point of view. Middle are probes that are straight up in the surface coordinate system, but tilted 45° toward the viewer. Finally, front are surface normals seen facing almost directly toward the viewer.

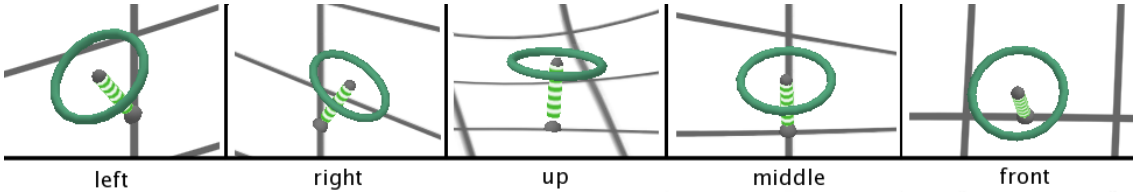



*Figure 111. Canonical Probe Directions. The four extreme probe directions and the middle probe direction.*

Surface normals that fit into each category (within about 15° of the canonical direction) were averaged over all subjects and textures, and are shown in green in Figure 112. The accompanying probe directions were also averaged, and are shown in red. The numbers above each plot are the average errors for the normals in each category. These numbers are not the same as the angle between the normal and probe directions shown because those directions are averages themselves. Therefore, the plots show the average error direction, but the numbers above are a better measure of average error magnitude. All vectors are shown as they would appear from the point of view of the subject. One pattern is immediately clear: subjects tend to over-rotate the probes away from the up-direction, and this over-rotation is larger the farther the true surface normal is from vertical. However, subjects seem to be very accurate about the rotation in the x-z plane, since all the surface normal-probe direction pairs are very nearly coplanar with the up-vector. If we translate the traditional definition of slant and tilt to mean tilt away from, and slant around the up-vector, this means that subjects overestimate tilt, but judge slant very accurately. We note that the only remotely non-coplanar vector pairs are the middle and front vectors. This is investigated later in comparing the camera direction differences.

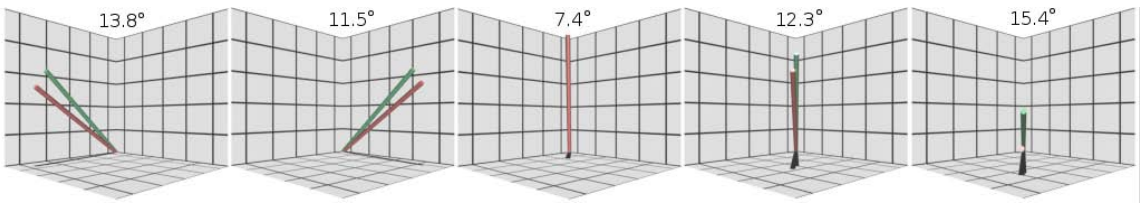
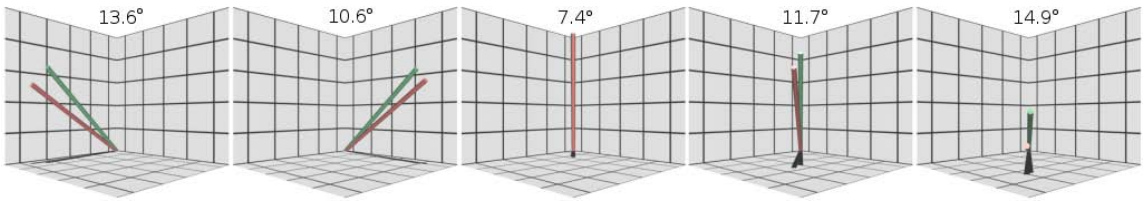
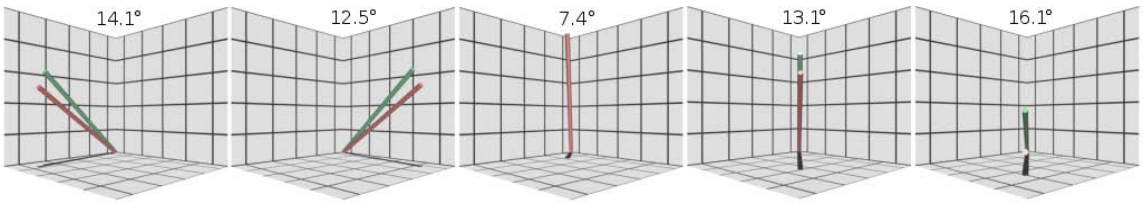



*Figure 112. Average Errors in Canonical Directions. The average normal (green) and probe (red) directions are shown for each of five canonical angles: left, right, up, middle, and front. The numbers above each plot are the average angular error.*

In the above ANOVA analysis, significant error differences were found for different camera rotations, and different texture rotations. Since it was not immediately obvious why one camera or texture orientation might be better than another, the data was split up by these groupings, and the vector averages were computed for each rotation setting. The results for camera rotation are shown in Figure 113. Two major differences can be seen between the 0° and 45° viewpoints. First, the 0° case has a left-right bias seen in larger errors and a leftward lean for left normals and smaller errors on the right, and a leftward lean in the middle and front vectors. The 45° viewpoint does not have this left bias, and the difference in left and right probe error magnitudes is half as large. However, the 45° view does have larger overall errors in every direction except up, which account for the significantly worse error rate in the ANOVA measurement.




a) 0° Viewpoint




b) 45° Viewpoint

*Figure 113. Camera Viewpoint Errors in Canonical Directions. Canonical vectors are shown separately for a) 0° and b) 45° camera rotation.*

The reasons for these differences can be understood by looking at the locations of the surface normals from each camera viewpoint. First, we look at the stronger left-right bias seen in the 0° camera angle data. Figure 114 shows approximate locations for probes that would fit the left, right, up, middle and front normal directions for both of the camera rotations. Note that ‘middle’ is actually shown five times because this surface normal is found on all peaks, valleys and saddles. One thing that can be seen is that there is actually a left-right bias in the surface itself as seen from the 0° viewpoint. This is because the saddle is seen as having positive curvature going from the front left to the back right, but negative curvature from the back left to front right. If the middle bias is a result of this, then there is a tendency to rotate the probe too far down hill in the direction of positive curvature in the case of a saddle. The biases in left, right and front surface normals can be accounted for by noting that the sampling of these vectors is biased to be closer to the central saddle than to the tiled (and opposite direction) saddles on each side and to the front and back. The up vector had no noticeable bias, likely because it was unaffected by saddle curvature because the surface normal was so well constrained by foreshortened line crossings.

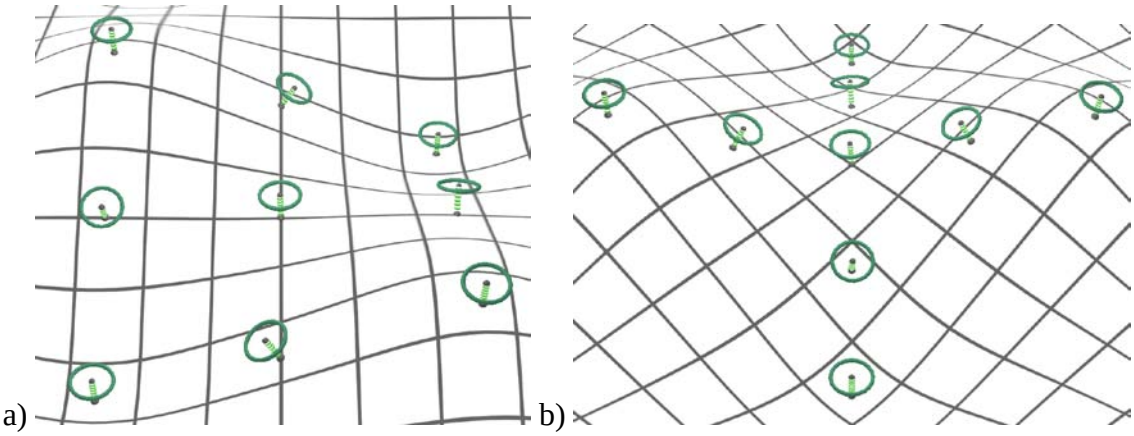
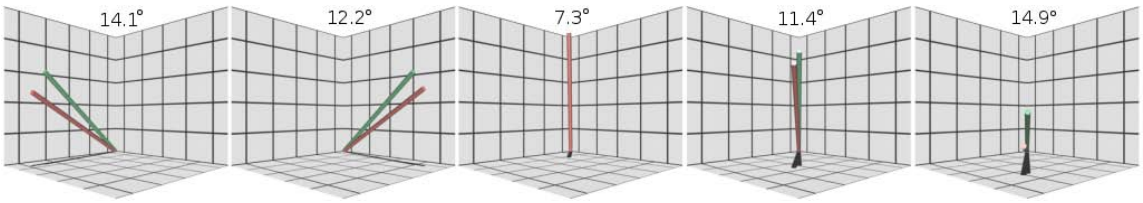


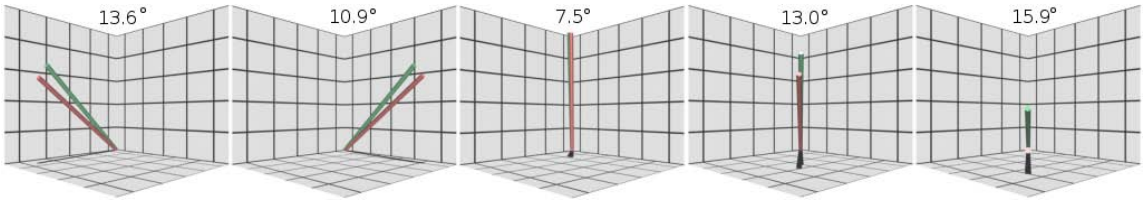

*Figure 114. Canonical Direction Dependence on Camera Direction. Examples of left, right, up, middle and front probes for the two camera rotations of a) 0° and b) 45°.*

Also, there is a noticeable difference in magnitude of error between the 0° and 45° camera rotations, for all probe directions except the up-direction. Interestingly, this was not due to an increase in the average tilt toward the viewer; the average surface tilt was almost identical for the 0° and 45° cases. What is different about the two viewpoints is the overall direction of changes in the surface. In the 0° case, the surface changes rapidly in both the side to side and up directions from the subject's point of view. In contrast, the 45° case has stronger variation along the line of sight, something that does not show up well in the curvature of the projected lines. As a result, the uncertainty associated with forward tilting surface normals is harder to constrain.

Although an ANOVA found no differences between the texture types, analysis of error magnitude for the five canonical directions showed that there were a few differences between the texture types. Figure 115 shows that principal direction textures generally had larger errors for left and right surface normals, while projected grid texture had larger errors for middle and front-facing normals. Both cases had a left-right bias as found above.




a) Principal Direction




b) Projected Grid

*Figure 115. Texture Type Errors in Canonical Directions. Canonical vectors are shown for principal direction and projected grid textures.*

The result that projected grids do not work as well for middle and front-facing normals suggests that subjects might incorrectly read the line junction distortions produced when lines are projected on slopes. If subjects were reading the line junctions as perpendicular, it would explain the larger over-estimation of the surface normal rotations. However, the projected grids worked better for the left, right and up cases, suggesting that in these cases subjects were correctly reading the curves as parallel projected lines, which gives more information than the individual principal direction line junctions. For comparison, Figure 116 shows an example principal direction texture and a projected texture.

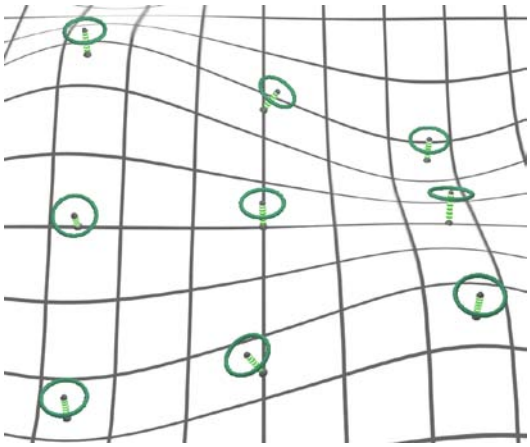
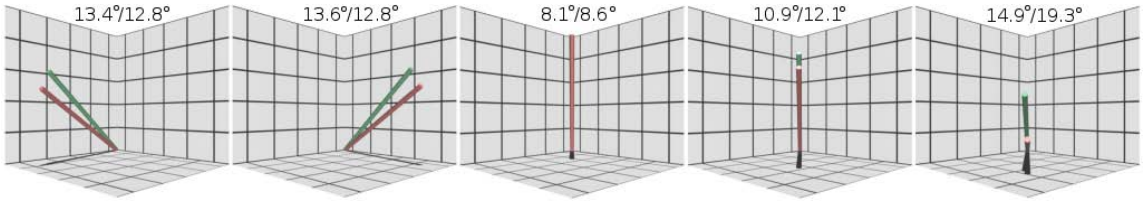
a)  b) 

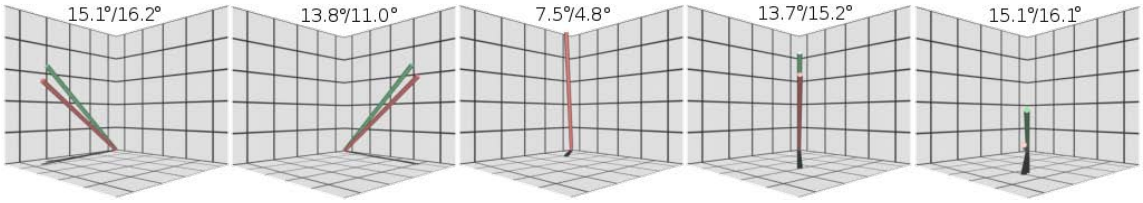
*Figure 116. Canonical Probes for Different Texture Types. Canonical probe directions are shown for comparison between principal direction textures and projected grid textures.*

The last factor that had a significant ANOVA difference was texture rotation. In the interaction analysis, we found that for the 45° camera rotation, a 45° texture rotation was significantly worse than a 0° texture rotation. Figure 117 shows the canonical vectors for the two texture rotation cases for the 45° camera angle. The numbers above are separated by texture type (principal direction / projected grids) because there were interesting differences in the values. The 0° case works significantly better for the middle surface normal direction. We think this is true for both the projected grids and the principal direction texture because the projected grids offer angled lines to better constrain the surface normal, while the principal direction marks are better placed because of the interaction with the saddle area and the texture rotation. The left and right normal directions are similar, though there is a strong left-right bias for the case of projected grids with a 45° texture rotation and a 45° viewpoint. We traced this bias back to two surfaces with significant height differences between the left right bumps, where the two subjects had extreme left-right biases. Therefore, we will treat this difference as surface-specific and non-significant. There is also a difference between the up-direction errors for 0° and 45° projected grids; the 45° case is better. From looking at the images, we guess this might be due to some interaction between the line curvature and the concave surface curvature in the area. Finally, the projected grids are much worse for the

front facing surface normals. We presume that this is because the angled lines exaggerate the downward tilt of the surface, and make the subject more likely to over-rotate the probe.




a) 0° Texture Rotation, 45° Viewpoint




b) 45° Texture Rotation, 45° Viewpoint

*Figure 117. Texture Rotation Errors in Canonical Directions. Canonical vectors are shown for texture rotations of 0° and 45° where the camera has a 45° rotation.*

Figure 118 shows the four texture-rotation and texture-type combinations. In line with the error results, we would say that the surface in a) seems more strongly tilted in front than in c), although the surfaces shown in the figure are identical. Also, the principal direction marks toward the center are nicely spaced in b), but overlap unattractively in d). The overlapping of the marks could have made foreshortening of lines more difficult to judge, contributing to increased errors.

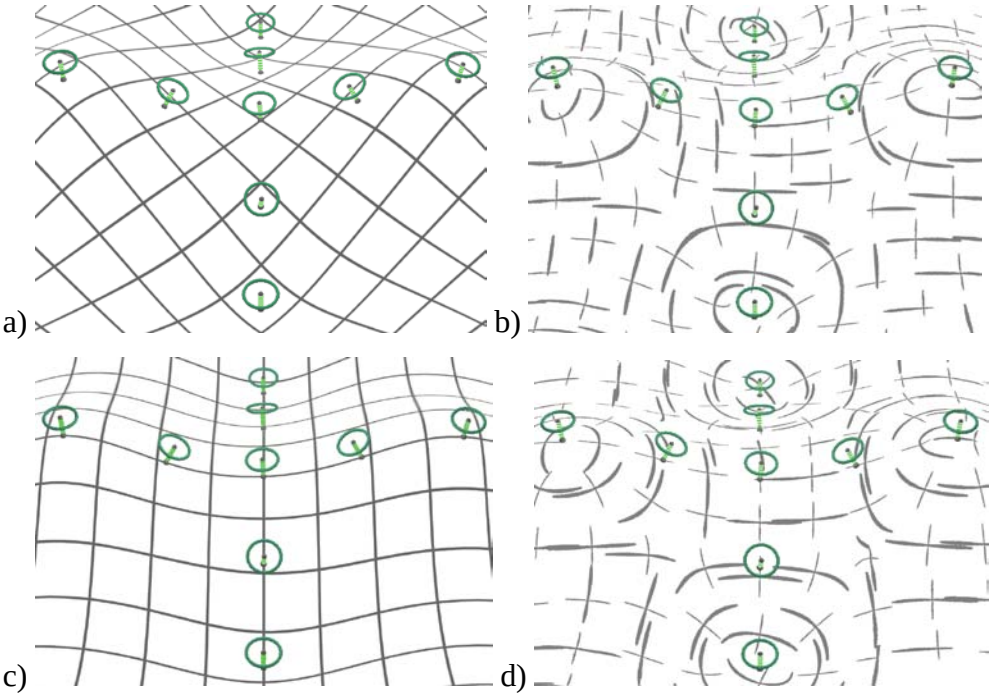



*Figure 118. Canonical Probes for Different Texture Rotations. The two texture rotation cases of 0° (a and b) and 45° (c and d) are shown for a 45° camera rotation and both principal direction and projected grids.*

We wanted to model how much of the error could be explained by the simple tendency to over-rotate the surface normal probe in proportion to the surface tilt toward the viewer. If we use a viewer-based coordinate system where y is up, x is left to right and z is toward the viewer, we can model the over-rotation of the probe as shortening the y-component and lengthening the x and z-components of the true surface normal vector to generate the probe vector. Figures 119-121 show linear regression models fitted to the surface normal and probe direction components in the x, y and z coordinate system just described. What we see in Figure 119 is that the x vector components are hardly biased from the ideal line. We do note that there are two visible modes in the graph around the 0-value for the probe x-component. The tendency is to see a curve fit through the data as shown in the plot labeled ‘Not a good fit’. However, this is a visual trick of the eye. The shape seen is due to a smaller variance in probe x-components when the actual normal

pointed straight toward the viewer than when it pointed to the left or right. The incorrect third-order fit shown gives significantly worse residual errors. Higher order regression showed that there was no significant cubic component, but a fourth-order regression had a significant quartic component. However, even the fourth-order fit did not significantly improve the residuals. Therefore, for this analysis, we use only linear regression.

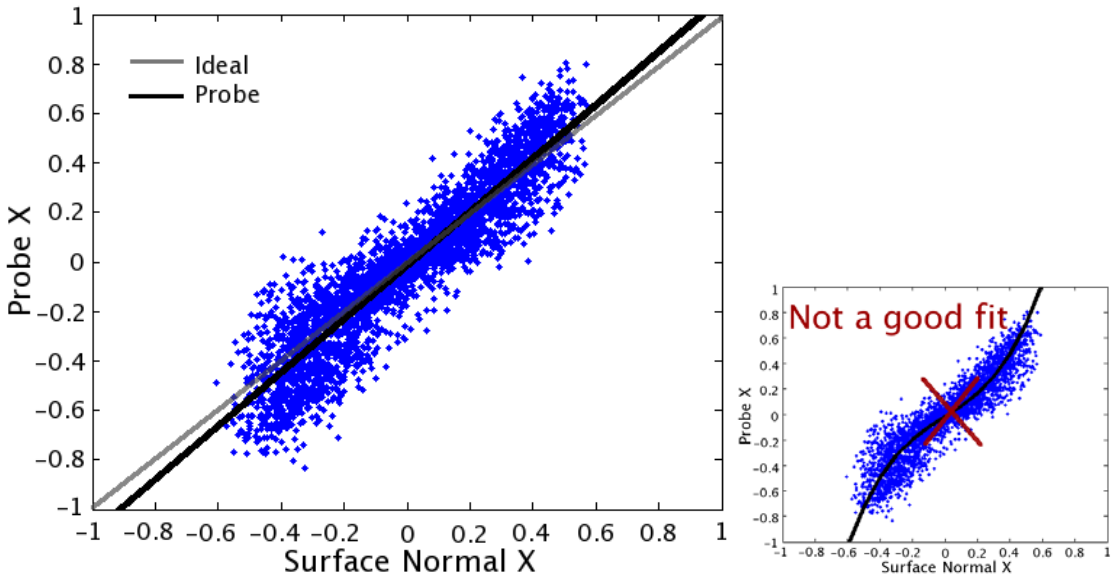



*Figure 119. Probe X-component Regression. On average, subjects had little bias in the x-component, as can be seen comparing the ideal grey line with the actual fit line in black.*

As expected, the probe y-component was strongly biased. Figure 120 shows the linear regression of y-probe components in black, and the ideal observer line in grey. Subjects strongly underestimated the true surface normal y-component. Vectors, whose surface normals have a y-component of 0.27, have an estimated probe direction straight toward the viewer. That is, probes with a tilt of about 30° away from the view direction would be seen as pointing straight at the subject.

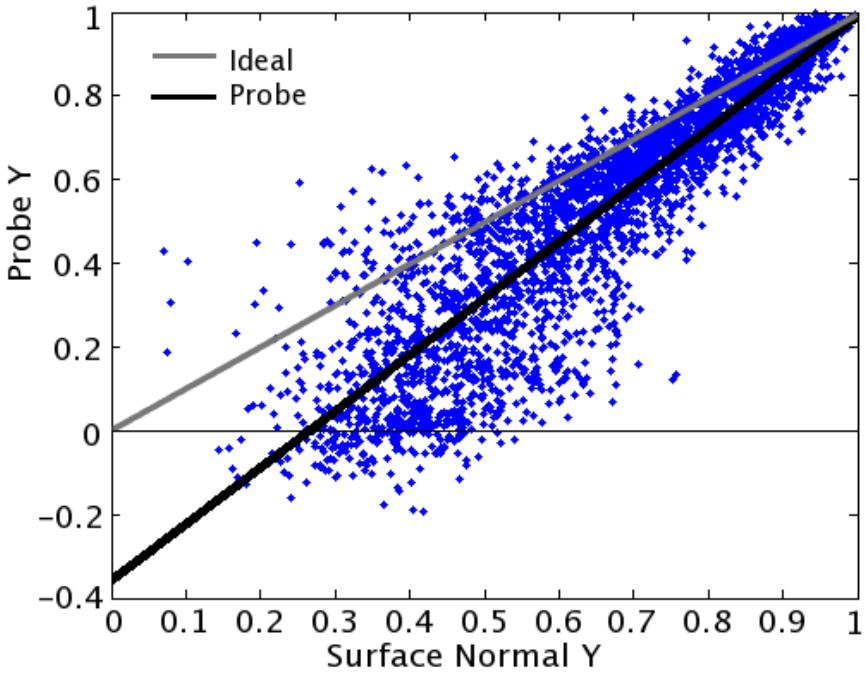



*Figure 120. Probe Y-component Regression. Subjects had significant bias in the y-component, as can be seen comparing the ideal grey line with the actual fit line in black.*

As shown in Figure 121, the z-component of the vector appears to be overestimated when the surface z is large. This is correlated with the y-component result; for the surfaces in this experiment, surface normals with small y-components must have relatively large z-components, and underestimation of the y-component will overestimate the z-component. Therefore, we focus only on the x-component and y-component regressions in the rest of the analysis.

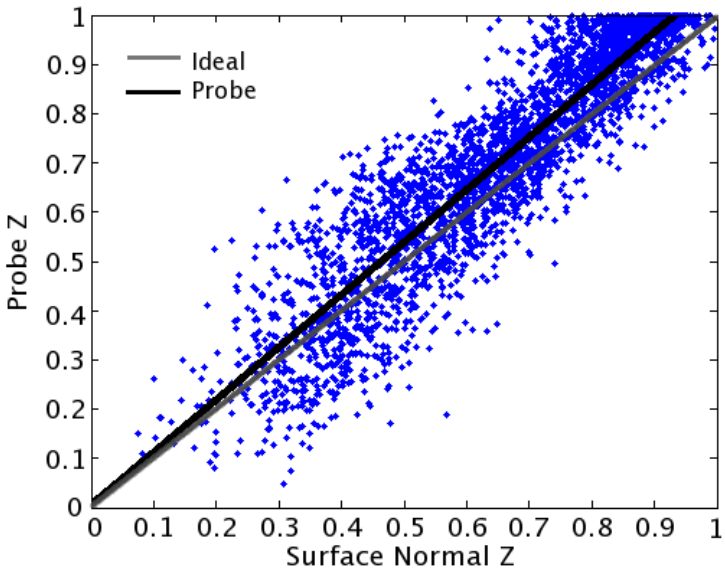



*Figure 121. Probe Z-component Regression. Subjects had some bias in the z-component, which is correlated to the bias in the y-component.*

Using a naïve approach and assuming that all subjects had the same slope of y-component bias, a regression was able to account for 23% of the overall probe error across all subjects and textures. However, looking at individual subjects' plots of the probe components vs. surface normal components showed that different subjects indeed had different biases. Figure 122 shows plots of the x-component and y-component data from all five subjects. Subject 1, the author, had relatively little y-component error, and essentially no x-component error. Although experienced subjects are not normally included for fear of biasing the experimental results, this is interesting information. It would seem that training and practice could perhaps lessen or remove one of the larger error components: over-rotation downward. The other subjects generally were very accurate for y-components near 1, and tended to underestimate the y-component when it was smaller. Interestingly, three of the subjects tended to overestimate the x-components slightly. This is equivalent to over-rotating around the y-axis, so negative x-components were seen as more negative, and positive as more positive. One subject was dramatically

different. Subject 4 had a tendency to underestimate the x-component, opposite of the other subjects. The y-component trend was different as well. Subject 4 tended to tilt the probe lower than correct when the surface was seen nearly edge-on and higher than correct when the surface normal was nearly facing the viewpoint.

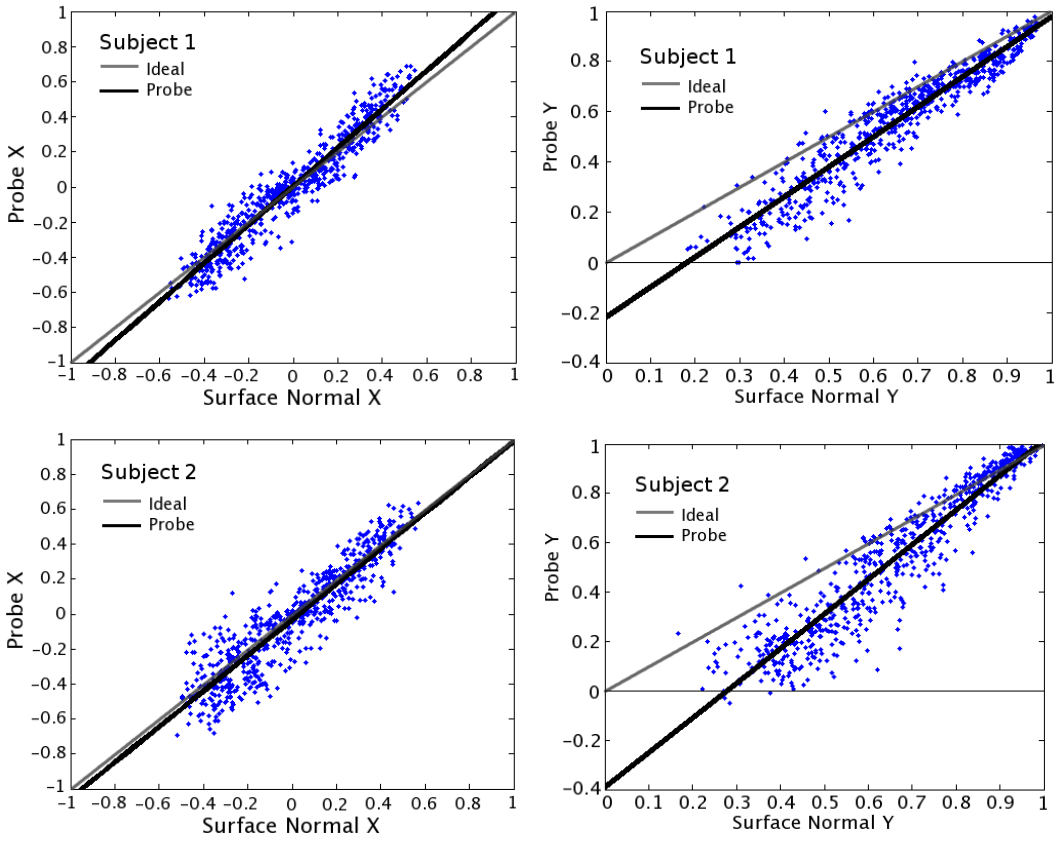



*Figure 122. Individual Subject Component Regressions. Results for linear regression on each subject's probe data show different trends depending on the other subjects.*

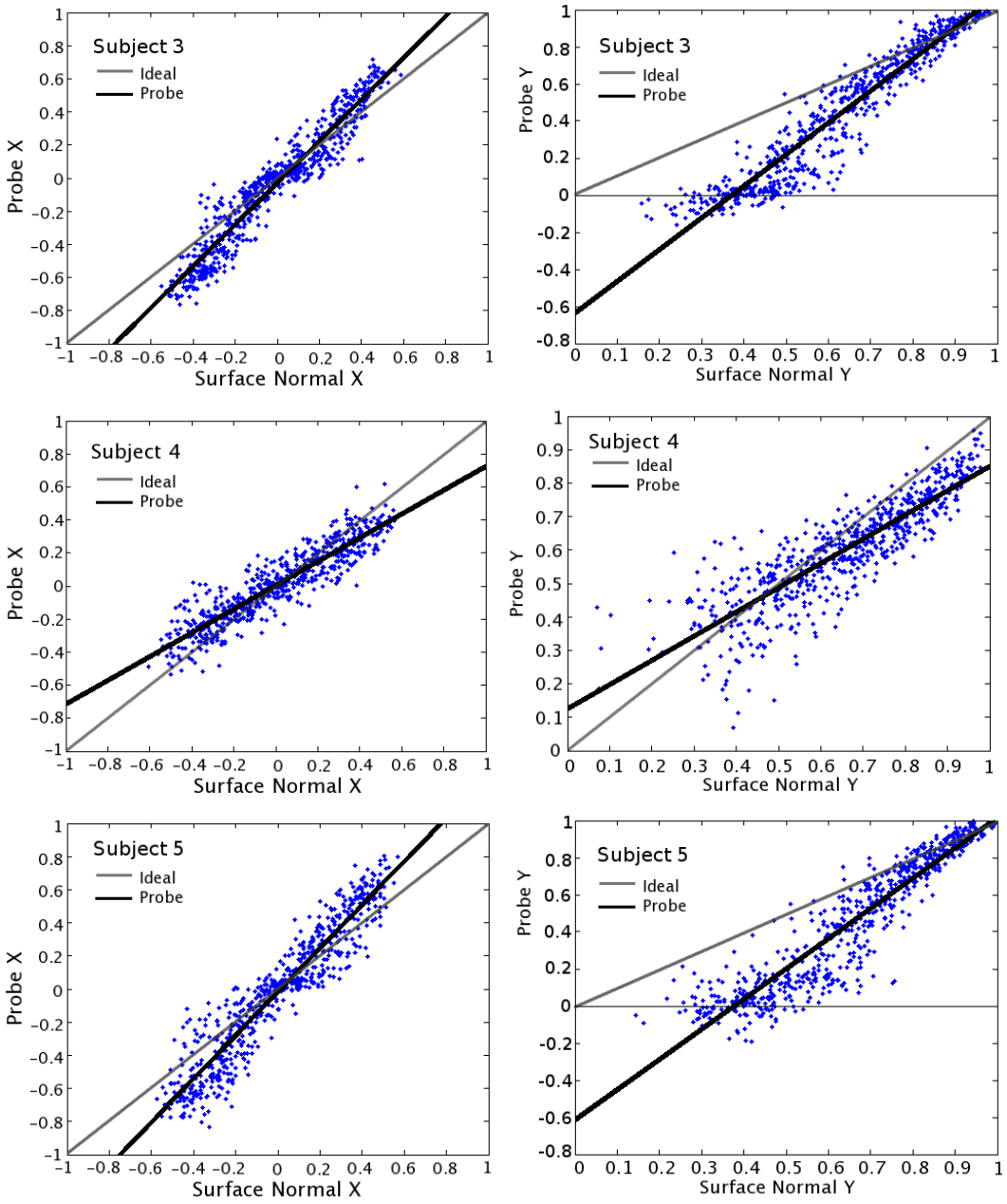



*Figure 122. Continued.*

We conclude that subjects seem to have individual biases in terms of over-rotation, though most tend to exaggerate rotations rather than de-emphasize them. However, it is worth noting that the probe errors of all subjects seemed to have a very

linear relationship between the true surface normal components and the estimated probe direction components. Data that was fitted to each subject's regression line for both y and x errors was able to explain 34% of the total angular error.

As a final check to look at how consistent the biases were for individual data points as well as averages, we show actual probe data in Figure 123. Each pair of cylinders represents a single data point for a given subject, surface, texture and viewpoint. The patterns are strikingly consistent, and were useful in forming hypotheses on how to analyze the data.

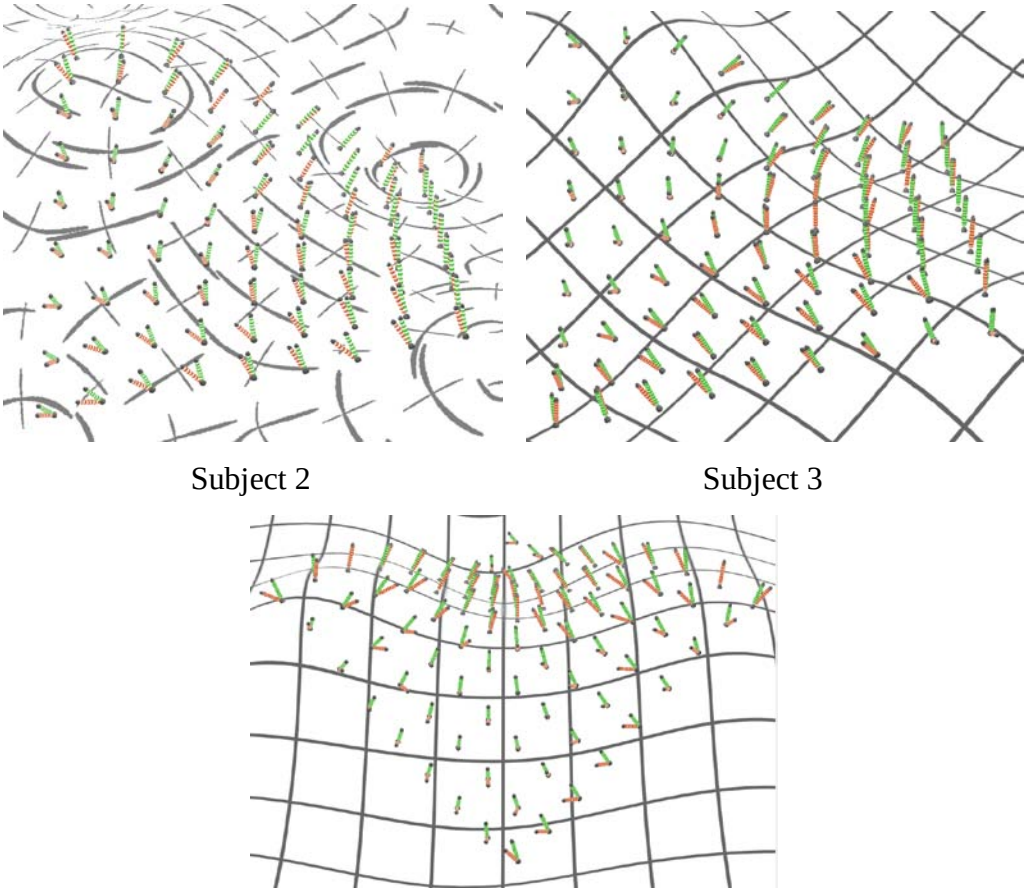



*Figure 123. Individual Probe Data Points. The green cylinders represent the correct surface normal, and the red are the probe positions. Each pair of cylinders represents a single probe measurement by a single subject.*

Another factor that was originally expected to have an effect on error was proximity to line junctions. Since junction information is so important, it was expected that probes placed close to a line junction would be more accurate than ones farther away. However, the data shows that this was not the case. A linear regression on angular error over maximum distance from a junction showed no relationship. Figure 124 shows the fit, which had a very slightly positive slope. It seems that subjects were able to correctly interpolate information from nearby line junctions to estimate the local surface direction.

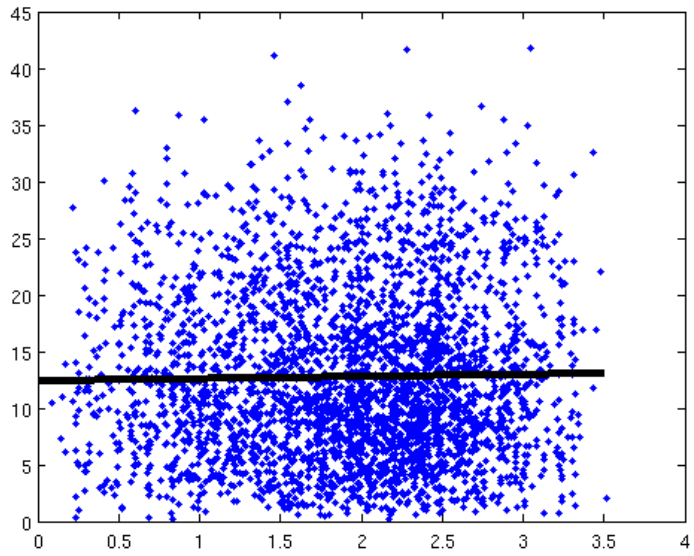



*Figure 124. Line Junction Proximity Regression. Proximity to line junctions is not a good indicator of error magnitude.*

## 8.8 Discussion

The largest effect found was actually due to the camera viewpoint. This is an interesting result because changing viewpoint is rather simple and can be done interactively. It suggests that for still images, choosing an optimal viewpoint is an

important consideration for surface visualization. In cases similar to this experiment, where silhouettes are not shown, the preferred viewpoint seems to be the direction that maximizes visible surface shape change in the directions of the image plane, rather than the depth dimension. Since silhouette contours are such a powerful shape cue, optimizing them would have to be considered as well in cases where they are shown.

Results show that subjects do seem to use line information as expected from the introductory analysis. That is, they assume that line junctions are perpendicular, but are able to use cues like parallel correspondence for additional information in some cases. Subjects were apparently little affected by the actual direction of the lines. A strong bias for most subjects to over-rotate probes down suggests a preference for the canonical 'forward' direction. The tendency for some subjects to over-rotate in the x-direction as well suggests that there might be a bias toward canonical horizontal angles as well. Since there tend to be per-subject biases, some techniques might be useful in balancing the biases. For example, showing a textured, front-facing surface patch alongside the projected surface might make subjects less likely perceive surfaces as facing them when it is not the case.

It seems that overall projected grid textures work just as well as principal direction textures. One caveat to this result is that projected textures naturally break down as the surface shape turns perpendicular to the projection direction. Since our surfaces were height fields, none of the local surface patches were close to being perpendicular to the projection direction, but higher errors were noted when this angle was large. Arbitrary surfaces have a full hemisphere of surface normals, and therefore include cases where projected textures might cause significant errors and should not be used. In this case, the more difficult to compute principal direction textures are the better solution.

One interesting question is how much the strong bias in over-rotating the probes was due to the height-field nature of the experiment. That is, surface normals are highly constrained in areas of the surface whose normals are nearly perpendicular to the view direction. Since the surfaces used in this case were height fields, the locations where the

surface normals were tightly constrained were correlated with locations that faced upward. For instance, if the viewer were looking up at a height field, we would expect the subjects to over-rotate the probes in the same way, by underestimating the magnitude of the surface normal's negative y-component.

However, if the subjects were seeing a more cylindrical object rather than a height field, they would see well-constrained surface normals facing both up and down, as well as fully front-facing surface normals. We suspect that this would cause a minimum of rotation bias for front-facing normals because subjects could identify all three extremes, and interpolate between them. However, we believe that there might still be a tendency to over-rotate probes that have a near-zero y-component, especially for textures like the 45° projected grids that emphasized surface tilt toward the viewer.

Lastly, we consider how probe direction might be biased with a fully three-dimensional object. In this case, subjects would have constrained surface normal information for every direction in the image plane, as long as silhouettes of the object were visible. This should give significantly more surface information, but we hesitate to assume that all biases would disappear. In this case, texture information would remain the primary means of interpolation between well-constrained silhouette surface normals and interior surface normals. Hopefully, future research can illuminate the complexities of this case.

There are several practical guidelines that can be built based on the results found here. The first, mentioned above, is to choose good viewpoints when showing surfaces. The second is that practice can improve surface perception accuracy and reduce biases, though how much training is required to be helpful is unknown. Thirdly, showing surface edges or silhouettes might significantly reduce the biases seen in this study. If neither can be shown for some reason, displaying patches of surface facing fully forward with the given texture could give subjects a basis for surface normal interpolation. Finally, some texture choices, such as a projected grid angled relative to the viewpoint, seem to exaggerate surface normal biases and should be used with care.

# 9. CONCLUSIONS

## 9.1 Summary and Conclusions

The purpose of this line of research is to investigate layered surface texturing. Visualization of layered surfaces can be useful both for scientific exploration of layered data and for instructional purposes like communicating well-understood layered data to other scientists and the general public. Showing layered data simultaneously should allow viewers to make connections between the data layers that would be more difficult to see if the data were presented separately. Texture was known to be a strong cue for surface shape, and has been shown in these studies to be an important cue for showing layered surfaces as well. The studies presented here used a broad to narrow approach to find what factors are important for layered textures, what values are optimal, and what interactions are important between the factors. Each study used experiments run on human subjects to find optimal texture conditions for layered surface visualization. Each experiment gave subjects a task to complete for a set of textures that varied according to a given texture parameterization. The initial experiments were broad, using many parameters, while the later experiments were more specific, varying only a few select texture parameters.

There were two main types of task used to measure perception of the layered surfaces. The task used in most of the experiments was a measure of surface direction perception; subjects had to align a probe to the surface normal at a particular location. However, one experiment used a feature-finding task; subjects had to identify bumps on a noisy surface. Optimal textures were found to vary depending on the task.

Texture was found to be extremely important for the surface normal estimation task. In a study that compared textured layered surfaces to un-textured but shaded layered surfaces, the errors on the un-textured surfaces were almost as bad as a random guess, far worse than any of the texture conditions studied. Overall, it was found that clear textures with strong directionality, such as grids or lines are the most important for showing shape information.

On the other hand, texture seemed less critical for the feature-finding task. In this case, optimal textures were subtle, with very high luminance. In most cases, the best textures had only one surface with a noticeable texture, leading us to suspect that texture was primarily being used to distinguish surfaces. Also, a bright texture allows for clear shading information: a strong cue for identifying features on a surface. Therefore, if a visualization is being used simply for identification of important features, textures should be subtle and bright.

Since surface normal estimation is a more difficult task than feature finding, and probably gives a better estimate of shape understanding, it was used for the finer-grained studies of texture optimization. Texture opacity was found to have an unexpectedly wide range of usable values; in a two-layer visualization, the top surface opacity could vary from 30% - 60% without any significant error increase. Texture sizes had a more complicated effect. The pilot studies suggested that larger top textures seemed to work better. Two studies explicitly varied texture size. Results from these showed that top texture size had a significant effect, but it involved a trade-off between performance on the top and bottom surfaces. Larger openings in top texture grids make the bottom surface easier to see, but the top surface harder to see.

A feasibility experiment was run to estimate how many surfaces could be layered before the visualization is too complicated for subjects to get any worthwhile information. Results from this study suggest a limit of 5-6 surface layers, at which point the visualization is too complicated to understand. Probably 3-4 layers are the most that could be practically shown at once. User interaction could possibly increase this limit. See the discussion for more ideas on user interaction for multiple layers.

Textures with lines in the principal curvature direction of a surface have been thought to be optimal for surface shape perception, but most of the research has focused on developable surfaces, a very limited set of all possible surface shapes. Therefore, our final experiment compared projected grid textures with principal direction textures. On average, projected grid and principal direction textures were found to work equally well for shape perception. However, the projected grid textures were worse in areas where the

surface normal pointed toward the viewer and significantly away from the projection direction. Overall, subjects used texture line information as expected for estimating the surface direction. Errors were lowest for edge-on surface patches, and highest for front-facing surface patches. A rotational bias was found on a per-subject basis. If the probe rotation can be thought of in terms of latitude and longitude, where the north pole corresponds to the up viewing vector, most subjects tended to over-rotate the surface normal probe toward the equator. The magnitude of this over-rotation was in proportion to how close the probe was to the equator. This suggests that steps could be taken to reduce the overall error by minimizing this bias.

### 9.2 Discussion

Only rather simple textures were used in this line of research. These included projected grids, dots and hatches, and principal direction textures made with discrete marks. Many other texture types are available, and might prove useful for shape perception. Among these are surface-dependent textures and viewpoint-dependent textures. Principal-direction textures are an example of surface-dependent textures, but others might include textures based on the surface height, curvature magnitude, or the direction of the surface relative to another point such as a light source. One example of a viewpoint-dependent texture is an edge enhancement algorithm. Surface boundaries are a strong shape cue. That is, surface shape on a boundary is more constrained than in other areas, and the shape seen at the boundaries can be interpolated to improve shape perception elsewhere. Silhouette shaders are already commonly used in non-photorealistic rendering. To make these shaders work for an interactive surface visualization, they would need to render fast enough for interactive exploration of the scene. Also, the algorithm would need to have temporal coherence. Since edge and boundary lines are viewpoint specific, the lines that are drawn would change as a user rotates a camera to explore a surface. A good algorithm would ensure that no unnatural temporal jumps in the lines would occur to distract the user. Since humans are very sensitive to motion, any abrupt changes in the overall line shape and location would

probably make the method unusable. But if the temporal coherence issues could be overcome, view direction-dependent textures could probably provide stronger surface shape cues. Ideally, an optimal texture would be one that finds optimal surface and viewpoint-dependent locations for showing shape information, adds clear and bias-free texture marks in those locations, and leaves the rest of the surface empty to allow space for information from other surfaces room to show.

The methods used to measure shape perception in this line of research were relatively simple. We included subjective ratings of surfaces, feature finding tasks and surface normal estimation. Surface normal estimations were able to measure not only magnitude, but also direction of error, which led to several interesting findings. However, one thing that has not been measured directly is how the subjects are actually using texture information. Eye-tracking is one method that could show what information subjects are actually focusing on when looking at the surfaces. For example, it would be interesting to learn how much surface area around any probe was actually looked at to perceive the surface. This might vary depending on the task. Certainly feature finding would probably involve a quick scan of the entire surface, with longer focus time spent on each feature. We hypothesize that surface normal estimation would only use the several closest line junctions to estimate a direction. Curvature estimation might require more data: synthesis of the global properties of the surface, or it might simply use the local curvature of a few line junctions.

Curvature has been measured before using a forced-choice categorization task, where subjects had to say if a surface was concave, convex, cylindrical or saddle-shaped. Subjects have also been asked to draw curves to match the perceived curvatures of singly curved surfaces. However, it would be interesting to find a way to get a quantitative measure of perceived local curvature for arbitrary surfaces. One possible way to do this might be asking the subject to align several probes in an area, and taking the difference in surface normals as a measure of the perceived curvature. However, since we don’t yet have a good understanding of how much local vs. global information is used in estimating surface normals, it would be hard to know if the probe directions

were being set relative to each other based on curvature perception, or independently based on surface normal perception at each location. Still, since curvature can be derived from change in direction, and vice versa, such an experiment should give useful results.

A second possibility would be to develop a ‘curvature probe’ similar to the surface normal probe to directly measure perceived curvature. For example, at a given surface location, the subject could manipulate a small surface patch to look like the actual surface. The subject could vary two curvature magnitudes and a rotation to try to match the surface patch with the real surface. This would measure the perceived minimum and maximum curvature directions, as well as their magnitudes. Since the surface patch direction would be set to correctly match the actual surface normal direction, biases in surface direction would not affect the curvature estimations. We note based on experience, that if stereo cues are used the surface patch should be translated away from the actual surface. This is because subjects can be very accurate at matching the relative heights of the surface patch with the real surface even without accurate perception of the surface curvature or direction. Thus care should be taken to make sure that the surface patch is seen from the same camera angle and field of view, despite being positioned away from the actual surface. Also, the surface patch should have visible edges to maximize the understanding of its curvature.

Surface normal and curvature perception both measure the perception of a single surface, no matter how many surface layers are being displayed at a particular time. However, the reason that multiple surfaces would be shown in a single scientific visualization is to find relationships between the surfaces. Relationships might include whether there are correlations between the directions or curvatures of the surfaces, or identifying locations where the surfaces are very close together or far apart. Finding relationships might work well as a feature-finding task, where subjects are given a pattern or relationship to look for in the surfaces. However, under the assumption that subjects are able to perceive the individual surfaces accurately, finding relationships becomes a problem of understanding the relative locations of points on each surface. Therefore, probes to measure perceived direction and distance between random points

on each surface might be a good quantitative measure of relationships between surfaces.

Monocular, static images of layered surfaces are much more difficult to see than visualizations with stereo and motion cues. Although we have done no studies directly comparing visualizations with and without the stereo and motion cues, subjects averaged less than 10° of error for a single surface in a study that used both cues, while subjects had an average error near 13° in a similar study that used only stereo cues. Since static images have fewer cues to surface shape, we would expect that static, monocular viewing conditions would have larger errors overall, and any biases would probably be exaggerated. Also, some parameters that were not particularly important for moving, stereo images might be more important without those cues. For instance, choosing an optimal viewpoint is probably quite important, because viewers do not get a variety of viewpoints. Also, differentiation of different texture layers using texture parameters like hue are also likely to be more important. Stereo and motion cues are very helpful for separating one surface layer from another. Therefore, cues that distinguish one surface from another should be useful for static images.

Not all three-dimensional data naturally fits a surface model. For example, in medical data, because of the nature of the scanning technology, most data is in the form of volume data. However, most of the tissues and organs of the body do have well-defined surfaces. Careful segmentation algorithms can successfully extract these surfaces from the volume data for texturing. There is no reason why volume rendering and surface rendering methods cannot be combined. Volume rendering can change depending on the transfer function. The transfer function is used to transfer the volume data along a ray direction into a color at a pixel. One of the most common transfer functions simply adds the densities at points along a ray, so that the image shading is an indication of the object thickness and material type. Treating the colors produced from volume rendering as surface shading means that they can be combined with texturing on the object's surface. Therefore, there is no reason why texturing cannot be used equally well with surface and volume data.

The subjects in our experiments had relatively little ability to interact with the

surfaces they were seeing. Although the surfaces rocked in most of the experiments, the subjects could not move around as they wished, and the camera was at a fixed distance from the surfaces. Many types of interaction might be added to improve overall perception of the surfaces. Certainly, allowing the users free movement within the scene would improve surface perception because the users could move to optimal camera angles depending on the shape and direction of the surface. Turning a difficult-to-see shape so that it shows as a silhouette would allow subjects to better interpret the 3D shape in that area.

Also, selective display of surfaces and parts of surfaces could help subjects understand single surfaces or pairs of surfaces at a time, so that a more complete mental picture of all the surfaces can be built. Based on the difficulty of seeing six surfaces together in the feasibility study, layering of a large number of surfaces is not likely to work without significant user interaction. A simple interface to turn the display of individual surfaces on and off would let users inspect any combination of layers at a time. This would require a user to remember the surface shapes and relationships between selected surfaces. However, we guess that the most important relationships are often between individual pairs of surfaces, so this simple method might work.

A slightly more complicated method could allow users to vary the opacity of portions of surfaces. That is, users could select areas of interest where they want to see lower level surfaces clearly. This portion of the upper level surfaces would become fully or mostly transparent so the user could see through them, but the rest of the upper surfaces would still provide context. This is similar to the artistic methods of using selective opacity only where the texture significantly helps show the upper surface, or where the lower surface is not important, and could be either user-driven, or possibly automatic.

Interpenetrating surfaces are another problem that was not considered in this line of research. Surfaces that have no clear ordering present several interesting problems for both the implementation and the theory of visualization. For one, if the surfaces are represented as polygonal meshes, a layered visualization must be drawn from the most

distant polygons at any pixel to the closest polygons for the composition operations to work correctly. This means that intersections of the surfaces must either be pre-computed, or the polygons must be sorted on a per-fragment basis. For surfaces based on distance fields or other functions, this is less of an issue.

However, the implications for how to texture interpenetrating surfaces are more interesting. Certainly, the textures should be consistent enough over each surface so that the continuity of each surface is clear. Yet, the bottom-most surface will likely be fully opaque, while any upper surfaces will be partially transparent. It would be interesting to test whether an abrupt or a smooth transition between the opacities works better. Based on the importance of silhouette and contour lines for perception, we anticipate that emphasizing the interpenetration contour might be useful for shape perception. Other style parameters might vary between the surfaces, such as color, size, frequency, line direction and texture type. Since several top-layer factors have been shown to affect bottom-surface perception, it would be important to consider which texture parameters were consistent along a particular surface, and which were consistent depending on the layer depth.

All of this research has been on showing surface shape, but other types of data might exist on the surfaces, and using texture to show that data could allow users to see relationships between the data and the surfaces themselves. Mapping data onto the textured surfaces is an interesting problem. Since surface shape is shown best with directional textures, any data mapping would have to work with the directional texture. Data is often shown with glyphs, using variables like shape, color, density and direction of glyphs to encode data visually. One workable method might be to have texture shape marks (probably lines of some sort), and glyph marks on the same surface. However, changes in glyph compression, size and density could be interpreted as changes in surface angle, or distance from the camera. Therefore care would have to be taken that the effects of viewpoint and surface direction did not interfere with interpretation of the data encoding. Hue would actually be a good glyph parameter. A single dimension of data on each surface could be encoded using different hues for each surface. Mark blur,

and line parallelism are other parameters that are independent of viewpoint and surface direction. Symbols would certainly also be useful for marking feature locations.

Color, blur and line parallelism are a rather restrictive set of glyph parameters. We believe that if the surface shape texture has strong regular distance cues, such as would be found in a regular projected grid, then glyph size, shape, density and direction could also be used to encode data. With the strong shape cues provided by grid contours, subjects are likely to correctly interpret shape changes as changes in the actual glyph shape rather than a rotation of the surface itself. Motion and stereo cues would help support this interpretation so that subjects could correctly read both the surface shape and changes in data glyphs.

Given how important texture has been found for surface shape perception, it is surprising that it is hardly ever used in scientific applications. Certainly care must be taken that textured do not obscure or confuse the surface data being shown, and scientists are not always known for their artistic abilities. However, we feel that with the guidelines presented in these studies, even people unfamiliar with visualization principals could successfully use texture to show surface shape. One use for texture is medical visualization. In many cases, disease diagnosis is dependent on feature finding of abnormalities, such as tumors. For these tasks, any textures used should be very subtle so that the texture does not interfere with any feature data. However, in some cases the shape is very important, such as determining the size and shape of a tumor to be surgically removed. In this instance, we would strongly recommend the use of texture to improve shape perception. Like any tool, texturing must be used with care, but when used to best advantage, texturing can provide huge benefits to the complicated process of shape perception.

# REFERENCES

Acevedo, D. and Laidlaw, D. H. 2006. Subjective quantification of perceptual interactions among some 2d scientific visualization methods. *IEEE Transactions on Visualization and Computer Graphics* (Proceedings Visualization / Information Visualization) *12*, 5, 1133-1140.

Albers, J. 1975. *Interaction of Color*. Yale University Press, New Haven, CT.

Bair, A., House, D., and Ware, C. 2005. Perceptually optimizing textures for layered surfaces. In *Proceedings of Symposium on Applied Perception in Graphics and Visualization*. 67-74.

Bair, A., House, D., and Ware, C. 2006. Texturing of layered surfaces for optimal viewing. *IEEE Transactions on Visualization and Computer Graphics* (Proceedings of Visualization 2006), *12*, 5. 1125-1132.

Bair, A. and House, D. 2007. A grid with a view: optimal texturing for perception of layered surface shape. *IEEE Transactions on Visualization and Computer Graphics*, *13*, 6, 1656-1663.

Barrow, H. and Tenenbaum, J. 1981. Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, *17*, 75-116.

Campbell, F. and Green, D. 1965. Optical and retinal factors affecting visual resolution. *J Physiol*, *181*, 3, 576-593.

Chapman, H. 2006. *Michelangelo*, Yale University Press, New Haven, CT.

Costa Sousa, M. and Prusinkiewicz, P. 2003. A few good lines: Suggestive drawing of 3d models. *Comput. Graph. Forum 22*, 3, 381-390.

Craven, M., Shavlik, J. 1997. Using neural networks for data mining. *Future Generation Computer Systems*, *13*, 211-229.

Cummin, B., Johnston, E., and Parker, A. 1998. Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Res. 33*, 5-6, 827-38.

Curran, W. and Johnston, A. 1996. The effect of illuminant position on perceived curvature, *Vision Res. 36*, 10, 1399-1410.

De Vries, S., Kappers, A., and Koenderink, J. 1993. Shape from stereo: A systematic approach using quadratic surfaces. *Perception & Psychophysics. 53*, 1, 71-80.

dos Ries Rivotti, V. C., La Pais Proenca, J. R., Jorge, J. A., and M. Costa Sousa. 2007. Composition principles for quality depiction and aesthetics, *Proceedings of Computational Aesthetics 2007*, 37-44.

Feichtinger, H. and Strohmer, T. 1998. *Gabor Analysis and Algorithms: Theory and Applications*. Birkhäuser, New York, NY.

Gabor Filter. 2009. en.wikipedia.org/wiki/Gabor_filter.

Gooch, A., Gooch, B., Shirly, P., and Cohen, E. 1998. A non-photorealistic lighting model for automatic technical illustration. In *Proceedings of International Conference on Computer Graphics and Interactive Techniques*, 447-452.

Gorla, G., Interrante, V., and Sapiro, G. 2003. Texture synthesis for 3d shape representation. *IEEE Transactions on Visualization and Computer Graphics, 9*, 4, 512-524.

Grey, L. 2007. Stereo photographs of Texas A&M fountain. Unpublished, Texas A&M University.

Hagh-Shenas, H., Interrante, V., Healey, C., and Kim, S. 2006. Weaving versus blending: A quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, 164-164.

Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, NJ.

Hoffman, D. 1983. The interpretation of visual illusion. *Scientific American 249*, 6, 154-162.

House, D. and Ware, C. 2002. A method for the perceptual optimization of complex visualizations. In *Proceedings of Advanced Visual Interfaces (AVI' 02)*, 148-155.

House, D., Bair, A., and Ware, C. 2005. On the optimization of visualizations of complex phenomena. In *Proceedings of IEEE Visualization 2005*, 87-94.

House, D., Bair, A., and Ware, C. 2006. An approach to the perceptual optimization of complex visualizations. *IEEE Transactions on Visualization and Computer Graphics, 12*, 4, 509-521.

Interrante, V., Fuchs, H., and Pizer S.M. 1997. Conveying the 3d shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on Visualization and Computer Graphics, 3*, 2, 98-117.

Interrante, V. and Grosch, C. 1998. Visualizing 3D Flow. *IEEE Computer Graphics and Applications, 18,* 4, 49-53.

Interrante, V., and Kim, S. 2001. Investigating the effect of texture orientation on the perception of 3d shape. *Human Vision and Electronic Imaging VI,* SPIE 4299, 330-339.

Interrante, V., Kim, S., and Hagh-Shenas, H. 2002. Conveying 3d shape with texture: Recent advances and experimental findings. *Human Vision and Electronic Imaging VII*, SPIE 4662, 197-206.

Johnson, J. 1982. *Treasury of American Pen-And-Ink Illustration: 1881 to 1938.* Dover Publications, Inc, New York.

Kim, S., Hagh-Shenas, H., and Interrante, V. 2003. Showing shape with texture: Two directions seem better than one. *Human Vision and Electronic Imaging VIII*, SPIE 5007, 332-339.

Kim, S., Hagh-Shenas, H., and Interrante, V. 2004. Conveying shape with texture: Experimental investigations of the texture's effects on shape categorization judgments. *IEEE Transactions on Visualization and Computer Graphics*, *10*, 4, 471-483.

Kindlmann, G. and Westin, C. 2006. Diffusion tensor visualization with glyph packing, *IEEE Transactions on Visualization and Computer Graphics 12*, 5, 1329-1335.

Kirby, R, Keefe, D., Laidlaw, D. 2004. Painting and visualization, In *Visualization Handbook,* Hansen and Johnson, Eds. Elsevier Academic Press, 873-891.

Knill, D. 1998. Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research,* 38, 1655-1682.

Knill, D. 2001. Contour into texture: Information content of surface contours and texture flow. *Journal of the Optical Society of America. A, 18,* 1, 12-35.

Koenderink, J. and van Dorn, A. 1991. Affine structure from motion. *Journal of the Optical Society of America A, 8,* 2, 377-385.

Koenderink, J., and van Dorn, A., Kappers, A. 1992. Surface perception in pictures, *Perception and Psychophysics, 52*, 5, 487-496.

Laidlaw, D., Kirby, M., Davidson, S., Miller, T., DaSilva, M, Warren, W., and Tarr, M. 2001. Quantitative comparative evaluation of 2d vector field visualization methods. In *Proceedings of IEEE Visualization 2001*, 143-150.

Li, A., Zaidi, Q. 2000. Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research 40*, 217-242.

Li, A., Zaidi, Q. 2001. Veridicality of three-dimensional shape perception predicted from amplitude spectra of natural textures. *Journal of the Optical Society of America. A. 18*, 10, 2430-2447.

Lohan, F. J. 1983. *Pen & Ink Sketching: Step by Step.* Contemporary Books, Inc., Chicago.

Mamassian, P., and Kersten, D. 1996. Illumination, shading and perception of local orientation. *Vision Research, 36,* 15, 2351-2367.

Mamassian, P., Knill, D., and Kersten, D. 1998. The perception of cast shadows. *Trends in Cognitive Sciences*, *2*, 8, 288-295.

Mandelbrot, B. 1982. *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco.

Neber, A. 2007. Sediment layer exploded view. http://www.eurotecbroker.com/upload/bilder/Erganzungen%20ETB%20Borse/produkt_071.jpg, Germany.

Nefs, H., Koenderink, J., and Kappers, A. 2006. Shape-from-shading for matte and glossy objects, *Acta Physologica*, *121*, 297-316.

Ramachandran, V. 1988. Perceiving shape from shading. *Scientific American*, *259*, 2, 76-83.

Saunders, J. and Knill, D. 2001. Perception of 3D surface orientation from skew symmetry. *Vision Research*, *41*, 3163-3183.

Saunders, J. and Backus, B. 2006. Perception of surface slant from oriented textures. *Journal of Vision*, *6*, 882-897.

Singh, M. 2007. Image of a ’32 Dodge. Unpublished, Texas A&M University.

Squillacote, A. 2006. *The Paraview Guide.* Kitware Inc, Clifton Park, NY.

Stevens, K. 1981. The visual interpretation of surface contours, *Artificial Intelligence*, *17*, 47-73.

Stone, M. 2003. *A Field Guide to Digital Color.* AK Peters, Ltd, Natick, MA.

Sweet, G., Ware, C. 2004. View direction, surface orientation and texture orientation for perception of surface shape. In *Proceedings of the 2004 Conference on Graphics Interface.* ACM International Conference Proceeding Series, *62*, 97-106.

Taylor, R. 2002 Visualizing multiple fields on the same surface, *IEEE Computer Graphics and Applications*, *22*, 3, 6-10.

Todd, J. and Mingolla, E. 1983. The perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 583-595.

Todd, J. and Reichel, F. 1990. Visual perception of smoothly curved surfaces from double-projected contour patterns. *Journal of Experimental Psychology: Human Perception and Performance. 16*, 3, 665-674.

Todd, J., Farley Norman, J., and Koenderink, J. 1997. Effects of texture, illumination and surface reflectance on stereoscopic shape perception, *Perception*, *26*, 807-822.

Todd, J. and Oomes, A. 2002. Generic and non-generic conditions for the perception of surface shape from texture. *Vision Research 42*, 837-850.

Todd, J., Oomes, A., Koenderink, J. and Kappers, A. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psychological Science, 15*, 1, 40-46.

Urness, T., Interrante, V., Longmire, E., Marusic, I., O'Neill, S., and Jones, T. 2005. Strategies for the visualization of multiple co-located vector fields. *University of Minnesota Tech. Report 05-032*.

Viola, I., Kanitsar, A., and Gröller, M. E. 2004. Importance-driven volume rendering, In *Proceedings of IEEE Visualization 2004*, 139-145.

Ware, C. and Frank, G. 1996. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics 15*, 2, 121-140.

Wheatstone, C. 1838. Contributions to the physiology of vision. Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision, *Philosophical Transactions of the Royal Society*, *128*, 371-394.

Zhang, R., Tszi, P., Cryer, J., and Shah, M. 1999. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*, 8, 690-706.

Zhang, L., Curless, B., Hertzmann, A., and Seitz, S. 2003. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV), 2*, 618.

# VITA


Name: Alethea Scattergood Bair

Address: Department of Visualization, Texas A&M University, C108 Langford 3137 TAMU, College Station, TX 77843-3137

Email Address: abair@viz.tamu.edu


Education: B.A., Physics, University of Illinois at Urbana/Champaign, 2003
Ph.D., Architecture, Texas A&M University, 2009

Publications:

Bair, A. and House, D. 2007. A grid with a view: Optimal texturing for perception of layered surface shape, *IEEE Transactions on Visualization and Computer Graphics, 13*, 6, 656-1663.

House, D., Bair, A. and Ware, C. 2006. An approach to the perceptual optimization of complex visualizations, *IEEE Transactions on Visualization and Computer Graphics, 12*, 4, 509-521.

Bair, A., House, D., and Ware, C. 2006. Texturing of layered surfaces for optimal viewing. *IEEE Transactions on Visualization and Computer Graphics* (Proceedings of Visualization 2006), *12*, 5. 1125-1132.

House, D., Bair, A., and Ware, C. 2005. On the optimization of visualizations of complex phenomena. In *Proceedings of IEEE Visualization 2005*, 87-94.

Bair, A., House, D., and Ware, C. 2005. Perceptually optimizing textures for layered surfaces. In *Proceedings of Symposium on Applied Perception in Graphics and Visualization*. 67-74.

Awards: First place in the 2007 IVC Data Visualization Competition
Finalist in the 2007 Microsoft DreamBuildPlay Competition