# Towards effective evaluation of geometric texture synthesis algorithms

Zainab AlMeraj*
University of Waterloo
Kuwait University

Craig S. Kaplan
University of Waterloo

Paul Asente
Adobe

## Abstract

In recent years, an increasing number of example-based Geometric Texture Synthesis (GTS) algorithms have been proposed. However, there have been few attempts to evaluate these algorithms rigorously. We are driven by this lack of validation and the simplicity of the GTS problem to look closer at perceptual similarity between geometric arrangements. Using samples from a geological database, our research first establishes a dataset of geometric arrangements gathered from multiple synthesis sources. We then employ the dataset in two evaluation studies. Collectively these empirical methods provide formal foundations for perceptual studies in GTS, insight into the robustness of GTS algorithms and a better understanding of similarity in the context of geometric texture arrangements.

**CR Categories:** I.3 [Computer Graphics]: ;— [I.5]: Pattern Recognition—Design Methodology Pattern Analysis;

**Keywords:** non-photorealistic rendering, texture synthesis, 2D vector graphics, 2D visual perception, user studies, qualitative and quantitative evaluation methods

## 1 Introduction

Example-based Geometric Texture Synthesis (GTS) refers to a class of algorithms that generate a large arrangement of vector elements from a small input arrangement called an *exemplar*. Roughly speaking, the goal is the same as it is with raster-based texture synthesis: the output arrangement should be judged by a human viewer to be "similar" to the exemplar. The challenge is to define similarity in a way that is rigorous enough to be formalized as an algorithm, while still conforming to human perceptual judgments.

We have seen a positive trend of applying formal evaluation methods in the validation of new algorithms in non-photorealistic rendering (NPR), but this trend has not caught on in the field of GTS. Many GTS algorithms have been proposed, all of which seem to produce reasonable results across a range of inputs. But at best, authors run their algorithm on an exemplar from a previous paper by others, and show the old and new outputs side by side. We believe that there is a need for effective evaluation strategies in GTS, which can be applied to compare existing algorithms and validate new ones. Hence our high-level goal in this paper is to establish a practical evaluation methodology for GTS algorithms.

AlMeraj et al. [2011] conducted the first study that probed the nature of similarity in the perception of geometric textures. Their investigation resulted in a descriptive list of visual features that people use to explain the similarity between synthesized arrangements and exemplars. Building on their work, this paper attempts to push our understanding of texture similarity even further. We gather a comprehensive dataset of geometric textures (Section 3) from several different *synthesis sources*: expert human designers, state-of-the-art synthesis algorithms, and simple randomly generated textures. We then conduct two user studies based on this dataset (Sections 5–6), in order to see whether human judgments of similarity between synthesized textures and exemplars can be used to assess the performance of different synthesis sources. Using results from the studies we attempt a small evaluation (Section 7). We believe that the dataset and the evaluation methodologies will be useful to others in the GTS field, and will suggest analogous studies that could be applied in other areas of NPR.

## 2 Related work

### 2.1 Geometric texture synthesis

Current GTS algorithms use various combinations of procedural growth, statistics and perceptual foundations to gather layout information about individual motifs from exemplars and utilize them to synthesize larger similar arrangements.

Barla et al. [2006] were the first to contribute a 2D geometric texture synthesis algorithm. Their method adopts a non-parametric statistical method on an exemplar to capture the spatial distribution. Hurtut et al. [2009] devised a statistical appearance-based approach to GTS modelling concepts from gestalt grouping theory.

Alves dos Passos et al. [2010] and Ijiri et al. [2008] use similar procedural growth approaches to enhance the appearance of results for a variety of texture styles. The method by Jenny et al. [2010] synthesizes regular and irregular arrangements while simultaneously resolving overlaps and appearance issues.

The algorithm by Ma et al. [2011] is able to synthesize 2D and 3D results using a complex energy-based optimization process designed to mimic both appearance and distribution properties found in exemplars. A subsequent geometric synthesis algorithm by AlMeraj et al. [2013] uses a patch-based method to achieve global and local distributions similar to those in the exemplar.

A recent statistical approach by Öztireli and Gross [2012] uses a second-order statistic called the Pair Correlation Function (PCF) as a guide to achieve global similarity. Given one or more exemplar inputs, they are able to synthesize 2D and 3D arrangements either by using a generalized dart throwing routine, or by fitting an arrangement to the PCF by gradient decent.

Öztireli and Gross offer quantitative evidence for their claims of similarity by including charts showing PCF curves and irregularity measures for synthesized and target arrangements. These quantitative measures reduce subjectivity in comparing synthesized arrangements, and move us a step closer towards understanding similarity in GTS. However, proving whether or not these statistical measures give an effective account of how humans judge similarity is difficult. In this paper, we address the subjectivity involved in similarity judgements and hope that our insights help researchers develop an appropriate definition of similarity for GTS in the future.

*e-mail: z.almeraj@gmail.com


NPAR 2013, July 19 – 21, 2013, Anaheim, California.

### 2.2 Evaluation in texture synthesis

Evaluation has always been a challenging problem in the graphics community. Due to the broad nature of the algorithms and the different sub-areas involved, very few useful evaluation methods have been proposed. In pixel-based texture synthesis, evaluation involves running quantitative metrics on synthesized results to measure the amount of pixel-level similarity to example inputs [Wei et al. 2009]. These types of measures rely on a uniform spatial domain, in which a synthesized texture can be analyzed as a sampled signal. They cannot generally be adapted to a freeform arrangement of geometric primitives, as in GTS. Below we list some previous work on the subject of evaluation in pattern recognition and non-photorealistic rendering.

Lin et al. [2006] present a quantitative evaluation of regular and near-regular image-based textures. They compare the performance of four synthesis algorithms to understand how much a near-regular texture's global regularity and local randomness affects human judgement. They develop a statistical score to measure regularity through user-defined translation vectors. In addition to this quantitative evaluation, Lin et al. conducted a supporting subjective evaluation to determine the significance of global regularity of textures on participant similarity ratings. Participants are presented with an exemplar and two textures on a computer screen and asked to provide a similarity ranking of 1 to 4. The findings suggest a bias in favour of one of the synthesis algorithms adopted. The results also support the regularity metric as a reliable evaluation measure of structural similarity.

Isenberg et al. [2006] investigate the quality of automated pen-and-ink algorithms by comparing computer-generated to hand-drawn (artist) images. In their study, participants were given collections of images printed on paper and instructed to separate them into piles according to their own criteria. The results highlight differences between hand-drawn and computer-generated images, as well as positive aspects of both. A similar pile-sorting strategy has been used in computer vision for classifying natural textures into meaningful categories [Balas 2008]. In these studies, the unrestricted comparisons allow participants to accomplish the task at their own pace without external influences. We believe that this experimental strategy shows promise for the analysis of geometric arrangements.

Recent research by AlMeraj et al. [2011] offers the first step in the NPR literature towards understanding geometric arrangements in light of human visual perception. They conducted two psychophysical studies, and through analysis of the results they produced a collection of qualitative strategies and visual cues used in making texture similarity judgments. The results are valuable in understanding similarity between arrangements and we refer to them throughout this work. AlMeraj et al. also highlight a growing need to establish a plausible suite of benchmark samples that future algorithms can use to evaluate similarity of geometric synthesis results.

## 3 A geometric texture benchmark

To allow for more effective comparisons of GTS algorithms we collect a dataset of synthesized arrangements. Our goals are to use this collection as a benchmark for evaluating existing and future GTS algorithms; to further elucidate the meaning of "similarity" in the context of geometric textures; and to determine the progress and shortcomings of geometric texture synthesis as a research area.

To select our exemplars, we chose to adapt four source arrangements from the US Geological Survey (USGS) Digital Cartographic Standard for Geologic Map Symbolization [US FGDC 2006]. This resource contains textures used to indicate different features in geological maps. Jenny et al. [2010] designed a tool that helps cartographers generate maps with similar features. Similar artificial textures have also been used as input by recent GTS algorithms [Alves dos Passos et al. 2010], making the dataset a suitable candidate for future experimentation.

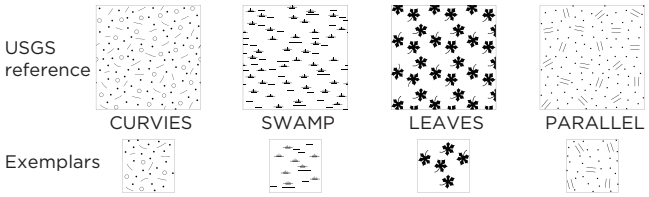


**Figure 1:** *The original source arrangements and the extracted (preprocessed) exemplars.*

We identified four distinct *patterns* in the USGS standard that we take to be representative and that use relatively few distinct motif shapes. As shown in Figure 1, we name them CURVIES, SWAMP, LEAVES and PARALLELS. From each square texture, we constructed a smaller exemplar by extracting motifs contained entirely within a square sub-region with half the side length of the original.

Armed with these four exemplars, we set about collecting a diverse set of actual arrangements constructed from them. For each exemplar we gathered a set of eleven arrangement results from three sources: human experts (Section 3.1), existing GTS algorithms (Section 3.2), and a simple pseudorandom approach (Section 3.3). We describe the collection of this data in following subsections. To encourage others to add to this benchmark with the results of other algorithms, we have made our dataset publicly available.[1]

### 3.1 Arrangement collection from expert designers

In order to compare fairly between computer-generated and hand-generated arrangements, we recruited expert human designers to draw large arrangements from our four exemplars. Human designers have a keen eye for texture, composition, layout and design, providing us with a rich set of subjective interpretations of the synthesis task. We found experts by word of mouth and by advertising on a forum for expert users of vector illustration software. Participants were required to have extensive experience in their field and keen aesthetic judgement. A total of four people qualified for the study; each had over 9 years of experience. Hereinafter we identify them and their arrangements as **H1**–**H4**.

To collect human-generated arrangements we created a self-contained template in the form of an Adobe® Illustrator® document. A copy of this document is available in the supplementary materials for this paper. The template describes the synthesis task as follows: "Given a small sample of arranged symbols, place copies of the symbols into a large area so that the overall impression of the larger arrangement is like the smaller one". Below that, the template includes a completed example. Four empty regions appear below, one for each of the USGS exemplars. Next to each region is a copy of the exemplar, and copies of symbols for the distinct motif shapes used in that exemplar.

Each participant received an information letter to sign, the template (in PDF and Adobe® Illustrator® formats) and a questionnaire. Their results can be seen in Figure 2. Participants were compensated for their efforts.

[1] GTS dataset: `http://www.cgl.uwaterloo.ca/~zmeraj/GTS/geometric_arrangement_dataset.html`

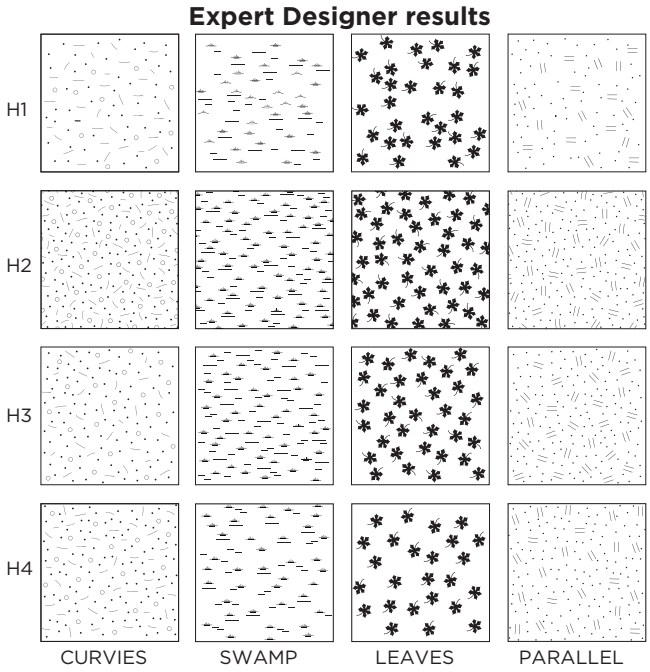


**Figure 2:** *Results gathered from our expert designers.*

| ID | Algorithm | ID | Algorithm |
|---|---|---|---|
| **A1** | Alves dos Passos et al. [2010] | **R1** | Pseudorandom |
| **A2** | Hurtut et al. [2009] | | (Section 3.3) |
| **A3** | Ma et al. [2011] | **R2** | Pseudorandom |
| **A4** | AlMeraj et al. [2013] | | (Section 3.3) |
| **A5** | AlMeraj et al. [2013] | | |

**Table 1:** *Algorithm labels and their corresponding authors.*

### 3.2 Arrangement collection from GTS algorithms

One problem with attempting a robust evaluation of GTS algorithms is the difficulty of acquiring and developing the actual implementations. Reimplementing existing synthesis algorithms is difficult because they often includes ad hoc fine tuning. Without the expertise of the original creators of these algorithms, their true value can be obscured. To make comparison results valid, it is important to use original algorithms to synthesize new arrangements [Lin et al. 2006].

We solicited generated arrangements from the four most recent GTS algorithms, which cover a spectrum of approaches [Alves dos Passos et al. 2010; Hurtut et al. 2009; Ma et al. 2011; AlMeraj et al. 2013]. Each algorithm author was sent the four exemplars via email, in the format required by their algorithm. The authors synthesized larger arrangements, adhering to the same criteria they used when generating their previously published arrangements. Their results are shown in Figure 3 and referred to as **A1**–**A5** as shown in Table 1. Synthesis sources **A4** and **A5** are both from AlMeraj et al. [2013]; they were generated using square and hexagonal arrangements of tiles, respectively.

To enable the gathering of this benchmark we were obliged to support the individual practices of each algorithm, as each had different input requirements. Some algorithms required text files with point locations and IDs of motifs, while others required specific vector formats. It would be easier to compare GTS algorithms if the community were to agree on a common input standard. We recommend the SVG-based format used by AlMeraj et al. [2013]. SVG supports the full range of standard vector primitives, as well as a simple instancing mechanism via the `<symbol>` and `<use>` tags.

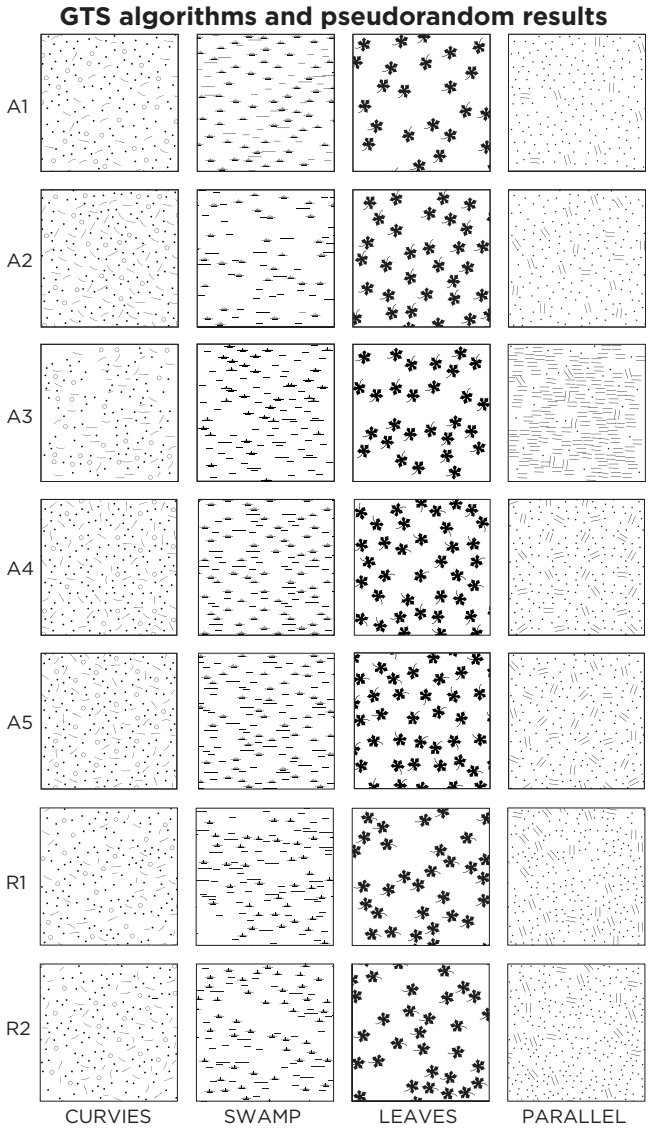


**Figure 3:** *Results gathered from the GTS and psuedo-random algorithms.*

### 3.3 Pseudorandom texture arrangements

To test whether random arrangements would be perceived differently from the results of other synthesis sources, we include two pseudorandom arrangements per exemplar in the dataset. We developed a simple randomized synthesis algorithm and used it generate arrangements labelled **R1** and **R2**.

Let $d$ refer to the minimum distance between centroids of motifs in the exemplar, and let $\rho$ be the density of the exemplar, i.e., the fraction of the exemplar covered by motifs. We choose a random point $P$ within the synthesis region, and a random motif from the exemplar to place there. We perform two tests on this proposed

motif placement:

- If the distance from this point to any other placed motif centroid is less than $d$, we reject $P$.
- Center a window on $P$ with the same shape as the exemplar. If the density of the synthesized arrangement within the window exceeds $\rho$, we reject $P$.

If the point $P$ passes these tests, we place the chosen motif there. We iterate this process until the overall density of the synthesized arrangement comes within a threshold of $\rho$.

## 4 Evaluating synthesized arrangements

Using the benchmark of synthesized texture arrangements shown in Figures 2 and 3, we now wish to compare the eleven synthesis sources. We move beyond the subjective practices currently used, and explore a more effective study-based methodology that supports rigorous investigations into perceptual similarity. Insights gathered from these investigations will help guide researchers towards effective practices for evaluating GTS algorithms.

Our investigation is divided into two parts, discussed in the two sections that follow. In the first study we conduct an observational pile-sorting study and watch how human subjects sort arrangements based on their similarity using printed cards on a flat surface (Section 5). In the second study we conduct a pairwise comparison test using a computer interface. Participants are given pairs of synthesized arrangements and asked to click on the arrangement they believe is most similar to an exemplar (Section 6).

For these studies we recruited twenty participants (undergraduate and graduate students), with no previous experience with geometric textures or with our research. Participants were compensated for their efforts.

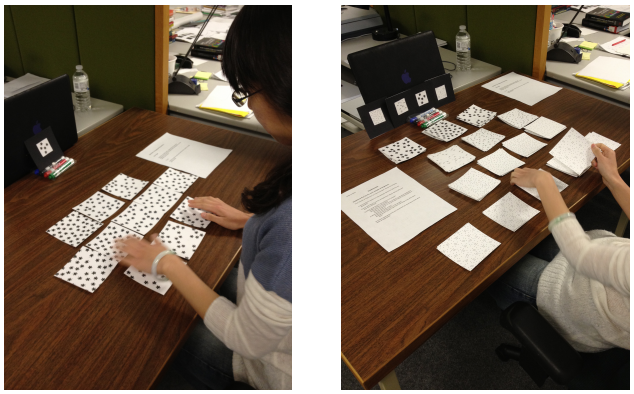

**Figure 4:** *The setup for the pile sorting study with a participant distributing the 11 piece card set while seated at a distance from the exemplar source (left). After sorting all arrangement into piles the participant discusses them with the investigator (right).*

## 5 Pile-sorting synthesized arrangements

Pile-sorting is effective for gathering qualitative data such as user observations [Weller and Romney 1988]. It can also be supported by feedback gathered through short semi-structured interviews. It is particularly suitable when there are few quantifiable measures suitable for analyzing the target material.

This qualitative style of analysis was previously adopted by Isenberg et al. [2006] to understand how people judge similarity between hand-drawn and computer-generated pen-and-ink drawings. We use a similar approach to compare between multiple synthesized arrangements and their sources.

*Card preparation*: We created 44 cards from our dataset, eleven for each of the four patterns. Each had the arrangement printed and glued to a $12\text{cm} \times 12\text{cm}$ square of cardstock. We created a card of the same size for each exemplar, printed at $6\text{cm} \times 6\text{cm}$ inside a black border.

*Setup*: Each participant was asked to sit on a chair in front of a large flat table surface. The exemplar source was placed approximately $100\text{cm}$ away from the participant, as shown in Figure 4. Because synthesized arrangements depend so strongly on their exemplars, and because of the diversity of arrangements for each exemplar, we opted to show the exemplar cards as a reference during the pile sorting study.

At the beginning of the study, we provided a set of cards and instructed the participants to read the task provided to them on a white sheet of paper, ask questions, and begin when ready. The sorting task was described as follows: "Using the provided cards, create piles that represent categories that show how similar each arrangement is to the sample input shown".

*The methodology*: We adopt an unconstrained pile-sorting task in which participants could make as many piles as they wanted without any time restrictions. They were encouraged to provide their thoughts during and after the study. To ensure that we collect enough data for comparisons we suggested that participants create at least two piles and minimize the number of singleton piles when possible.

At the start of the study we provided participants with a random card set (either CURVIES, SWAMP, LEAVES, or PARALLEL) and let them generate piles using their own criteria. Most participants distributed the cards across the table before piling them, making it easier to notice differences and similarities between the cards. Once they completed piling the first card set, the piles were pushed to the side of the table and the participants were handed a card set with a different pattern. This was repeated for each card set. Participants created an average of four piles with a standard deviation of one for each card set.

For the interview, the piled cards were moved closer to the participant in the same order they were presented. The investigator initiated a discussion by handing the participant a sheet of paper containing some questions.

*Data collection*: The piles of arrangements were recorded via note taking by the investigator. During the pile sorting task and semi-structured interview, participants were audio recorded. The pile sorting task took an average of 14 minutes, while the semi-structured interview and discussions that followed took an average of 8 minutes.

In the following subsections we analyze the results of the pile-sorting experiment in two parts. We first analyze the generated piles according to the four source patterns separately (Section 5.1) and then analyze the data according to the synthesis sources (Section 5.2). These are followed by a summary of findings gathered from participant interviews (Section 5.3).

### 5.1 Pile-sorting according to arrangement patterns

To understand the resulting piles we created a similarity matrix for each participant's piling of each pattern. Similarity matrices are created by tabulating the co-occurrences of synthesized arrangements found in each pile. If a participant grouped the cards from

| | H1 | H2 | H3 | H4 | A1 | A2 | A3 | A4 | A5 | R1 | R2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | 20 | 0 | 1 | 2 | 4 | 2 | 9 | 0 | 0 | 1 | 1 |
| H2 | | 20 | 3 | 5 | 1 | 11 | 1 | 8 | 16 | 4 | 6 |
| H3 | | | 20 | 9 | 10 | 4 | 4 | 2 | 2 | 5 | 5 |
| H4 | | | | 20 | 6 | 3 | 4 | 5 | 4 | 4 | 7 |
| A1 | | | | | 20 | 3 | 5 | 5 | 2 | 7 | 4 |
| A2 | | | | | | 20 | 2 | 9 | 8 | 4 | 6 |
| A3 | | | | | | | 20 | 1 | 1 | 5 | 2 |
| A4 | | | | | | | | 20 | 9 | 6 | 7 |
| A5 | | | | | | | | | 20 | 5 | 7 |
| R1 | | | | | | | | | | 20 | 9 |
| R2 | CURVIES | | | | | | | | | | 20 |
| Most | 4 | 1 | 7 | 8 | 8 | 2 | 1 | 6 | 3 | 7 | 6 |
| Least | 9 | 4 | 1 | 3 | 2 | 4 | 12 | 3 | 5 | 4 | 4 |

| | H1 | H2 | H3 | H4 | A1 | A2 | A3 | A4 | A5 | R1 | R2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | 20 | 1 | 0 | 6 | 0 | 13 | 5 | 1 | 2 | 6 | 6 |
| H2 | | 20 | 15 | 2 | 16 | 0 | 0 | 16 | 9 | 0 | 2 |
| H3 | | | 20 | 3 | 13 | 0 | 1 | 14 | 10 | 3 | 4 |
| H4 | | | | 20 | 2 | 3 | 7 | 4 | 2 | 8 | 5 |
| A1 | | | | | 20 | 1 | 1 | 13 | 9 | 0 | 3 |
| A2 | | | | | | 20 | 5 | 1 | 3 | 8 | 6 |
| A3 | | | | | | | 20 | 1 | 3 | 11 | 9 |
| A4 | | | | | | | | 20 | 10 | 2 | 2 |
| A5 | | | | | | | | | 20 | 3 | 5 |
| R1 | | | | | | | | | | 20 | 10 |
| R2 | SWAMP | | | | | | | | | | 20 |
| Most | 11 | 2 | 3 | 7 | 1 | 7 | 8 | 5 | 8 | 6 | 7 |
| Least | 1 | 16 | 13 | 2 | 16 | 3 | 2 | 11 | 6 | 1 | 2 |

| | H1 | H2 | H3 | H4 | A1 | A2 | A3 | A4 | A5 | R1 | R2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | 20 | 1 | 2 | 7 | 3 | 4 | 10 | 4 | 2 | 6 | 11 |
| H2 | | 20 | 14 | 1 | 3 | 6 | 2 | 7 | 16 | 1 | 2 |
| H3 | | | 20 | 2 | 1 | 6 | 2 | 10 | 11 | 2 | 1 |
| H4 | | | | 20 | 7 | 8 | 9 | 3 | 2 | 5 | 5 |
| A1 | | | | | 20 | 4 | 5 | 2 | 2 | 7 | 6 |
| A2 | | | | | | 20 | 5 | 11 | 4 | 5 | 5 |
| A3 | | | | | | | 20 | 2 | 3 | 9 | 5 |
| A4 | | | | | | | | 20 | 8 | 5 | 3 |
| A5 | | | | | | | | | 20 | 2 | 2 |
| R1 | | | | | | | | | | 20 | 11 |
| R2 | LEAVES | | | | | | | | | | 20 |
| Most | 10 | 2 | 3 | 6 | 10 | 1 | 8 | 2 | 2 | 5 | 10 |
| Least | 2 | 15 | 11 | 2 | 3 | 4 | 1 | 4 | 12 | 6 | 6 |

| | H1 | H2 | H3 | H4 | A1 | A2 | A3 | A4 | A5 | R1 | R2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | 20 | 2 | 2 | 4 | 7 | 5 | 1 | 2 | 2 | 2 | 3 |
| H2 | | 20 | 14 | 6 | 1 | 6 | 1 | 16 | 15 | 5 | 0 |
| H3 | | | 20 | 9 | 1 | 8 | 1 | 16 | 15 | 2 | 0 |
| H4 | | | | 20 | 5 | 17 | 0 | 6 | 9 | 3 | 3 |
| A1 | | | | | 20 | 6 | 2 | 1 | 1 | 7 | 10 |
| A2 | | | | | | 20 | 0 | 7 | 9 | 2 | 2 |
| A3 | | | | | | | 20 | 0 | 0 | 0 | 2 |
| A4 | | | | | | | | 20 | 15 | 2 | 1 |
| A5 | | | | | | | | | 20 | 2 | 0 |
| R1 | | | | | | | | | | 20 | 14 |
| R2 | PARALLEL | | | | | | | | | | 20 |
| Most | 3 | 10 | 9 | 8 | 4 | 9 | 0 | 9 | 11 | 5 | 4 |
| Least | 3 | 2 | 2 | 2 | 2 | 1 | 16 | 0 | 1 | 2 | 2 |

**Figure 5:** *Correlations showing the number of times arrangement patterns were grouped together. Pairings that occurred ten or more times are highlighted in red. Each table is labelled with the corresponding pattern name. The two rows at the bottom of each table indicate the number of participants who placed a given synthesis source into their Most similar or Least similar piles.*

two synthesis sources into a pile, we place a 1 in the corresponding matrix entry; otherwise we place a 0 there.

For each pattern, we combine the similarity matrices for all twenty participants, as shown in the tables of Figure 5. In the combined matrices, each entry represents the number of participants who placed a combination of sources into the same pile. Higher scores in these tables imply that the arrangements share similar characteristics, while lower scores imply dissimilarity.

Once pile-sorting was complete, we asked participants to indicate which piles of cards were most and least similar to the exemplar; answers are tabulated in the bottom two rows of each table in Figure 5. We discuss the reasons behind participants' choices in more detail later in Section 5.3.

In our analysis we found that some synthesis source correlations varied from one pattern to another, suggesting that similarities differed depending on the patterns. In SWAMP, for example, arrangements by **H2** and **A1** had the highest correlation while in CURVIES they were less correlated.

Common trends found in all patterns include high correlations between **H2** and **H3**, suggesting that these two experts recognized and used similar features in constructing their arrangements. Various low correlations amongst the four designer results highlight the subjectivity problem present in GTS research.

Interestingly, arrangements by **H2** and **H3** correlate highly with arrangements by **A4** and **A5** and with some arrangements by **A1**. This consistency implies that similar pattern characteristics were featured by these sources.

Even though we found high scores between **H2** and **H3** and between **H2** and **A2**, none were considered most similar to the exemplar. However, arrangements by **H3**, **A1** and **H4** had lower correlations but were chosen to be more similar. Comparable observations for the remaining tables suggest that the piling decisions are not random and that participant similarity judgements were unambiguous.

Pseudorandom LEAVES arrangements were chosen as most similar by ten participants. These similarity choices may have been influenced by this algorithm's strong emphasis on achieving the same density as the exemplar. This is an important observation that demands further investigation into the significance of pseudorandom algorithms and density for GTS.

Arrangements CURVIES and PARALLEL by **A3** stood out as least similar in the study. These arrangements are less uniform and contain different motif ratios to those present in the exemplar, explaining participant decisions.

### 5.2 Pile-sorting according to synthesis sources

The previous analysis provides us with an overview of common groupings that occurred when participants compared synthesized arrangements according to the four patterns. To get a more general intuition of the piles independent of the patterns we analyze the data in terms of the number of participants found to have piled synthesis source arrangements together. The goal is to highlight consistencies in the data and explain similarities and differences between the synthesis sources. A table of participant choices along with an accompanying visualization—a 2D dendrogram—is shown in Figure 6.

A dendrogram visualization is the result of hierarchical clustering performed on pairwise distances calculated from the table data. We calculate a chi-squared measure using pair average linkage to measure dissimilarity between every two sources of synthesis. This de-

| | H1 | H2 | H3 | H4 | A1 | A2 | A3 | A4 | A5 | R1 | R2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | 20 | 4 | 4 | 14 | 13 | 15 | 15 | 6 | 4 | 12 | 13 |
| H2 | | 20 | 19 | 11 | 16 | 15 | 4 | 19 | 19 | 8 | 9 |
| H3 | | | 20 | 14 | 15 | 13 | 7 | 19 | 20 | 9 | 11 |
| H4 | | | | 20 | 13 | 17 | 16 | 9 | 13 | 12 | 12 |
| A1 | | | | | 20 | 11 | 11 | 14 | 13 | 16 | 13 |
| A2 | | | | | | 20 | 8 | 16 | 15 | 11 | 15 |
| A3 | | | | | | | 20 | 6 | 6 | 16 | 14 |
| A4 | | | | | | | | 20 | 20 | 10 | 10 |
| A5 | | | | | | | | | 20 | 10 | 13 |
| R1 | | | | | | | | | | 20 | 18 |
| R2 | | | | | | | | | | | 20 |

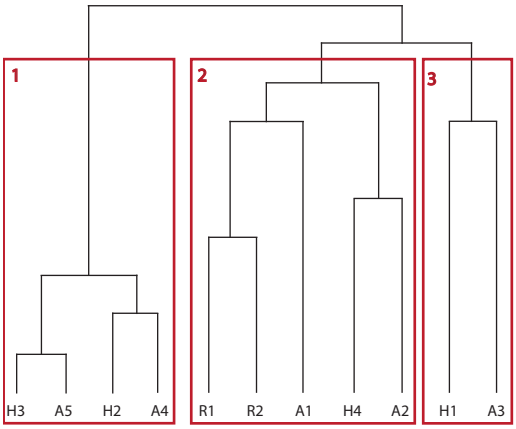


**Figure 6:** *Left: The pile-sorting correlation table shows the number of participants that have piled cards of arrangement sets together at least once. The highest correlation scores are highlighted in red. Right: A 2D dendrogram showing the cluster results of a hierarchical clustering analysis of sorting piles from the study based on the table.*

scriptive analysis method is common for interpreting values found in similarity matrices [Weller and Romney 1988].

The dendrogram shows how arrangements along the $x$-axis merge and divide. Along the $y$-axis we see how far apart the merging happens. Linked arrangements near the bottom of the $y$-axis imply frequent placements of arrangements in one pile. Those linked higher up the $y$-axis and farther apart are found together less often, hence less consistent. The linkages result in three clusters of arrangement sets which are derived purely from participants' similarity choices.

For example, in Figure 6 **H3** and **A5** were piled together by all twenty participants, so they are connected low on the $y$-axis. In contrast, **A3** and **H2** were piled together by only four participants, leading to a linkage high on the $y$-axis. In the following points, we discuss the contents of each cluster:

- **Cluster 1**: In this cluster we have four synthesis sources: **H2**, **H3**, **A4** and **A5**. Perceptual characteristics captured by all these sources result in a larger number of co-placements by a majority of the participants. This cluster differentiates these four sources from the rest of the synthesis sources in terms of their appearance.
- **Cluster 2**: This cluster contains five synthesis sources: **H4**, **A1**, **A2**, **R1** and **R2**. Notice that arrangements by **R1** and **R2** are consistently correlated by many participants, as are **H4** and **A2**. This shows that participants are meticulous at deciphering commonalities between arrangements causing them to distinguish arrangements by **R1** and **R2** as coming from a similar source. The same observation applies for **H4** and **A2**.
- **Cluster 3**: This cluster contains two synthesis sources: **H1** and **A3**. The linkage between these sources is higher up the y axis, implying that they are less consistent than their neighbouring sources. Although a total of 15 participants were found to pile arrangement from these sources together, the linkage suggests that some patterns could be correlated more than others. We notice that this is true for arrangement patterns CURVIES and LEAVES in the tables of Figure 5.

In summary, synthesis sources **H2** and **H3**, **A4** and **A5**, and **R1** and **R2** were more consistent in achieving higher similarity correlations with one another than other sources. The pseudorandom sources **R1** and **R2** are successfully distinguishable as originating from the same source, so are sources **A4** and **A5**. Arrangements by other GTS synthesis sources are harder to distinguish as coming from a similar source. The different patterns used for this study may have influenced these findings. For example, the dissimilarity between synthesis sources **H1** and **A3** was clearly evident for only two of the patterns.

### 5.3 Semi-structured interview

Once participants finished sorting all the card sets, the piles were brought back and placed across the table in four rows according to the patterns (Figure 4) and participants were provided a sheet containing the questions. We decided to leave questioning until after the pile-sorting task was complete to eliminate biases.

We asked three open-ended questions targeting the thoughts and decisions participants made during the study. The qualitative information gathered from our interview helped us elucidate the visual factors participants felt important when depicting similarity as well as their overall confidence during card sorting. We repeated the same questions, in order, four times for each participant (once per pattern). The answers to these questions are discussed in the three subsections that follow.

#### 5.3.1 How would you explain the rationale or logic behind the piles that you generated?

Over the course of the pile-sorting study, participants were observed to use different sorting criteria. The criteria reported are summarized in order from most to least common in Table 2.

Of the 20 participants, 19 singled out density as one of the main factors they used when sorting the cards. This observation is consistent with previous GTS studies [AlMeraj et al. 2011], in which density was identified as a crucial visual cue in texture perception. Variation in ratios of distinct motif shapes also influenced some participants in their decisions to group them separately. This was apparent for arrangement patterns PARALLEL, CURVIES and SWAMP but not for LEAVES which had only one motif. We believe that further research is needed to understand the importance of density and motif ratios in texture similarity judgments.

Twelve participants mentioned the identification of noticeable patterns in exemplars. This involved either holding the card out near the exemplar and deciding whether it was a good extension to the

small sample or locating small groups of motifs distributed in ways similar to groups in the exemplar.

Orientation cues were used occasionally, particularly when sorting the LEAVES and CURVIES cards. The leaf motifs in the LEAVES exemplar exhibit only three rotation angles, which some participants interpreted as significant. Lines in CURVIES appeared to have a principal orientation in some arrangements, which also influenced participant judgements. This behaviour was not noticed with SWAMP. Some arrangements were explicitly sorted according to how regular and chaotic their distributions appeared. Given that all arrangements are irregular/stochastic, a regular appearance did not connote similarity.

From analyzing visual cues used for each of the four patterns, we noticed that participants did not use distance between motifs as a measure of similarity for the CURVIES patterns. Since CURVIES had the largest number of different motif shapes, participants were more inclined to look at densities and motif distribution rather than local distances. This is an important finding since many GTS algorithms use distances between motifs to achieve similar distributions in their results.

| Rationale | PARALLEL | CURVIES | SWAMP | LEAVES | Any |
|---|---|---|---|---|---|
| Density | 10 | 15 | 15 | 17 | 19 |
| Motif ratios | 11 | 11 | 8 | 0 | 16 |
| Patterns | 5 | 5 | 5 | 3 | 12 |
| Orientation | 4 | 7 | 0 | 6 | 11 |
| White space | 4 | 3 | 5 | 5 | 11 |
| Sparsity | 1 | 3 | 3 | 5 | 10 |
| Regularity | 2 | 6 | 2 | 3 | 6 |
| Distances | 3 | 0 | 1 | 2 | 4 |

**Table 2:** *The rationale for similarity sorting and number of participants that used them.*

#### 5.3.2 How hard was it to sort the arrangements? Where was the difficulty? What was difficult?

In general, participants claimed that the study was not difficult and was in fact rather enjoyable. In some instances participants had difficulties sorting certain patterns. We report these below. Note that some participants had difficulties with more than one pattern.

Seven participants stated that the CURVIES arrangements were hard to sort into piles and five thought that SWAMP arrangements were hard to sort. In these cases, participants noted that it was harder to compare arrangements that had more than two motifs. It was easier for participants to judge similarity by comparing densities and motif ratios than by looking at local distances.

Two participants found the LEAVES arrangements hard to sort. One of them believed that having only one motif type in the arrangement made comparing them hard, while the other found it difficult to explicitly match the orientations of the leaves to the exemplar angles. Only one participant in the study mentioned that PARALLEL was difficult to sort, stating that the density was hard to estimate. All the participants who indicated a difficulty spent some extra time sorting the cards but successfully completed the task.

#### 5.3.3 Which pile is the most/least similar to the sample and why?

After choosing the most and least similar piles for each pattern set, participants were asked to provide the reasons for their decisions. Their answers were very concise. In addition to the criteria observed when sorting the piles (Table 2), participants indicated the following as contributors to their similarity decisions: repetition of the source pattern, groups of motifs, broken motifs at the borders, and overlapping motifs.

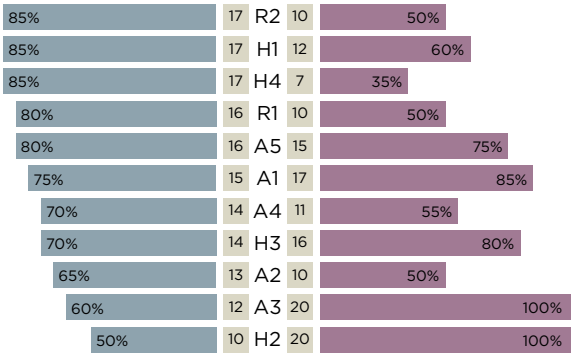


**Figure 7:** *Percentages of participant ratings of synthesis sources as most (grey) or least (purple) similar to the exemplar.*

To visualize which synthesis sources were chosen as more similar most often, we tabulated participant selections as shown in Figure 7. In our analysis, we divide the most and least results into three groups according to the percentage range they fall into (0–50%, 51–75% and 76–100%). The figure illustrates that synthesized arrangements perceived as most similar to their exemplars (as indicated by the bars on the left of the figure) had a correspondingly lower chance of being chosen as least similar (as shown on the right). Synthesis sources that are consistent with this observation are not discussed here.

In the 76–100% range of the most similar list, we find synthesis sources **R2**, **H1**, **H4**, **R1** and **A5**. The interesting observation here lies in the fact that arrangements generated by GTS algorithms are rarely selected as most similar. Despite all efforts made to develop more compelling GTS algorithms, there clearly exist missing pieces to the synthesis problem that need to be addressed. Note that the pseudorandom sources were frequently rated as being most similar to the exemplar. A closer investigation into the relevance of pseudorandom methods for irregular GTS may help us understand what is missing.

Source **A2** was chosen as most or least similar relatively rarely, indicating that participant choices were less consistent for this source. This source was found more similar in some instances for certain patterns, implying that the algorithm was better at reproducing the features of the exemplar in those cases (See the tables in Figure 5). Synthesis sources **A4**, **A5** and **A1** were selected as most similar approximately the same number of times they were selected as least similar. This finding suggests that regardless of the arrangement pattern similarity ratings were consistent, giving us a first hint of how to effectively determine dominance between algorithms.

### 5.4 Summary of findings

Based on the study and subsequent analysis, we have come to a better understanding of the distinctive nature of geometric arrangements and the synthesis sources that made them. The pile-sorting study we adopted led to (1) validating a set of visual cues proposed in earlier perceptual studies [AlMeraj et al. 2011] and (2) a strategy for classifying multiple geometric arrangements based on similarity.

The main observations from our analysis of the pile-sorting data include (1) different synthesis sources correlated with one another highly, (2) pseudorandom synthesis of irregular arrangements effec-

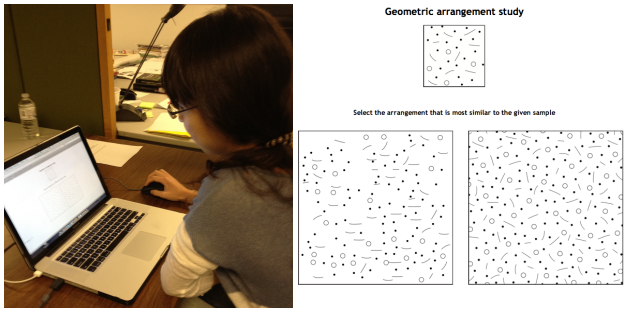


**Figure 8:** *The study setup and comparison interface.*

tively captures characteristics of irregular arrangements and could act as an alternative to GTS algorithms, and (3) none of the GTS algorithms provided us with consistent results for the arrangement patterns but some were more consistent than others.

We recognize that our findings are based upon a limited investigation of a small number of patterns, and do not claim that they are the last word on the relative merits of these synthesis sources. Adding more synthesized arrangements, participants and algorithmic sources to the study may reveal different results. In the same way, adding more arrangements by expert designers can benefit the whole study experience by providing us with a wide set of varying interpretations based on judgements of aesthetics and structure.

## 6 Pairwise comparisons of geometric texture arrangements

In the study described in the previous section we observed participants sort multiple card sets based on their similarities to an exemplar. To determine whether or not these findings are genuinely reproducible, we conducted a second psychophysical experiment. This time we asked participants to chose the most similar arrangement from a randomly presented pair.

Our goal here is to look for patterns in participant choices under brief presentation of the arrangements. We intend to show that these choices are consistent to the ones found in the previous pile-sorting study. Discovering similar patterns will demonstrate that both pile-sorting and pairwise comparison studies are effective for evaluating similarity in GTS results.

### 6.1 Design and setup

*Sample arrangement set*: We study the same synthesized arrangements as in the first study: four sets of patterns each containing 11 synthesized arrangements (Figures 2 and 3). Pairwise combinations of these arrangements result in a total of 440 comparisons, 110 for each pattern.

*Interface and methodology*: Participants were seated on a chair positioned beside a table with a laptop computer. The comparison interface as shown in Figure 8 contains one exemplar input along with two randomly selected geometric arrangements from the same pattern placed at corners of an equilateral triangle on the screen.

The task was described as follows: "Select the arrangement that is most similar to the given sample". Participants made their selection using a mouse. A trial session of twelve random comparisons was required by all participants. They were encouraged to ask questions during the trial before proceeding onto the study.

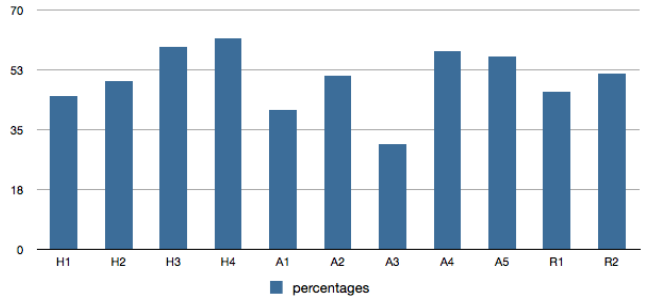


**Figure 9:** *The percentage of most similar ratings of arrangements according to the synthesis sources.*

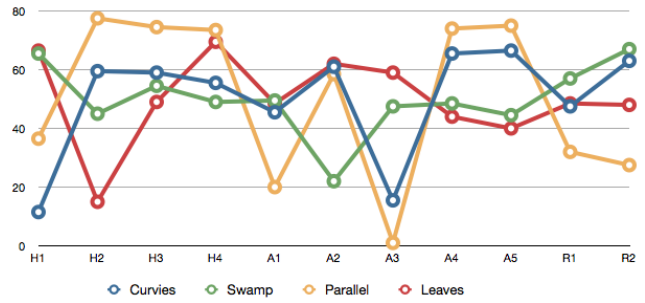


**Figure 10:** *A line chart showing the percentage of most similar ratings according to the arrangement pattern (*CURVIES, SWAMP, LEAVES, *or* PARALLEL*).*

Participants were then presented with 110 comparisons from a randomly chosen pattern. Each arrangement was compared with a result from each of the other sources, and each pair was shown twice, in both left-right orders. The result was that each arrangement appeared 20 times. We found that the left-right order did not effect the results, so we only used one selection from each of the presented pairs in our analysis. After completing this set, the investigator closed the interface, asked if the participant wished to take a break, and then opened a new screen containing the next set of patterns. This was repeated until all four sets were presented. To minimize the chance of participants receiving similar sequences of arrangements, all pattern sets and arrangements within the sets were randomly presented throughout the study.

*Data collection and analysis*: Logs of participants selections, selection times, and arrangements paired were recorded automatically. The average time it took participants to complete this part of the study was 13.5 minutes. To interpret the data we use simple quantitative analysis.

### 6.2 Quantitative analysis of comparisons

In this study we try to identify patterns in participant similarity selection ratings. We first look at the percentages of most similar arrangement ratings according to the generating sources (expert designers, GTS algorithms or pseudorandom) irrespective of the arrangement patterns.

In Figure 9, we find that more participants were inclined to select arrangements from the different sources except **A1** and **A3** as more similar to the exemplar. Note that the least similar choices made by participants for **A1** in the pile-sorting study are more significant when presented though comparisons (see Figure 7). Another observation is that synthesis source **H2** did much better in the com-

parisons than in the pile-sorting study. The remaining sources performed well in this study suggesting that participants were able to compare the differences in arrangement characteristics effectively and judge the similarity quickly.

To gain insight into why sources **A1** and **H2** received low ratings than those in the pile-sorting study, we analyzed the collected rating according to the type of pattern used. Figure 10 presents a breakdown of participant similarity selections. We find that source **A1** did worse for the PARALLEL and **H2** did worse for LEAVES. In Section 5.3 we discussed the different visual cues the participants used to decide similarity; both density and motif ratios are factors in the decisions made here. To understand where the problem areas are for the remaining sources, we analyze them below according to the patterns.

In CURVIES, two of the lowest rated arrangements include sources **H1** and **A3**. This finding is consistent with the previous study and suggests that characteristics captured by these sources are different to those found in other source arrangements. For SWAMP, synthesis source **A2** had the lowest ratings, lower than those found in the previous pile-sorting analysis. The low density exhibited in the arrangements synthesized by this source appears to be more noticeable in pairwise comparisons.

Of all the LEAVES arrangements, as mentioned above, synthesis source **H2** was least likely to be chosen as similar to the source. This result separates **H2** from **H3** and **A5**, though the three were highly correlated in the pile-sorting study. Participants were more likely to select arrangements that had lower densities and avoided overly dense ones as in **H2**.

The PARALLEL arrangements show that sources **H1**, **A1**, **A3**, **R1** and **R2** were more likely to be chosen as least similar than the other synthesis sources. This observation is also consistent with findings in the pile sorting study. The patterns found for the two pseudorandom source arrangements reveal that there is a difference even between two arrangements generated by the same source. This could be a coincidence attributable to the random number generator. Determining any statistical significance here would require generating multiple arrangements, testing them, and averaging the most similar choices.

Synthesis sources **A4** and **A5** were more consistent in their ratings regardless of the pattern. The same sources also had neutral ratings in the pile-sorting study. They achieve average standing in comparison to the other source, not always the best but never the worst.

We did not ask participants to comment on this part of the study. But the results show that participants prefer arrangements that appear to match the exemplar density. For example, we notice that in pile-sorting, **H2**, **H3**, **A4** and **A5** were often described as dense and were chosen as least similar more often than others. However in the pairwise comparisons, **H3**, **H4**, **A4** and **A5** are selected as similar to the exemplars more often indicating that density cues may be overlooked if paired with arrangements that are very different from the exemplar.

From both studies, we conclude that no single source of geometric texture synthesis works the best for all pattern types. This observation is consistent with our pile-sorting finding in Section 5. This is not surprising and hints to the fact that GTS algorithms still need to find better means of capturing the true essence of exemplar inputs even if they start with the comparatively simple case of irregular distributions. The results also suggest the importance of further investigating the visual cues that figure most prominently in human judgments of similarity for geometric textures.

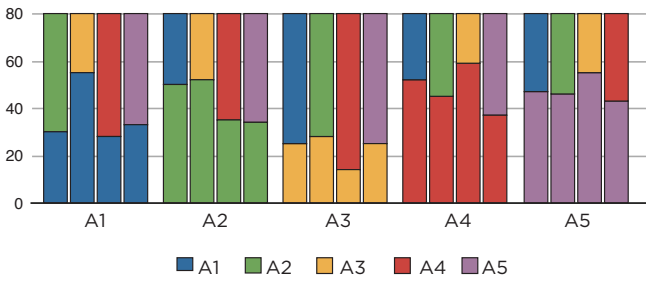


**Figure 11:** *Statistical probabilities of choosing one algorithm against another (only for GTS algorithms).*

## 7 Case report: Evaluating one GTS algorithm

Even though we are in the early stages of understanding how we should effectively evaluate the success of GTS algorithms, evidence from our two studies suggests apparent preferences in terms of algorithm consistency. In this section we attempt a comparison between the patch-based GTS algorithm by AlMeraj [2013] and the other GTS algorithms. Studying one algorithm in light of the others offers the area a first glimpse into the suitability of the evaluation strategies proposed in this work.

*Part 1—Pile-sorting study*: In the analysis reported for this study (Section 5), we found the following: Out of all other GTS algorithms, **A4** and **A5** correlated together most often. A4 also correlated in some cases with A1 and A2, but rarely with A3. LEAVES and SWAMP piles that contained **A4** and **A5** were more likely to be selected as least similar to the exemplar than other sources. In comparison to **A4** and **A5**, **A1** acquired a significantly higher number of least similar ratings for SWAMP. Source **A3** acquired even lower similarity ratings for its PARALLEL and CURVIES arrangements.

*Part 2—Comparison study*: We observed that similarity ratings for synthesized source arrangements **A4** and **A5** were the highest, and so were the ratings for source **A2** (Figure 9). In addition, the ratings for **A4** and **A5** deviated much less than those for other GTS algorithms for the different patterns.

To visualize probabilities of choosing each GTS algorithms regardless of the arrangement patterns, we constructed Figure 11. In it we show, for each algorithm source, the probability of choosing an arrangement from that source over arrangements from any of the other four algorithms.

A chi-squared test (at one degree of freedom, $\alpha = 0.01$) on the pairwise comparisons collected from this study shows a statistically significant bias in favour of **A4** and **A5** when tested against GTS algorithms **A1**, **A2**, **A3**. This means that participants were more likely to select **A4** or **A5** as more similar when presented with an arrangement from another GTS source. When shown a pair of sources from **A4** and **A5**, the decisions participants made were less significant indicating that they were equally likely to select either source as most similar. This explains the consistency noticed between these two sources demonstrated throughout the analyses in this paper.

## 8 Conclusion and future work

Despite having plenty of attractive and visually interesting interpretations of realistic data, NPR has always suffered from a dearth of evaluations to establish the validity of algorithms.

In GTS, a previous attempt to uncover perceptual principles that

cause algorithms to succeed or fail resulted in a concise set of visual cues used by study participants to generate and compare geometric arrangements. Our research takes a broader observational approach and looks at how people compare multiple arrangements generated from different sources (expert designers, GTS algorithms and a pseudorandom routine). This methodology offers us an effective evaluation strategy for gathering and assessing geometric texture arrangements. It also draws attention to relevant quantitative measures that can be explored in the future.

From two carefully coordinated psychophysical experiments described in this paper, we look closer at perceptual significances and similarity preferences. Limited to examples from our newly devised benchmark, our findings indicate strong similarity trends for certain synthesis sources. In order to further validate these trends, it would be valuable to enhance the GTS dataset with more examples and rerun a series of pile-sorting studies. We also believe that including the larger original USGS textures in future studies may shed light on whether human judgements are purely based on aesthetics, or tied to specific details of individual spatial distributions.

Establishing any form of evaluation for this area is evidently hindered by the lack of a formal definition of *similarity*. Note that the choice of words presented in the user study tasks is subjective in nature and may have been interpreted differently by the expert designers. We hope that the visual criteria mentioned throughout this paper will motivate more detailed investigations into their influence on GTS similarity judgements. Being able to describe similarity through a validated set of perceptual processes in the form of a formal definition is essential for the advancement of the GTS field.

Pile-sorting [Isenberg et al. 2006] and pairwise comparisons [Lin et al. 2006] have perviously been adopted as methodologies and have subsequently provided us with a stable experimental paradigm. An interesting next step would be to adopt similar strategies for other area in NPR first as exploration tools and inevitably as evaluation methods.

Most current GTS algorithms are heuristic in nature, and if tweaked, even slightly, could produce different arrangements. This will continue to be a major limiting factor when evaluating GTS algorithms unless standards are proposed. In this work we suggested standardizing the vector input style of the algorithms but further investigation into its practicality is required.

Our results also give rise to the significance of density as a visual factor when judging similarity. Upon testing the densities of LEAVES from the pile-sorting study we noticed that clusters conform to arrangements of relatively similar density. An interesting next step would be to verify this for the remaining patterns. Detecting significances will facilitate the development of a compact similarity definition for GTS.

Our experiments have shown that no GTS algorithm performs well for all the patterns we adopted. Future efforts should focus on developing a set of criteria to help researchers and designers decide which algorithm is best suited for their applications. Narrowing down to a succinct set of criteria would depend on collecting more arrangements and using effective study methodologies.

## 9 Acknowledgements

We would like to thank the expert designers and GTS algorithm authors for their contributions to this work. The benchmark of examples and studies would not have been possible without their sincerity and time. We would also like to thank Ed Lank for his guidance during the study design. Finally, we would like to thank the anonymous reviewers for their feedback which helped improve the final version of this paper. This research is supported by a doctoral scholarship from Kuwait University and Adobe.

## References

ALMERAJ, Z., KAPLAN, C. S., ASENTE, P., AND LANK, E. 2011. Towards ground truth in geometric textures. In *NPAR*, ACM, New York, NY, USA, 17–26.

ALMERAJ, Z. F., KAPLAN, C. S., AND ASENTE, P. 2013. Patch-based geometric texture synthesis. In *Expressive – CAe*, ACM.

ALVES DOS PASSOS, V., WALTER, M., AND SOUSA, M. C. 2010. Sample-based synthesis of illustrative patterns. In *Pacific Graphics '10*, IEEE Computer Society, Washington, DC, USA, 109–116.

BALAS, B. 2008. Attentive texture similarity as a categorization task: Comparing texture synthesis models. *Pattern Recognition 41*, 3 (Mar.), 972–982.

BARLA, P., BRESLAV, S., THOLLOT, J., SILLION, F. X., AND MARKOSIAN, L. 2006. Stroke pattern analysis and synthesis. *Computer Graphics Forum 25*, 3, 663–671.

HURTUT, T., LANDES, P.-E., THOLLOT, J., GOUSSEAU, Y., DROUILLHET, R., AND COEURJOLLY, J.-F. 2009. Appearance-guided synthesis of element arrangements by example. In *NPAR*, ACM, New York, NY, USA, 51–60.

IJIRI, T., MĚCH, R., IGARASHI, T., AND MILLER, G. 2008. An example-based procedural system for element arrangement. *Computer Graphics Forum. 27*, 2, 429–436.

ISENBERG, T., NEUMANN, P., CARPENDALE, S., SOUSA, M. C., AND JORGE, J. A. 2006. Non-Photorealistic Rendering in Context: An Observational Study. In *NPAR*, ACM, New York, NY, USA, 115–126.

JENNY, B., HUTZLER, E., AND HURNI, L. 2010. Point pattern synthesis. *The Cartographic Journal 47*, 3, 257–261.

LIN, W.-C., HAYS, J., WU, C., LIU, Y., AND KWATRA, V. 2006. Quantitative evaluation of near regular texture synthesis algorithms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, IEEE Computer Society, Washington, DC, USA, CVPR '06, 427–434.

MA, C., WEI, L.-Y., AND TONG, X. 2011. Discrete element textures. *ACM Trans. Graph. 30*, 4 (Aug.), 62:1–62:10.

ÖZTIRELI, A. C., AND GROSS, M. 2012. Analysis and synthesis of point distributions based on pair correlation. *ACM Transaction on Graphics 31*, 6 (Nov.), 170:1–170:10.

UNITED STATES FEDERAL GEOGRAPHIC DATA COMMITTEE, GEOLOGICAL DATA SUBCOMMITTEE. 2006. *FGDC Digital Cartographic Standard for Geologic Map Symbolization*. United States Geological Survey. Viewable online at `http://pubs.usgs.gov/tm/2006/11A02/FGDCgeostdTM11A2_web_all.pdf`.

WEI, L.-Y., LEFEBVRE, S., KWATRA, V., AND TURK, G. 2009. State of the art in example-based texture synthesis. In *Eurographics, State of the Art Report, EG-STAR*, Eurographics Association.

WELLER, S. C., AND ROMNEY, K. 1988. *Systematic data collection*. Sage Publications.