Yale University
Department of Computer Science

A Psychophysical Study of Dominant Texture Detection

Jianye Lu   Alexandra Garr-Schultz
Julie Dorsey   Holly Rushmeier

YALEU/DCS/TR-1417
June 2009

Contents

List of Figures

List of Tables

A Psychophysical Study of Dominant Texture Detection*

Jianye Lu, Alexandra Garr-Schultz, Julie Dorsey, and Holly Rushmeier

Abstract

Images of everyday scenes are frequently used as input for texturing 3D models in computer graphics. Such images include both the texture desired and other extraneous information. In our previous work, we defined dominant texture as a large homogeneous region in an input sample image and proposed an automatic method to detect dominant textures based on diffusion distance manifolds. In this work, we explore the identification of cases where diffusion distance manifolds fail, and consider the best alternative method for such cases. We conducted a psychophysical experiment to quantitatively study the human perception of dominant texture, by asking subjects to compare the analysis and synthesis results of dominant texture detection from different input images generated using different techniques. We then applied Analysis of Variance (ANOVA) to determine significant preference of one technique over another; paired comparison scaling (PC-Scaling) to quantitatively evaluate technique performance; multidimensional scaling (MDS) to further classify texture samples based on their paired comparison scales; and correlation estimation to reveal certain conventional texture descriptors that are suitable to predict dominant texture detection technique performance for texture samples. Our experiment confirmed that diffusion distance manifolds produce the best results for texture selection for a large number of image classes. Based on our experiment, we propose a technique for identifying cases where diffusion distance manifolds may fail, and suggesting the best alternative method for these cases.

*This material is based upon work supported by the National Science Foundation under Grant No. 0528204.

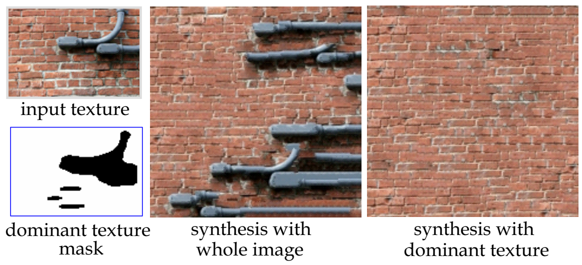


Figure 1: Dominant texture and homogeneous texture synthesis.

1 Introduction

Texture mapping is a standard technique for adding detailed variation of color and tone to three-dimensional digital models. Texture maps may be defined procedurally, or by image exemplars [18]. Images of natural scenes may be used as sources for exemplars. Finding an image that consists of a homogeneous texture alone – e.g. an area of grass with no anomalous weeds or flowers, or a brick wall with no singular cracks or holes – can be difficult. Natural images need to be edited to extract the desired texture and remove irrelevant regions of the image. In our previous work [22], we defined *dominant texture* as a large homogeneous region from an input texture sample; identifying such regions can improve the quality of texture synthesis results (see Figure 1.)

There are a variety of approaches to detect dominant textures. In Section 7 in [22], we provided digitally rendered results based on four different techniques, and required viewer's subjective judgement on quality evaluation. In this work, we explore the evaluation and classification of methods to automatically extract the region of dominant texture from an image. Based on the results of a psychophysical experiment, we propose a method for identifying the best method to perform such texture extraction. This work represents the last step in the process of our dominant texture detection system of appearance analysis, synthesis and validation.

In the rest of this report, we first review dominant texture and previous psychophysical studies in graphics in Section 2. Second, we introduce our experimental design in Section 3. Next, we detail data analysis with Analysis of Variation (ANOVA), multidimensional scaling (MDS), and scaling methods in Section 4. Then we study correlation between technique performance and conventional texture descriptors in Section 5. Finally, we conclude with several interesting points and possible improvement in Section 6.

2 Psychophysical Studies and Computer Graphics

One goal of computer graphics is to provide photo-realistic visual experience, where our eyes are the ultimate judge of photo-realism. However, currently we do not have complete knowledge of the human visual system. Therefore, we have to rely on the response from human visual systems using psychophysics studies. In this section, we briefly review several widely applied psychophysical experiments and their applications in computer graphics.

2.1 Diffusion Distance Manifolds

Recently we developed a new method for dominant texture detection [22]. The input is an image selected by the user that has a large area of the desired texture. The method has two phases. First, using Fourier analysis, the scale of the most common image feature is detected. Second, clusters of image patches of the size found in the first phase are formed. Each patch of n pixels is considered an n-dimensional vector. The image patches form a lower-dimensional manifold in n-dimensional space. The key insight into finding clusters of patches is to measure the distance between patches using the diffusion distance between the vectors in patch space. The diffusion distance accounts not only for the distance between patches within the manifold, but the number of paths between the patches. This property of the diffusion distance makes the cluster classification robust (e.g. see [5]). The patches in the largest cluster of patches form the dominant texture.

In [22], we presented comparisons among the diffusion distance manifold approach to the normalized-cut texture segmentation [29], and clustering with Euclidean distances and with using no dominant texture detection on input images. For images with obvious "contamination" of the dominant texture, the diffusion distance manifold approach appears to give the best results. However, for some classes of images, the dominant texture is not well detected. In this work, we explored in more depth identifying the cases where the diffusion distance manifold method fails, and considering the best alternative method.

2.2 Psychophysical Experiments

Ferwerda et al. presented fundamental rules about how to design, conduct, and analyze perceptual experiments in computer graphics [8, 9]. They also detailed several methods for data analysis.

In order to select the right psychophysical experiment and data analysis method, we first need to consider: *What to learn from our subjects?* There are two fundamental psychophysical quantities: absolute/difference thresholds (that measures the perception limits) and scales (that measures everything else). If we want to measure perception limits, we can use Fechner's classical threshold methods; if we want to reveal some quantitative properties, we can use scaling methods; or, if we want to study the structure of our data, we can use *Multidimensional Scaling* (MDS).

In most of the cases, we are interested in some quantitative measure (or, scales) of our samples. If participants can provide direct numerical values to their perception, we can use direct scaling methods (such as *Equisection*, *Magnitude Estimation*, and *Magnitude Production*); otherwise, we use indirect scaling methods (such as *Rating Scaling*, *Paired Comparison Scaling*, *Ranking Scaling*, and *Category Scaling*) where we only require participants to make simple judgments (such as sorting or grouping) and derive scale values using statistics.

We create a decision tree (as showed in Figure 2) that helps us narrow down the selection of scaling methods based on our answers to questions so far. We also need to consider some constraints when we design our experiments:

- *Rating Scaling* is known to subject to problems of range effects, frequency effects, and distribution effects;
- *Paired Comparison Scaling* requires significantly more trials when then number of sample set is large;
- *Ranking Scaling* requires relatively complicated user interaction;
- *Category Scaling* requires reliable predefined description of user perceptions (such as bright/dark);
- All direct scaling methods require continuous stimuli under samples;
- *Equisection* and *Magnitude Production* also require interactive-time sample generation given any stimuli, either online or offline.

In this study, we build on earlier work in computer graphics to evaluate alternative computational techniques and organize the results to form the basis of a new method. Following the work, such as that described in [15], we use a two-alternative forced choice (2AFC) method to compare computational techniques. Using the method given in [6] we convert these judgements into interval scales. Like [25] we use multidimensional scaling (MDS)

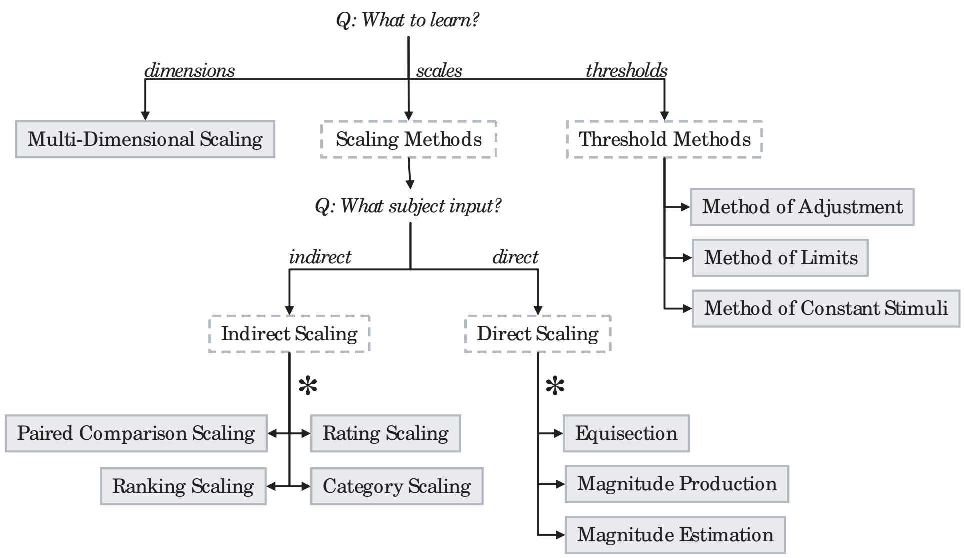


Figure 2: Decision tree that narrows down psychophysical experiment selection. $*$: See Section 2.2 for further comparison between these methods.

to form a 2D map to organize the results of the experiment. And similar to [28], which develops a new computational method from psychophysical results, we use the 2D organization from MDS as the basis of a classification method for determining the best technique for detecting dominant texture in a particular image.

2.3 Psychophysics Applications in Graphics

There are two challenges in realistic modeling and rendering.

1. It is difficult to build up a model that looks exactly the same as it physically should be, either because of our lack of understanding of the natural processes, or because of the limited computational resources.

2. It is difficult to quantitatively evaluate the performance of existing models because human perception is too subjective to quantify. For example, we demonstrated the quality of existing methods in weathered appearance generation is difficult to assess because of the rather limited validation data (see Table 1 in [23].) Many methods have not been validated at all. Where validation *has* been performed, it has generally not been applied to reproducing the full appearance caused

by aging effects, but has been focused on particular aspects of the phenomenon being studied, such as the time of onset of cracks, or the general shape of patterns being formed.

In either case, psychophysical studies help develop more effective model (by selecting the right features), more efficient implementation (by computing only what is necessary), and more effective application and user interfaces (by measuring user experience).

Psychophysical methods from experimental psychology have been widely applied in graphics, such as in research work about reflectance models [10, 24, 25], texture models [13, 14, 16], color-to-grayscale conversion [3], tone mapping [2,4,11,17], visual complexity [26,27], visual equivalence [28], user interface [19], visual realism [7], and so on (see Appendix A for detailed description of these work.) Closely related to our work, Longhurst et al. conducted psychophysical experiments to validate that adding visual imperfections (like dust, dirt and scratches) indeed made images perceived more realistic [20]. All those research topics either lack clear description or lack quantitative models.

In our work of dominant texture, we had a good description but no quantitative definition of what dominant texture is, especially when we are talking about a broad range of natural appearance samples. In addition, it is difficult to quantitatively evaluate the performance of existing models because human perception is too subjective to assign numbers. Therefore, we rely on a psychophysical experiment to quantitatively validate our selection of diffusion distance manifolds for dominant texture detection.

3 Experimental Method

To compare different dominant texture detection techniques, we apply the *Two-Alternative Forced Choice* (2AFC) method to collect a subject's response, where participants are required to select one over the other when images from the same source sample but with two different processing techniques are presented along with the source sample as reference. We detail a few experiment elements as follows:

Stimuli

We selected 17 texture sample images that capture natural variances, such as material weathering effects, or combination of different objects in a natural scene (see Figure 3.) These images include texture samples we already used

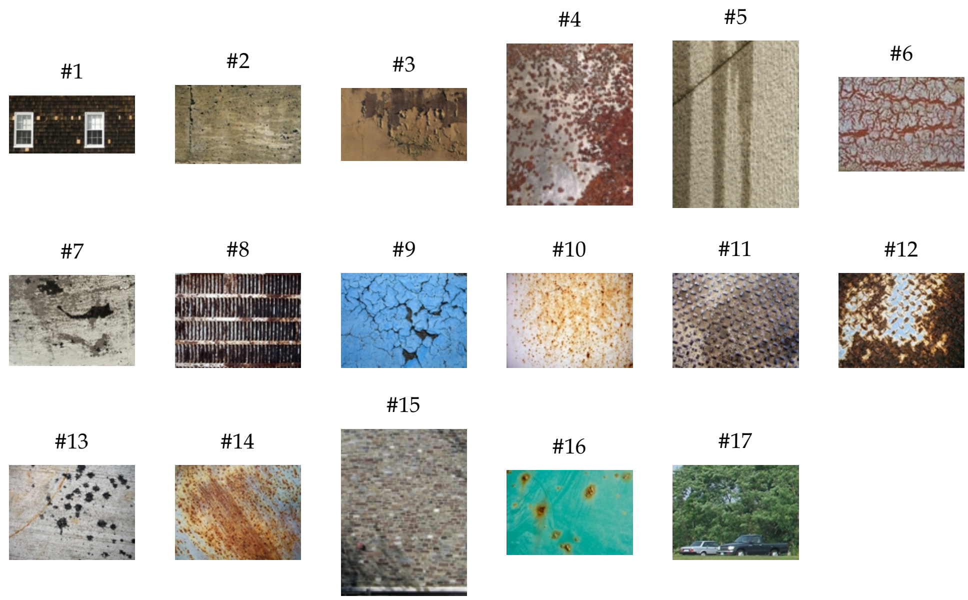


Figure 3: Texture samples we used in our psychophysical experiment.

in [22]. Given an input texture sample, we follow the processing pipeline in Section 5 in [22] and apply four different techniques to estimate dominant texture masks and generate larger uniform texture patches:

1. *nCut*: *Normalized-Cut*-based method as discussed in Section 7.4 in [22] and in [29];
2. *Diff*: Diffusion-distance-based method as detailed in Section 4 in [22];
3. *Eucl*: Euclidean-distance-based method as discussed in Section 7.3 in [22];
4. *None*: No preprocessing and use the whole input image for synthesis.

Graphical User Interface

Participants are shown a series of image sets on the screen. Each set contains three images: a source texture image at the center as reference, and two texture images generated based on different techniques on the sides. Participants are required to decide which of the two side images (left or right) he thinks better capture the underlying homogeneous texture of the reference

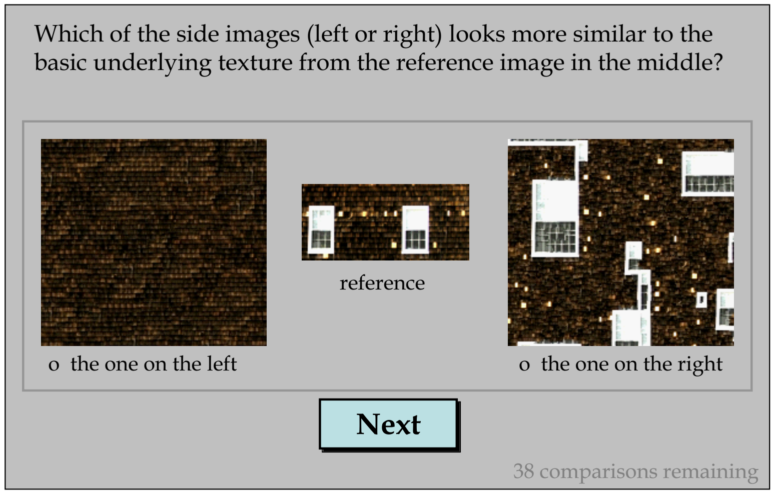


Figure 4: Graphical user interface for our psychophysical experiment.

image at the center, while preserving realistic appearance of the reference image, by simply clicking the radio button below that image. Then they click the "Next" button to proceed to the next set of images. Participants need to respond before going on to the next set of images, and cannot quit without responding to all sets of images. There is a prompting at the lower right corner of the dialog that tells participants how many comparisons are left to respond. See Figure 4.

Participants

We have 21 participants in our study who are all affiliated with Yale University. All participants are 18 years old or above, but no other demographic information has been collected due to privacy consideration.

Software and Hardware

The experiment was coded and conducted in Matlab R7.6 on a machine running Microsoft Windows XP (SP2). We use a 19-in Dell 1901FP flat panel color monitor and a USB mouse to simplify user interaction.

Procedure

Upon participation, subjects were given an informed consent form and a standardized written instruction sheet. During the experiment, image samples were arranged and presented in a way that any pair of images from the same source texture but with different processing techniques were displayed exactly once, and the orders of samples and processing technique pairs were completely randomized. After participants finished all image comparisons, they were given contact information in case of further questions.

Pilot Test

As a means of determining the actual experiment, we conducted a pilot test with our preliminary instruction sheets and test image samples within a small group of graduate students in the computer graphics lab. We addressed the following user comments to refine our experimental design:

1. Many participants considered our initial instruction confusing; thus, their response turned out to be away from our interests. We carefully redesigned the instruction sheet and asked participants to select the image that "*looks more similar to the basic underlying texture from the reference image in the middle.*" We also provided two visual examples on the instruction sheet to help further clarify our purpose;

2. Many participants complained that testing time was too long. We selected only 17 representative texture samples into our actual experiment, and reduced the testing time to around 10 minutes, which became acceptable for most participants;

3. Many participants complained that they felt lost not knowing how many comparisons were still left. We added a text indicator at the low right corner of the dialog, which eased the user anxiety;

4. Some participants suggested adding two more options for image comparison: "Both Images Look Good" and "Neither Image Looks Good Enough." However, we decided not to provide such extra options and force participants to select one image over another.

Data Collection

We maintain a 4-D binary matrix $R_{17\times 21\times 4\times 4}$ to collect subject responses: when subject p prefers the image with processing technique j over that with technique i with respect to texture sample t, we assigned 1 to $R(s,p,i,j)$ and 0 to $R(s,p,j,i)$, then 0 to all $R(s,p,i,i)$ for any s and p. This matrix is the starting point for our data analysis in the following sections.

4 Subject Response Analysis

4.1 Analysis with ANOVA

In statistics, analysis of variance (ANOVA) is a collection of methods to compare variables based on their variances. It has wide applications in data analysis in psychophysical experiments. In this section, we assume the responses from different subjects to a particular texture sample, with a particular technique are distributed as a normal function, and we study the statistical significance between techniques by a series of two-sample t-tests with a threshold of $\alpha = 0.05$, implemented by function `ttest2` in Matlab.

- **Across all participants and all texture samples**: We summed up responses for a particular technique j to a scalar $R_1^j = \sum_{s,p,i} R(s,p,i,j)$ and run the test between each of them. We found that textures processed by *Diff* were selected significantly more than those by *None* ($p < 0.001$), *nCut* ($p < 0.001$), and *Eucl* ($p < 0.001$);

- **Across participants but separated by texture samples**: We summed up responses for a particular texture sample s to a vector $R_2^s(j) = \sum_{p,i} R(s,p,i,j)$ and run the test between each of them. Figure 5(a) summarizes technique preferences, and Figure 6 details statistical significance we found with respect to different input texture samples. Most samples agree on the best performance of *Diff*;

- **Across texture samples but separated by participants**: We summed up responses from a particular participant p to a vector $R_3^p(j) = \sum_{s,i} R(s,p,i,j)$ and run the test between each of them. Figure 5(b) summarizes technique preferences, and Figure 7 details statistical significance we found with respect to different participants: the majority agreed on the best performance of *Diff*, while participants #1, #4, #6 as well as #7 showed no preference, and participants #2, #16 as well as #21 provided almost opposite replies from most others.

4.2 Analysis with MDS

In this section, we apply *Multi-Dimensional Scaling* (MDS) on subject response to evaluate response consistency across different texture samples and across different participants. MDS is a set of techniques to visualize information in high dimensional space. One of its variations, metric MDS, projects data points to a low-dimensional configuration while preserving pairwise

	nCut	Diff	Eucl	None
nCut		0	0	4
Diff	14		14	17
Eucl	9	0		14
None	0	0	0	

(a) ANOVA based on 17 textures

	nCut	Diff	Eucl	None
nCut		1	3	6
Diff	12		7	12
Eucl	1	0		7
None	3	2	3	

(b) ANOVA based on 21 testers

Figure 5: Technique preferences based on ANOVA. Each number in the cell indicates how many textures or testers prefer the technique to its left rather than the technique to its top. See Figures 6 and 7 for detailed statistical significance for each texture sample and each tester.

distances. We consider pairwise Euclidean distances between user response and apply metric MDS (shortened as "MDS" hereafter) for subject response analysis.

- **MDS across participants**: We define a 2-D distance matrix $D^{participant}_{21\times21}$ with entries of pairwise distances between any two participants, as

$$D^{participant}_{m,n} = |R(s, p_m, i, j), R(s, p_n, i, j)|_2, \quad m, n = 1, 2, 3, 4; \qquad (1)$$

  where $|\cdot,\cdot|$ is the Euclidean distance. Then we apply MDS on this matrix to extract a 2-D configuration. Figure 8 shows the pairwise distance matrix $D^{participant}$ and its 2-D configuration, where we can clearly see that participants #2, #16, and #21 form a small cluster away from all others. This confirms our observation in Section 4.1 that those participants have quite different response patterns from all others.

- **MDS across texture samples**: Similarly, we define another 2-D distance matrix $D^{sample}_{17\times17}$ with entries of pairwise Euclidean distances

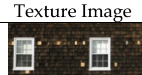
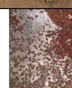
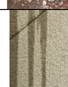
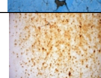
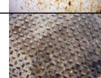
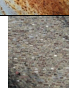
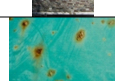
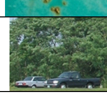
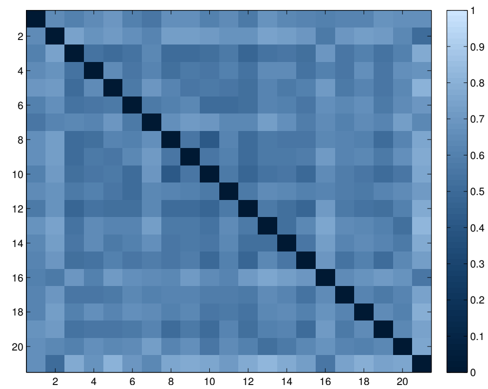
Texture	Texture Image	*nCut*	*Diff*	*Eucl*	*None*
#1		C > N	D > N	E > C E > N	
#2		C > N	D > N	E > N	
#3		C > N	D > N	E > N	
#4			D > C D > E D > N		
#5			D > C D > E D > N	E > N	
#6			D > C D > E D > N		
#7			D > C D > E D > N	E > N	
#8			D > C D > E D > N	E > C E > N	
#9			D > C D > E D > N	E > C E > N	
#10			D > C D > E D > N	E > C E > N	
#11			D > C D > E D > N	E > C E > N	
#12			D > C D > E D > N	E > C E > N	
#13			D > C D > E D > N	E > C E > N	
#14			D > C D > E D > N	E > C E > N	
#15			D > C D > E D > N		
#16			D > C D > E D > N	E > C E > N	
#17		C > N	D > C D > E D > N	E > N	

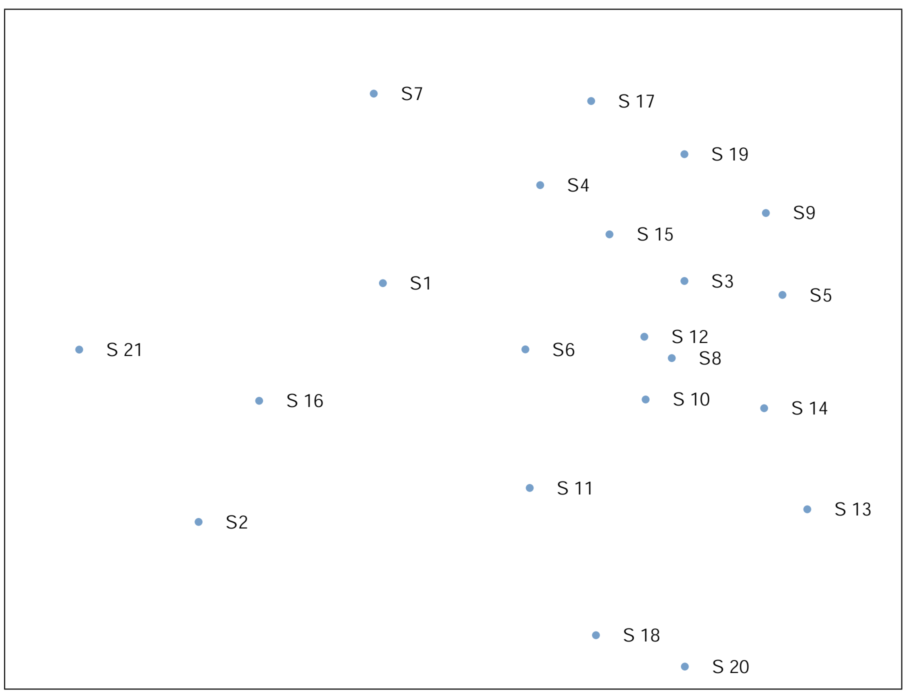
Figure 6: Statistical significance between techniques *nCut*, *Diff*, *Eucl*, *None* across participants but separated by texture samples. Notation of $X > Y$ means that texture sample generated by technique X is significantly preferred than that by technique Y. Comparison not listed were found not statistically significant.

Participant	*nCut*	*Diff*	*Eucl*	*None*
#1				
#2	C > E	D > E		N > C N > D N > E
#3		D > C D > E D > N	E > N	
#4				
#5	C > N	D > C D > N	E > N	
#6				
#7				
#8		D > C D > E D > N		
#9	C > N	D > C D > E D > N	E > N	
#10	C > N	D > C D > N	E > N	
#11		D > C D > E D > N		
#12	C > N	D > E D > N		
#13	C > N	D > C D > N	E > C E > N	
#14		D > C D > N	E > N	
#15		D > N		
#16	C > E	D > E		N > C N > E
#17		D > C		
#18	C > N	D > C	E > N	
#19		D > C D > N		
#20		D > C D > N		
#21	C > D C > E			N > C N > D N > E

Figure 7: Statistical significance between techniques *nCut*, *Diff*, *Eucl*, *None* across texture samples but separated by participants. Notation of $X > Y$ means that texture sample generated by technique X is significantly preferred than that by technique Y. Comparison not listed were found not statistically significant.



(a) Pairwise distance between participants (brighter colors indicate larger differences between responses)



(b) Participants on a 2-D MDS configuration

Figure 8: Participants distribution on a 2-D MDS configuration. We can clearly see outliers of participants #2, #16 and #21 from bright rows/columns in (a) and as an isolated cluster in (b).

between any two samples, as

$$D_{m,n}^{sample} = |R(s_m, p, i, j), R(s_n, p, i, j)|_2, \quad m, n = 1, 2, 3, 4. \qquad (2)$$

We then repeat our analysis as above, and plot pairwise distance matrix D^{sample} and texture sample distribution on MDS 2-D configuration in Figure 9(a). It is not clear how such texture images are organized: there is no obvious image clustering or sorting pattern we can find from this figure.

- **MDS across selected texture samples**: Finally, we exclude response from participants #2, #16, and #21, then repeat the previous analysis. The resulted texture sample distribution in the new 2-D configuration (as showed in Figure 9(b)) seems not significantly different from Figure 9(a) with all participants. We argue that our MDS-based analysis approach is very robust; its analysis result cannot be skewed by up to $3/21 = 15\%$ of outliers among our test participants.

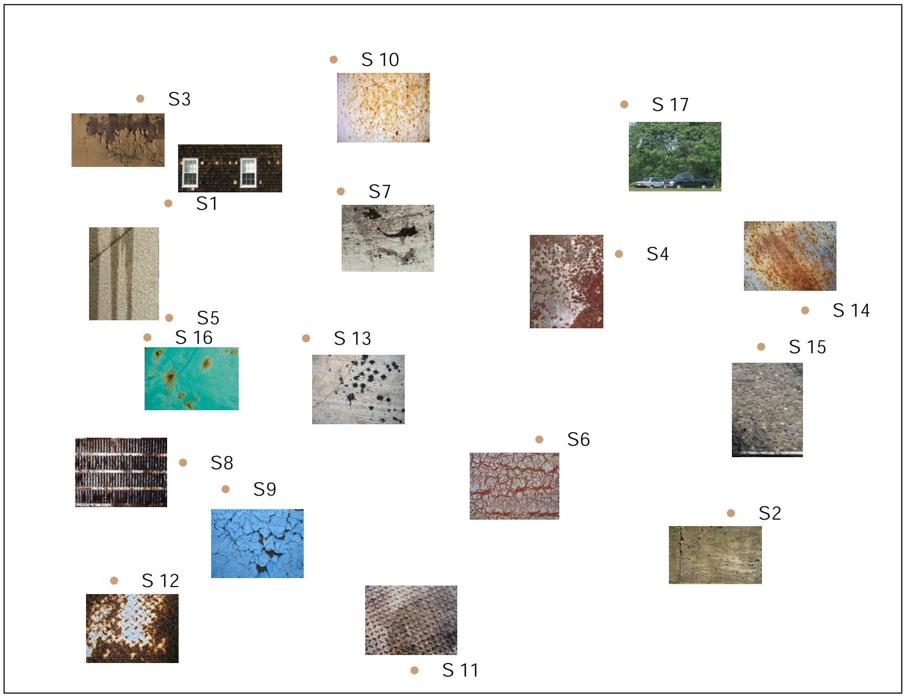
4.3 Analysis with Scaling

In this section, we assume subject response distributed as a normal function and apply Thurstone's Law of Comparative Judgments, Case V, to derive interval scales as quantitative measure of technique performance from pairwise ordinal comparisons [6].

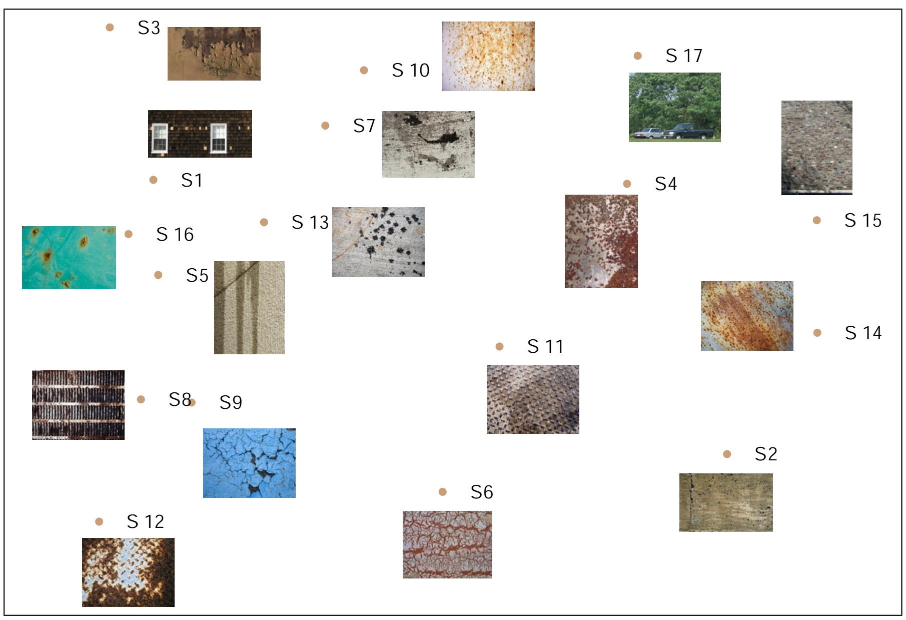
Figure 10 illustrates the work flow of paired comparison scaling, using texture sample #10 in Figure 3 as an example: we first tabulate user response into a frequency matrix M_f (as shown in (a)), where each entry $M_f(i,j)$ shows how many participants prefer synthesis result based on technique j than that based on technique i given two results presented side-by-side; then we convert M_f into a proportion matrix M_p by dividing each entry in M_f by the number of participants (as shown in (b)), and estimate a matrix M_{iCDF} where each entry is the inverse cumulative distribution function (CDF) associated with the standard normal distribution (as shown in (c)); the column averages of M_{iCDF} provide a quantitative measure of relative performance between all four techniques on this texture sample (as shown in (d)): higher the better.

Figure 11 shows scale values of the four techniques with respect to 17 texture samples, from which we can draw some preliminary conclusions:

- *Diff* has the best or above average performance for most texture samples;



(a) Textures on a 2-D MDS configuration with all participants



(b) Textures on a 2-D MDS configuration without participant outliers

Figure 9: Texture samples distribution on 2-D MDS configurations (a) with, and (b) without responses from participant outliers #2, #16, and #21.

	nCut	Diff	Eucl	None
nCut		15	5	5
Diff	6		10	2
Eucl	16	11		5
None	16	19	16	

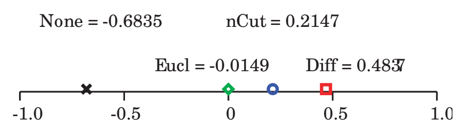
(a) frequencies

	nCut	Diff	Eucl	None
nCut		0.7143	0.2381	0.2381
Diff	0.2857		0.4762	0.0952
Eucl	0.7619	0.5238		0.2381
None	0.7619	0.9048	0.7619	

(b) proportions

	nCut	Diff	Eucl	None
nCut	0	0.5659	-0.7124	-0.7124
Diff	-0.5659	0	-0.0597	-1.3092
Eucl	0.7142	0.0597	0	-0.7124
None	0.7142	1.3092	0.7142	0
col. avg.	0.2147	0.4837	-0.0149	-0.6835

(c) inverse CDF



(d) scale values

Figure 10: Paired comparison analysis on texture sample #10. Diffusion-distance-based technique (*Diff*) is the best among the four techniques.

- *Eucl* and *nCut* have mixed performances;
- *None* has poor performance for more than half of the samples, except for samples that have smooth transition across regions (#4, #9, #12, #14) and that has few outliers (#15);
- Participants cannot tell much difference between all techniques on sample #6 ("cracks").

To further analyze texture samples based on their responses to different techniques, we assign a four-element vector to each texture sample by concatenating scale values from the four techniques, and use MDS to study their distribution on a 2-D configuration (as showed in Figure 12). We also list correlation coefficients between these vectors in Table 1, based on which we have the following observations:

- *Diff* successfully covers most texture samples and is the clear winner among the four techniques. This is further validated from the correlation matrix showed in Table 1, where no significantly positive correlation is found between *Diff* and any other technique;
- *nCut* and *Eucl* have almost *opposite* classifiers, which means they prefer very different input texture samples and can serve as complements with each other (!). This observation is further validated by the matrix showed in Table 1, where correlation coefficient between *nCut* and *Eucl* is close to -1;

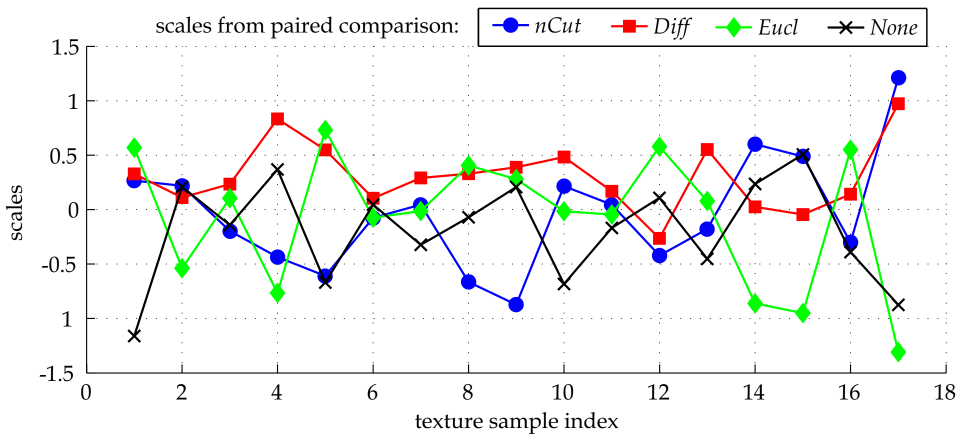


Figure 11: Paired comparison scaling values of the four techniques with respect to all 17 texture samples.

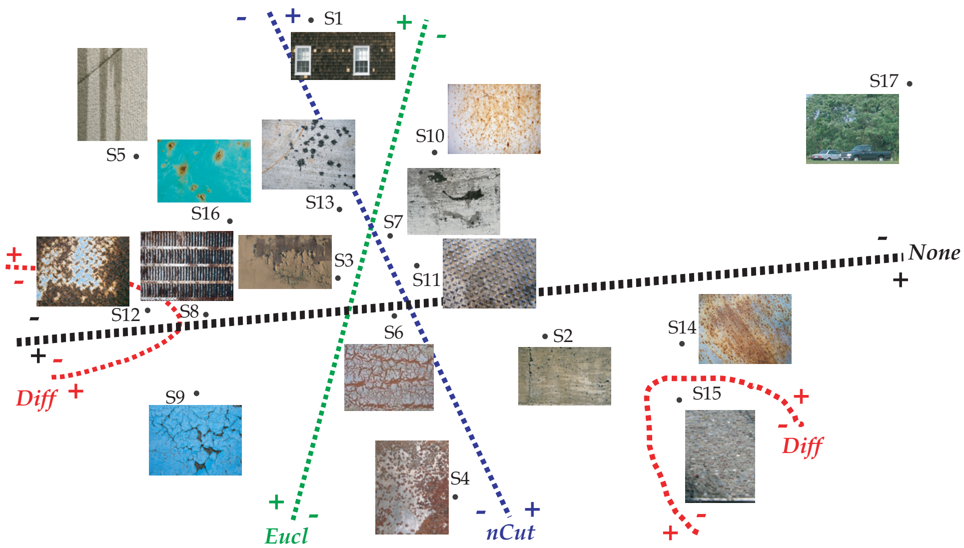
	nCut	**Diff**	**Eucl**	**None**
nCut	-	0.1059	-0.7237	-0.2426
Diff	0.1059	-	-0.2532	-0.4534
Eucl	-0.7237	-0.2532	-	-0.3279
None	-0.2426	-0.4534	-0.3279	-

Table 1: Linear correlation coefficient matrix between paired comparison scaling values using different techniques, with diagonals of 1's omitted. The only significant correlation is -0.7237 between techniques $nCut$ and *Eucl*.

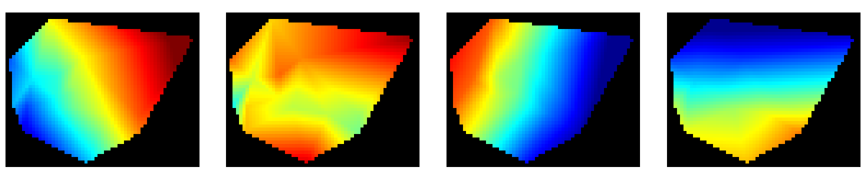
- *None* has poor performance for most of the texture samples. It is also negatively correlated with all three other techniques, which put it at the bottom among the four;
- Texture samples around the crossing area of three dotted lines are difficult to call: scaling values do not have much variances among the four techniques for texture samples #6 and #11. Those are crack pattern and a pattern with smooth transition texture, similar to failure cases we listed in Figure 8 in [22].

5 Texture Features

In Figure 12, we lay out texture samples on a 2-D plane based on human responses to different dominant texture detection techniques. We would like



(a) 2-D configuration of texture samples with MDS based on scaling values



(b) *nCut* (c) *Diff* (d) *Eucl* (e) *None*

Figure 12: Texture samples distribution based on scaling-value-based MDS. (a) 2-D configuration of texture samples. Dotted lines are classifiers that roughly divide texture samples into two parts: those with high scaling values (on the side of the dotted lines with "+" marks) and those with low scaling values (on the side of the dotted lines with "−" marks). Lines in different colors indicate classifiers for different techniques: blue for *nCut*, red for *Diff*, green for *Eucl*, and black for *None*. (b)–(e) Scaling values interpolation in the 2-D configuration space with respect to different techniques. Colors change from red, yellow, green, cyan, to blue as scaling values decrease. Classifiers in (a) roughly correspond to the 0-value contours in (b)–(e).

to identify regions where *Diff* is likely to fail (e.g. textures $S15$ and $S12$), and suggest an alternative technique (perhaps *None* for textures similar to $S15$ and *Eucl* for textures similar to $S12$.) However, it is not clear how to align texture distribution patterns we found with conventional texture features. Solving this problem would likely suggest a way to identify the best technique to detect dominant texture for a new texture sample, and provide more insights about the black box through which a human perceives natural appearance.

To address this problem, we first select a few texture descriptors:

- 6 color descriptors: we consider the average red, green, blue, hue, saturation, and intensity values of each image;
- 9 dimensionality descriptors: we consider the patch size, linear dimensions, diffusion manifold dimensions, and Gaussian kernel sizes as discussed in [22];
- 484 co-occurrence matrix descriptors: we convert input textures into 8-level grey images, create co-occurrence matrices with offsets up to 10 pixels from each direction, and study the *Contrast*, *Correlation*, *Energy*, *Homogeneity* properties of those co-occurrence matrices [12];
- 120 FFT descriptors: we apply 2-D Fourier transform to the grey image and collect response for the first 30 pixel units along the directions of 0, 45, 90 and 135 degrees.

Next, we assume there exists a monotonic function that describes the relationship between a texture descriptor and the performance of a technique, without making any other assumptions about the particular nature of the relationship. To verify this assumption, we estimate Spearman's rank correlation coefficient, ρ, between any pair of texture descriptor and technique performance, as shown in Figure 13.

We can see that the performance of *Eucl* is easier to predict using some dimensionality descriptors. For example, we define the following quantity to describe the nonlinear nature of a texture image:

$$DimNonlinear = 1 - DimDiffmap/DimPCA999, \tag{3}$$

where *DimDiffmap* is the manifold dimension defined in [22], and *DimPCA999* is the number of eigen-patches needed to preserve 99.9% energy based on PCA; the larger the value of *DimNonlinear* is, the more likely a nonlinear manifold is embedded in the feature space. The negative correlation between *DimNonlinear* and *scale-Diff*, as found in Figure 13, reveals

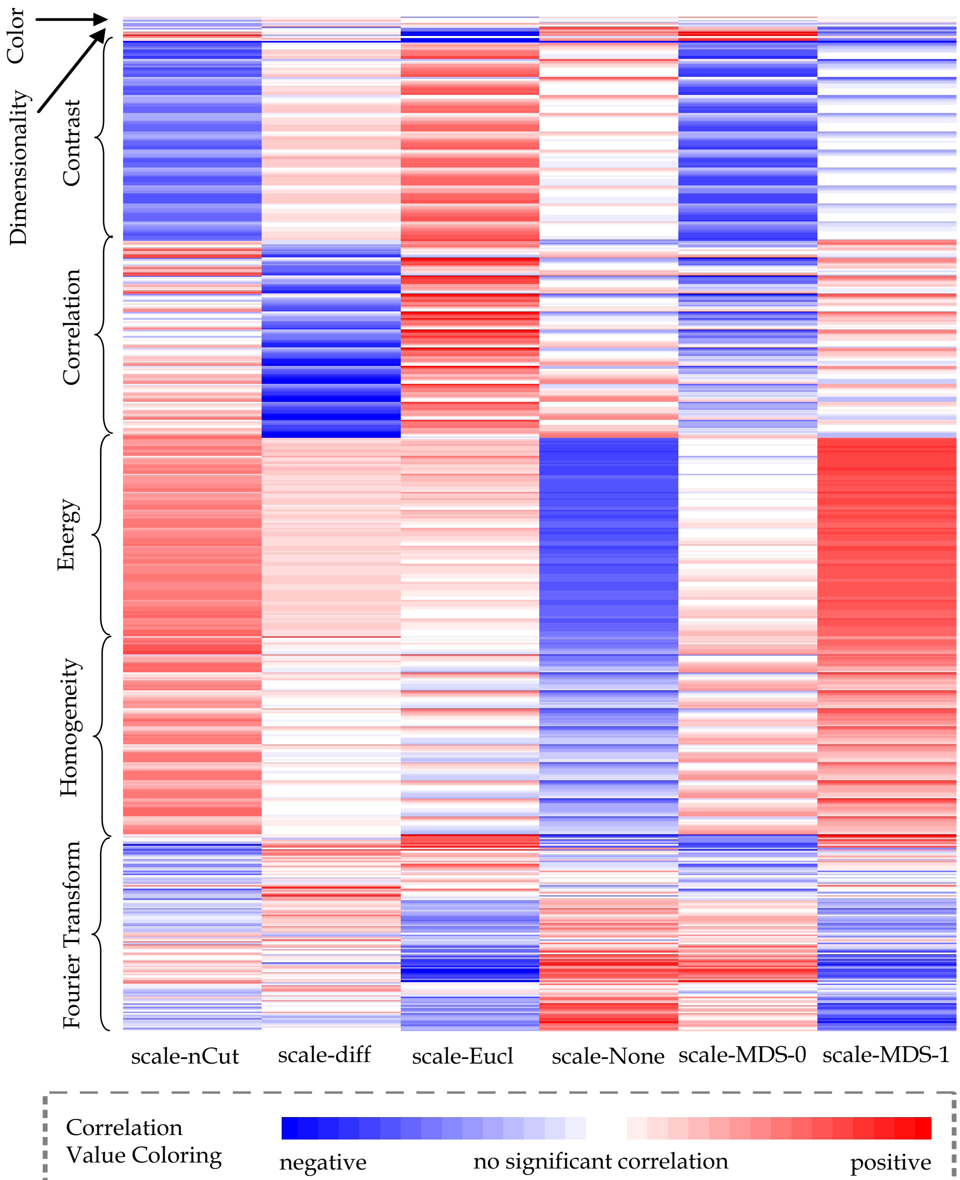


Figure 13: Spearman's rank correlation between texture descriptors and technique performance (*scale-MDS-0* and *scale-MDS-1* are the coordinate of each texture image in Figure 12.) Coefficients are colored from white to red with increasing positive values, and from white to blue with decreasing negative values.

	nCut	Diff	Eucl	None	MDS-0	MDS-1
DimNonlinear			−		+	
Contrast	−		+		−	
Correlation		−	+			
Energy	+			−		+
Homogeneity	+			−		

Table 2: Qualitative correlation between texture descriptors and technique performance for dominant texture detection. "+" indicates positive correlation, "−" indicates negative correlation, and an empty cell indicates that no significant correlation is found.

that the Euclidean-distance-based technique will not work for texture images characterized by manifolds, which is consistent with our intuition.

Moreover, we notice from Figure 13 that scaling values of all four techniques have relatively high correlation with co-occurrence matrix properties, as summarized in Table 2. Especially for *nCut* and *Diff*, which involve complicated and non-linear operations, those co-occurrence matrix properties better predict their performance while simple color descriptors and Fourier transform coefficients fail.

These observations suggest a method for determining an appropriate dominant texture detection technique for a given texture image by estimating technique performance based on proper texture descriptors as studied in Figure 13. Furthermore, as shown in the two rightmost column in Figure 13, we can also use the co-occurrence matrix property values to predict the position of a texture image in the 2-D configuration in Figure 12, using the average *Contrast* value for the first coordinate and the average *Energy* value for the second. Once we place a texture image in such a 2-D configuration, we can select the appropriate technique based on its position with respect to all of the classifiers.

6 Conclusion

Based on our analysis of ANOVA and scaling, we quantitatively confirmed that the diffusion-distance-based technique (*Diff*) proposed in our previous work [22] works for a broad spectrum of texture samples and is the best among all four techniques. Textures of cracks and those with smooth transitions are difficult to analyze; none of the four techniques gives us significantly better result. Also, the combination of paired comparison scaling and multidimensional scaling (MDS) provides quantitative and robust measurement of technique performance in our psychophysical experiment.

Furthermore, we studied the correlation between scaling values and different texture descriptors, and discovered certain dimensionality descriptors and co-occurrence matrix properties that can be used to predict the technique performance. The correlation is qualitative, but it suggests a way to select the appropriate dominant texture detection technique, given a new texture image.

Our work is preliminary and suggests the need for additional studies with more texture descriptors and a wider range of natural texture images. In addition, we can design a more reliable experiment that avoids texture synthesis and focuses only on difference that is due to distance metric selection.

This experiment is our attempt to include the study of the perception process (as illustrated in Figure 1-1 in [21]) which was simply ignored in most previous work. We can further extend such validation throughout the pipeline of appearance capture, analysis, and transfer. Such psychophysical study will help us gain insights before model development, select algorithm and tweak parameters during model development, and evaluate our models after model development.

A Collection of Psychophysical Studies in Graphics

Below, we list a few studies in graphics where researchers designed and conducted psychophysical experiments to gain insights of human perception, to build computational models, to accelerate certain computation, and to validate quantitative models.

- *Reflectance Models*: Pellacini et al. developed a psychophysically-based reflectance model with two perceptually meaningful uniform dimensions [25]; Matusik et al. showed that there are consistent transitions in perceived properties between large set of BRDF samples [24]; Filip et al. studied BTF perception quality and its data compression parameters, and proposed a perception-related metric to automatically determine compression rate on given materials preserving visual fidelity [10];

- *Texture Perception*: Julesz examined whether there are texture channels in human visual system by showing observers computer generated patterns, and found our vision systems are tuned to pick up features of different orientations and spatial frequencies [16]; his work was later exploited for texture synthesis by Heeger et al. [13]; Holten et al. derived a three-dimensional model (including contrast, spatial frequency, and spectral purity) for isotropic textures that is perceivably comparable to the RGB model [14];

- *Color-to-Grayscale Conversion*: Čadík evaluated the accuracy and preference of seven state-of-the-art color-to-grayscale conversions with subjective experiments, using multifactorial analysis of variance (ANOVA) and paired comparison scaling [3];

- *Tone Mapping*: Several researchers conducted separate work of tone mapping algorithms comparison in terms of overall preference and rendering accuracy for high dynamic-range (HDR) scenes [2, 4, 17]; Grave et al. compared tone mapping operators specifically for road visibility [11].

- *Visual Complexity*: Ramanarayanan et al. developed psychophysical experiments to explore overall properties of aggregates (including numerosity, variety, and arrangement), derived metrics to predict when two aggregates have similar appearance, then applied the result to reduce modeling complexity [26,27]. Although the actual problem might be more complicated than a few axes they introduced, their research work, along with other practical applications under development, is believed to be an important step towards realistic rendering [1];

- *Visual Equivalence*: Ramanarayanan et al. explored how object geometry, material, and illumination interact to provide information about appearance, then proposed visual equivalent predictors to improve rendering efficiency by properly blurring and warping illumination maps [28];

- *User Interface*: Lécuyer et al. validated and fine-tuned the *Size and Speed Technique* for simulating pseudo-haptic bumps and holes by a series of psychophysical experiments [19].

- *Visual Realism*: Elhelw et al. investigated the perception of visual realism of static images with different visual qualities based on eye tracking, and categorized image attributes affecting the perception of photorealism [7].

References

[1] Dicussion with Professor Kavita Bala, 2008.

[2] M. Ashikhmin and J. Goyal. A reality check for tone-mapping operators. *ACM Transactions on Applied Perception*, 3(4):399–411, 2006.

[3] M. Čadík. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum*, 27(7):1745–1754, 2008.

[4] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Evaluation of hdr tone mapping methods using essential perceptual attributes. *Computers & Graphics*, 32:330–349, 2008.

[5] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: multiscale methods. *Proceedings of the National Academy of Science*, 102:7432–7437, May 2005.

[6] H. A. David. *The Method of Paired Comparisons*. Oxford University Press, 2nd edition, 1988.

[7] M. Elhelw, M. Nicolaou, A. Chung, G.-Z. Yang, and M. S. Atkins. A gaze-based study for investigating the perception of visual realism in simulated scenes. *ACM Transactions on Applied Perception*, 5(1):1–20, 2008.

[8] J. A. Ferwerda. Psychophysics 101: how to run perception experiments in computer graphics. In *SIGGRAPH'08: ACM SIGGRAPH 2008 classes*, pages 1–60, New York, NY, USA, 2008. ACM.

[9] J. A. Ferwerda, H. Rushmeier, and B. Watson. Psychometrics 101: How to design, conduct, and analyze perceptual experiments in computer graphics. ACM Siggraph Course, 2002.

[10] J. Filip, M. J. Chantler, P. R. Green, and M. Haindl. A psychophysically validated metric for bidirectional texture data reduction. *ACM Trans. Graph.*, 27(5):1–11, 2008.

[11] J. Grave and R. Bremond. A tone-mapping operator for road visibility experiments. *ACM Transactions on Applied Perception*, 5(2):1–24, 2008.

[12] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 3(6):610–621, 1973.

[13] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238. ACM Press, 1995.

[14] D. Holten, J. J. V. Wijk, and J.-B. Martens. A perceptually based spectral model for isotropic textures. *ACM Transactions on Applied Perception*, 3(4):376–398, 2006.

[15] R. Jagnow, J. Dorsey, and H. Rushmeier. Evaluation of methods for approximating shapes used to synthesize 3d solid textures. *ACM Trans. Appl. Percept.*, 4(4):1–27, 2008.

[16] B. Julesz. Visual pattern discrimination. *IRE Transactions on Information Theory*, 8(2):84–92, 1962.

[17] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluating HDR rendering algorithms. *ACM Transactions on Applied Perception*, 4(2):9, 2007.

[18] V. Kwatra and L.-Y. Wei. Example-based texture synthesis. In *SIGGRAPH'07: ACM SIGGRAPH 2007 classes*. ACM, 2007.

[19] A. Lécuyer, J.-M. Burkhardt, and C.-H. Tan. A study of the modification of the speed and size of the cursor for simulating pseudo-haptic bumps and holes. *ACM Transactions on Applied Perception*, 5(3):1–21, 2008.

[20] P. Longhurst, P. Ledda, and A. Chalmers. Psychophysically based artistic techniques for increased percieved realism of virtual environments. In *AFFRIGRAPH 2003, Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, pages 123–131. The Association for Computing Machinery, Inc., February 2003.

[21] J. Lu. *Studies in Texture Generation of Weathered Appearance from Captured Data*. Phd thesis, Yale University, May 2009.

[22] J. Lu, J. Dorsey, and H. Rushmeier. Dominant texture and diffusion distance manifolds. *Computer Graphics Forum*, 28(2):667–676, 2009.

[23] J. Lu, A. S. Georghiades, A. Glaser, H. Wu, L.-Y. Wei, B. Guo, J. Dorsey, and H. Rushmeier. Context-aware textures. *ACM Transactions on Graphics*, 26(1):3, 2007.

[24] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, 2003.

[25] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 55–64, 2000.

[26] G. Ramanarayanan, K. Bala, and J. Ferwerda. Perception of complex aggregates. *ACM Transactions on Graphics*, 27(3):43, 2008.

[27] G. Ramanarayanan, K. Bala, J. A. Ferwerda, and B. Walter. Dimensionality of visual complexity in computer graphics scenes. *SPIE Human Vision and Electronic Imaging (HVEI) '08*, 6806, 2008.

[28] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. Visual equivalence: towards a new standard for image fidelity. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 76, New York, NY, USA, 2007. ACM.

[29] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.