An Appearance Uniformity Metric for 3D Printing

Michael Ludwig
University of Minnesota
Minneapolis, Minnesota
mludwig@cs.umn.edu

Gary Meyer
University of Minnesota
Minneapolis, Minnesota
meyer@cs.umn.edu

Ingeborg Tastl
HP Inc. Laboratories
Palo Alto, California
ingeborg.tastl@hp.com

Nathan Moroney
HP Inc. Laboratories
Palo Alto, California
nathan.moroney@hp.com

Melanie Gottwals
HP Inc. Laboratories
Palo Alto, California
melanie.gottwals@hp.com

ABSTRACT

A method is presented for perceptually characterizing appearance non-uniformities that result from 3D printing. In contrast to physical measurements, the model is designed to take into account the human visual system and variations in observer conditions such as lighting, point of view, and shape. Additionally, it is capable of handling spatial reflectance variations over a material's surface. Motivated by Schrödinger's line element approach to studying color differences, an image-based psychophysical experiment that explores paths between materials in appearance space is conducted. The line element concept is extended from color to spatially-varying appearances–including color, roughness and gloss-which enables the measurement of fine differences between appearances along a path. We define two path functions, one interpolating reflectance parameters and the other interpolating the final imagery. An image-based uniformity model is developed, applying a trained neural network to color differences calculated from rendered images of the printed non-uniformities. The final model is shown to perform better than commonly used image comparison algorithms, including spatial pattern classes that were not used in training.

CCS CONCEPTS

• **Computing methodologies** → **Perception**; *Appearance and texture representations*;

KEYWORDS

Appearance uniformity, spatially-varying appearance perception, 3D printing, neural networks

ACM Reference Format:
Michael Ludwig, Gary Meyer, Ingeborg Tastl, Nathan Moroney, and Melanie Gottwals. 2018. An Appearance Uniformity Metric for 3D Printing. In *SAP '18: ACM Symposium on Applied Perception 2018, August 10–11, 2018, Vancouver, BC, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3225153.3225169

SAP '18, August 10–11, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5894-1/18/08…$15.00
https://doi.org/10.1145/3225153.3225169

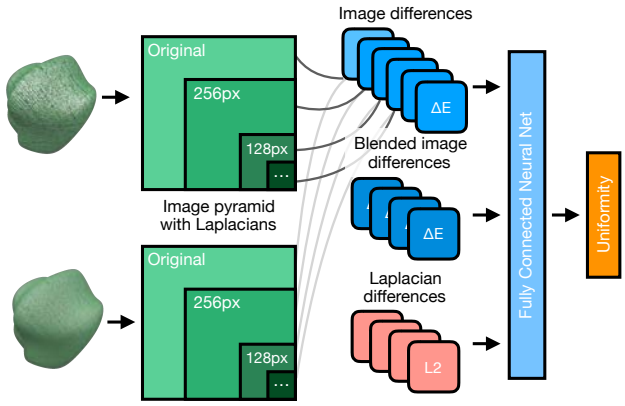

Figure 1: Visual representation of our neural network model that estimates perceived uniformity between two input images. The two images must be of similar appearances, where one represents the uniformity reference point.

1 INTRODUCTION

Appearance uniformity is a perceptual measure of a material's apparent spatial homogeneity. It can depend on how the material is illuminated, and from where and at what distance it is viewed. A material's uniformity was cast as an appearance matching problem by measuring how well it matched an ideal, smoothed target [Ludwig et al. 2018]. In this way, studying perceived uniformity is an important step towards developing a general, spatially-varying appearance metric. It moves past homogeneous studies of color [Wyszecki and Stiles 1982] and gloss [Ferwerda et al. 2001; Hunter and Harold 1987; Wills et al. 2009] to incorporate spatial variations of color, gloss, and texture.

Uniformity is of particular interest to us because of the role it plays in developing high quality 3D printing technologies and processes. Outputting a smooth surface with homogeneous color requires the device to have high precision and consistency. The printing of colored patterns or surface texture can in fact mask issues concerning the quality of the output. The input appearances shown in Figure 1 highlight two physically-correct rendered appearance images that differ in their spatial uniformity.

While most 2D consumer color paper printers are able to achieve satisfactory levels of uniformity across a page, a 3D printer's control over both macroscopic surface curves and mesoscale bumps and texture present new challenges in measuring and quantifying effects on appearance uniformity. Current 3D printers have not yet reached a comparable quality level to their 2D counterparts, and they frequently introduce unintentional non-uniformities. This work seeks to characterize the perceptual dissimilarities in color, surface, and other appearance non-uniformities that can occur in 3D printing. The resulting model, whose architecture is shown in Figure 1, acts as a metric that correctly handles the effects of spatially-varying color, texture, and gloss on perceived uniformity. It can be used to provide consistent judgments of print quality and of how well a digital design matches the resulting printed output.

When developing any approximation to human perception, it is necessary to validate it against actual human data. Utilizing the 3D printed materials created in [Ludwig et al. 2018], we present a psychophysical experiment that explores paths between the materials in appearance space. The design and motivation for the study comes from Schrödinger's line-element approach to studying color differences [Schrödinger 1920]. Perceptual line elements are vectors through color space that encode the relative compactness and similarity of the region around a specific color, and they represent a just-noticeable change from the reference. Section 3 presents the details of the experiment, and Section 4 summarizes the key results.

Following that, we present our uniformity model in Section 5 and evaluate it in comparison to existing image-based models. By relying solely on the imagery of the 3D printed surfaces, this model can easily be extended to operate on appearances with unknown material descriptions. We show that our model performs better than past image-based algorithms, particularly when the uniform target differs by more than just diffuse color. Our approach is also more consistent at matching human responses under different lighting conditions compared to prior work. We conclude with a discussion of the limitations of this work and the broader challenges of developing a complete perceptual appearance metric.

2 BACKGROUND

This work lies at the intersection of a number of fields. While it is motivated by additive manufacturing and 3D printing, the focus on appearance requires color science, the study of material modeling and simulation, and psychophysics. We briefly review these topics below.

2.1 3D Printing

A range of additive manufacturing techniques have been developed over the last few decades, such as fused deposition modeling (*FDM*), selective laser sintering (*SLS*), and most recently multi-jet fusion (*MJF*). Additional approaches have been developed that work with metals and ceramics, broadening the range of materials usable with 3D printing. While 3D printing has revolutionized prototyping, it has also become prominent for final parts production. In these domains, the appearance of the 3D printed object becomes paramount. Visual quality can also improve impressions of mechanical quality.

Researchers have been advancing the appearance capabilities of 3D printers. A relatively simple extension to FDM printers is to make it two-toned [Reiner et al. 2014]. Full color 3D printing further improves upon the realism of the printed object [Brunton et al. 2015]. Combining knowledge of human perception with additive manufacturing has led to noticeable improvements in visual quality. [Wang et al. 2015] uses a perceptual saliency model to control the resolution of the printer to minimize the visible layering in complex areas, while still minimizing the overall printing time.

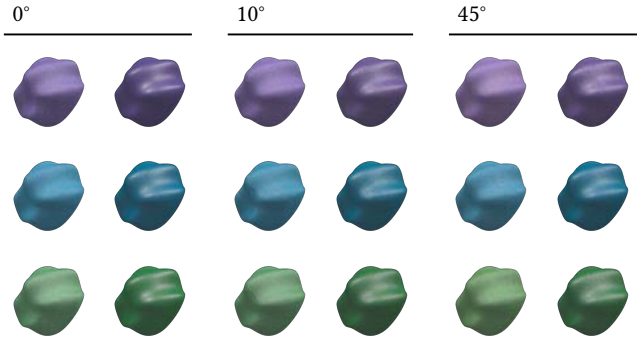

Figure 2: Renderings of the originally 3D printed tiles from [Ludwig et al. 2018]. Within each angle group, the left column represents the raw printed appearance, and the right column represents the appearance after post-processing the material. Note the qualitative similarities in spatial patterns within each column and group.

2.2 Measured and Perceived Appearance

Eventually printers will have the capability to print materials that exhibit a broad range of reflectance profiles. The future of 3D printing represents a way to print and manufacture full spatially-varying materials on arbitrary curved surfaces. Reaching that point requires a thorough understanding of both how to physically measure what is printed and how people perceive the result.

Materials can be scanned with a myriad of techniques that combine multiple calibrated lights and cameras to solve for spatially-varying bidirectional reflectance distribution function (*SVBRDF*) parameters. Examples of this include spatial gonioreflectometry [Mcallister 2002] and linear light scanning [Gardner et al. 2003]. An SVBRDF produced by such scanning techniques, or by a device like X-Rite's TAC7[1], provides two dimensional maps of parameters such as diffuse color, specular color, roughness, and surface normal that can be used to make physically-based renderings. SVBRDF scanners provide substantially more information compared to commonly used devices in the industry–such as colorimeters, spectrophotometers, and glossmeters–that sample a point or average a small area.

Appearance and its perception involves much more than physical measurements. It combines an object's color, glossiness [Ferwerda et al. 2001; Wills et al. 2009], shape [Havran et al. 2016; Vangorp et al. 2007], the lighting environment [Fleming et al. 2003], and more. These dimensions interact in complex manners, such as gloss and color [Dalal and Natale-Hoffman 1999], and gloss and texture [Marlow et al. 2012]. A constraint on these works has been

[1] https://www.xrite.com/categories/appearance/tac7

the assumption of a homogeneous reflectance model over the entire stimuli. A necessary step forward for appearance research is to account for spatial variations in all reflectance parameters; we consider the study of uniformity to be an effective means to pursue this challenge.

2.3 Appearance Uniformity

Uniformity is a spatially-varying appearance attribute that characterizes the perceived homogeneity of a surface. A perceived non-uniformity can be due to spatial variations in surface color, roughness, texture, or any other material property. It is also dependent on the lighting and viewing conditions, as they can mask or highlight differences over a surface. Actual 3D printed materials are produced with various imperfections and spatial distortions in surface height, color, and roughness. Quantifying the perceived uniformity in the presence of these variations can be used to guide improvements to the 3D printing process.

Perceptual uniformity can be quantified as the dissimilarity between the appearance under question and a homogeneous or near-homogeneous ideal material [Ludwig et al. 2018]. Specifically, for an appearance, A, and its ideal, $\mathrm{I}(A)$, the non-uniformity with respect to that ideal is given by the perceptual distance function $\mathcal{U}(A, \mathrm{I}(A)) \rightarrow \mathbb{R}$. While $\mathcal{U}$ could be described by a general appearance metric, the constraint that $\mathrm{I}(A)$ be similar yet more spatially uniform than A makes modeling $\mathcal{U}$ substantially easier in this context.

We use the appearance images and uniformity judgments from [Ludwig et al. 2018], so we describe the dataset in more detail here. A total of 288 appearance images, consisting of 144 different SVBRDFs of 3D printed material rendered with two lighting conditions, were prepared from a set of 9 actual printed 3D printed tiles. The source tiles were printed in three base colors–purple, cyan, and green–and oriented at three different angles with respect to the z axis of the printer. The 9 materials were scanned using an X-Rite TAC7 device, then put through a post-processing procedure to reduce surface roughness and increase glossiness, and then re-scanned to create a total of 18 high-quality SVBRDFs. Renderings of these initial appearances are shown in Figure 2. Each of the initial 18 SVBRDFs were processed to generate an ideal material definition with minimal color or surface abnormalities, as well as 7 variants with increased non-uniformities in bumpiness, glossiness, diffuse albedo, or combinations thereof.

The virtual generation process resulted in sets of 8 appearances with consistent spatial patterns produced by the original printing process. Within each set, the appearances shared an ideal appearance description. On their own, 3D printers are currently incapable of producing an ideal appearance and do not have the precision to exactly reproduce the spatial patterns seen in each of the images in Figure 2. However, this virtual representation can conveniently produce novel appearance images to be used in psychometric studies. Appearance uniformity was measured for the original 288 appearances by using the visual search paradigm [Arun 2012]. In the next section we present another approach designed to work with appearance differences closer to the visual threshold.

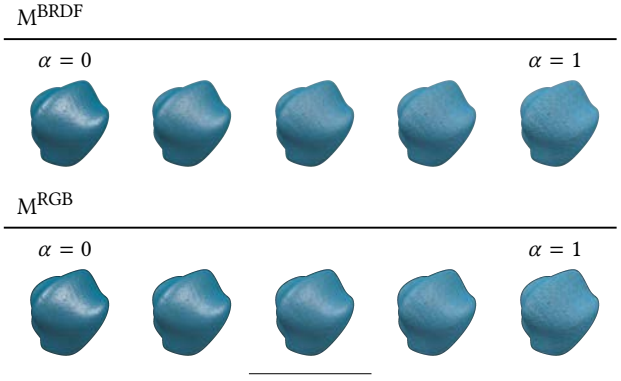

Figure 3: An appearance curve for $\mathrm{M}^{\mathbf{BRDF}}$ (top row) and $\mathrm{M}^{\mathbf{RGB}}$ (bottom row). The most significant difference between the two paths is the shape of the specular highlight, which is most visible in the underlined column.

3 MEASURING UNIFORMITY NEAR THRESHOLD

The range of appearances produced by these material variations was inspired by the discrimination ellipse model for a perceptual color space [MacAdam 1942]. By exploring colors very similar to reference points, the model identified perceptual scaling and principle dimensions. Discrimination ellipses served to answer the question of when colors looked the same, which has been of great importance in 2D printing and quality control. The appearances shown in Figure 2 sample the appearance space around a number of references to similarly learn how that space behaves, and they address the question of when a 3D printed surface matches its intended appearance.

We continue the approximation that the appearances, A and $\mathrm{I}(A)$, can be represented as color images and evaluated on a monitor. Although not as optimal as seeing the real object, the use of images for appearance comparisons has been employed to measure BRDF accuracy [Pereira and Rusinkiewicz 2012], translucency [Gkioulekas et al. 2013], and gloss [Wills et al. 2009]. Additionally, in the context of 3D printing, the constraints imposed by an image are similar to industry standards that specify viewing orientation and lighting, but images are still able to convey spatial variations in appearance.

In order to properly model how perceived uniformity changes near the ideal, we present a new study inspired by [Schrödinger 1920]'s line element approach to measuring perceived color differences. We use this experiment to augment the uniformity measurements from the existing 3D printing appearance dataset.

3.1 Line Elements and Appearance

Given two appearances, A and $\mathrm{I}(A)$, a curve between the two can be represented as $\mathrm{M}_A(\alpha)$ such that $\mathrm{M}_A(0) = A$ and $\mathrm{M}_A(1) = \mathrm{I}(A)$. A curve like this can be broken into perceptual line elements such that the perceived distance between the end points of each line element corresponds to a single just noticeable difference. The perceived length of the path is the number of line elements it contains, and the perceived distance between the endpoints is the minimum path length.

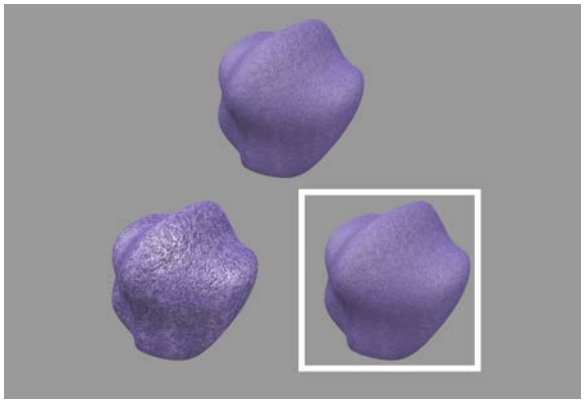

Figure 4: A cropped screenshot of the user interface for the forced-choice threshold measuring task. The top appearance is the target, subjects must select the left or right image in the bottom row that most closely matches the target. The subject's selection is shown with the white border.

We propose an extension of this concept to make it continuous. First, we introduce the just noticeable difference function for a path as $\Delta_{\mathrm{M}_A}(\alpha) : [0,1] \rightarrow [0,1]$. The Δ function returns the positive delta to the nearest point on the curve after α that is just noticeably different from the appearance of $\mathrm{M}_A(\alpha)$. Second, instead of counting discrete elements, we define the path length in terms of a line integral, so $|| \mathrm{M}_A || = \int_0^1 \left\| \frac{\mathrm{d}\,\mathrm{M}_{ij}}{\mathrm{d}\alpha} \right\| \mathrm{d}\alpha$. In a perceptual space, the norm of the derivative of M_A can be approximated using the Δ function: $\left\| \frac{\mathrm{d}\,\mathrm{M}_A}{\mathrm{d}\alpha} \right\| \approx \frac{1}{\Delta_{\mathrm{M}_A}(\alpha)}$. Thus, the distance between A and $\mathrm{I}(A)$ is the minimum path length,

$$\mathcal{U}(A, I(A)) = \min_{\mathrm{M}_{ij}} \int_0^1 \frac{1}{\Delta_{\mathrm{M}_A}(\alpha)} \,\mathrm{d}\alpha \tag{1}$$

By integrating over arbitrary sub-intervals $[a,b] \subset [0,1]$, the distance between $\mathrm{M}_A(a)$ and $\mathrm{M}_A(b)$ can be computed in terms of the original path. With the linear mapping, $(b-a)\alpha + a$, and the substitution rule, it can be shown that

$$\mathcal{U}(\mathrm{M}_A(a), \mathrm{M}_A(b)) = \int_a^b \frac{1}{\Delta_{\mathrm{M}_A}(\alpha)} \,\mathrm{d}\alpha \tag{2}$$

Equations 1 and 2 form the basis of the psychometric experiment presented in this work. By sampling the Δ function at specific intervals along a curve between a 3D printed appearance and its ideal, it can be approximated and used to estimate $\mathcal{U}$ for appearances on the curve arbitrarily close to the ideal.

We defined two different curve functions to explore how the specific function for M impacts the final calculated $\mathcal{U}$ for an appearance pair. $\mathrm{M}^{\mathrm{BRDF}}$ interpolates all SVBRDF parameters between the surfaces using perceptually linear scales for diffuse and specular colors [Sharma et al. 2004] and gloss [Ferwerda et al. 2001]. The output image is the rendering of the interpolated SVBRDF parameters. $\mathrm{M}^{\mathrm{RGB}}$ forms output appearance images by interpolating the per-pixel colors of the two input appearance images. An example comparison between the two curves' behaviors is shown in Figure 3.

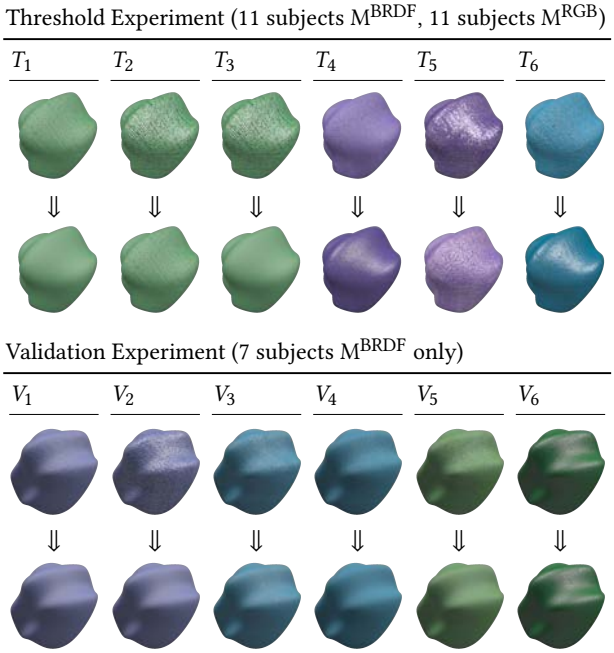

Figure 5: The top block displays the six appearance pairs whose Δ functions were in the primary experiment. The image above each arrow specifies $\alpha = 0$ and the image below, $\alpha = 1$. The bottom block displays the six appearance pairs that were evaluated in a follow-up experiment. The bottom appearances have novel spatial patterns because the tiles were printed at 60° and 90° with respect to the z axis.

3.2 Psychometric Experiment

The purpose of the experiment was to measure the Δ function for the six appearance pairs at the top of Figure 5. The top row represents the non-uniform appearance and the bottom represents the target ideal appearance. The appearances T_1 and T_3 correspond to dissimilarity comparisons originally evaluated in [Ludwig et al. 2018] and are included to check consistency between the two measurement procedures. The ideal materials of those two pairs were formed virtually by smoothing the diffuse albedo and height maps of the original, scanned SVBRDFs. For appearances T_4 through T_6, the target appearance is intentionally less ideal, such as when a material's ideal reflects what a printing device is capable of producing, even if that is not yet perfect.

Each subject was assigned either $\mathrm{M}^{\mathrm{BRDF}}$ or $\mathrm{M}^{\mathrm{RGB}}$, but evaluated all six appearance pairs using the assigned curve function. To measure each curves' Δ function, the paths were sampled at $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8\}$ to form a set of 30 reference appearances. A forced-choice paired-comparison task was then developed to identify the just noticeable difference threshold for each reference. The thresholds define the Δ function at the reference samples points of the curve; once all thresholds were measured, the full Δ function can be estimated by fitting a smooth function to the data.

Figure 4 presents the user interface for the forced-choice comparison task. The top image represents the reference, and the left and

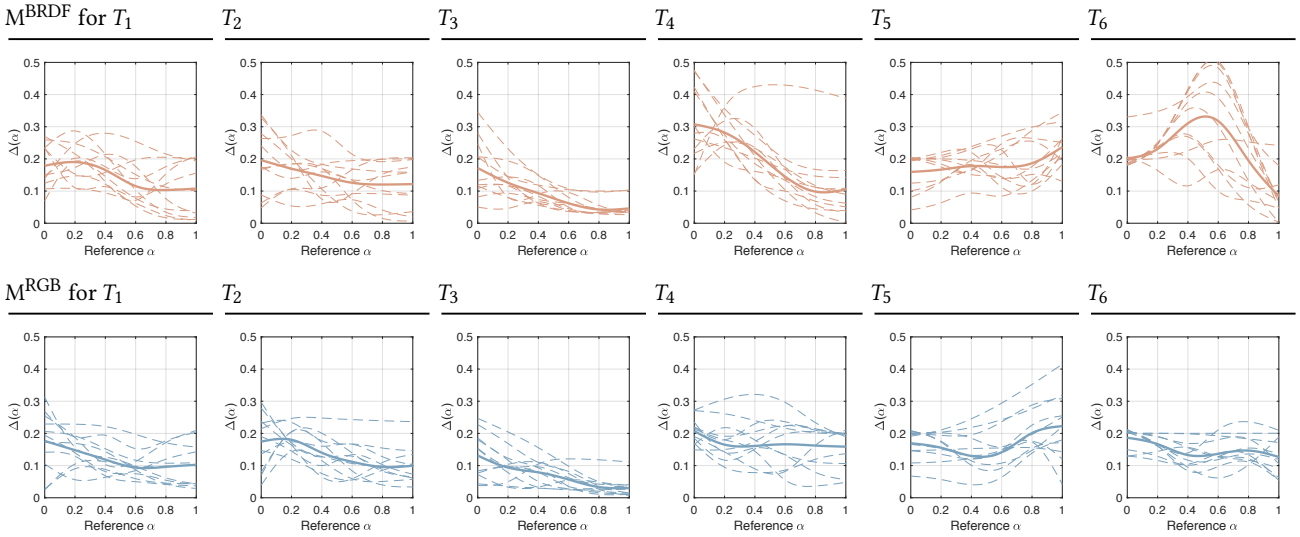

Figure 6: Plots of the measured Δ functions for the two appearance paths for each T_i. Each dashed line represents a smoothed curve fit to the subjects' thresholds at each reference. The solid line is $E(\Delta)$, equal to the mean of the dashed lines.

right images were randomly assigned a copy of the reference and the current test appearance image. Subjects had up to 3 seconds to make a choice between the left and right images, specified using the keyboard. A Bayesian stair-stepping process based on [Kontsevich and Tyler 1999] was used to determine the test appearances for each reference. Depending on a subject's response correctness, the next test image would be more or less similar to the reference. When the test image is within visual detection threshold of the reference, the probability of picking the correct side approaches $p = 0.5$. Beyond the threshold, p tends towards 1.

The adaptive procedure would continue iterating with new test positions along the path until the threshold estimate converged. Stochastic simulation of this procedure indicated that the threshold for a specific reference was reliably estimated within 20 trials. This was used as an upper bound for estimation to help cap the time length of the experiment to 30 minutes per subject. To avoid ordering effects, trials for all 30 reference images were randomly intermixed.

Experimental Setup. A total of 22 subjects participated. 11 subjects were assigned to M^{RGB} and the remaining 11 to M^{BRDF}. All subjects had normal or corrected to normal visual acuity and were tested to be color normal before participating. Subject ages ranged between 18 and 55, with 12 subjects being male and 10 female.

The experiments were conducted in a dark room on a 24 inch Dell monitor with a resolution of 1920×1080. The monitor was confirmed calibrated to the sRGB industry standard. The three images shown in each trial were downsampled to 300×300 and each occupied a visual angle of approximately 6.4°.

Section 4 analyzes the measurements of the 12 Δ functions sampled by this experiment.

Validation Study. The 3D printed appearances used in the above study and in the original visual search task from [Ludwig et al. 2018] were limited to using SVBRDF measurements from tiles printed at 0°, 10°, or 45° degrees. The print angle of the tiles produced qualitatively similar spatial patterns within each group. We produced a small set of new appearance images from tiles printed at 60° and 90° degrees to act as a validation test for our developed model. These materials are visually similar to the original appearances but feature new spatial patterns produced by the different print bed orientation.

The six appearance pairs in the bottom block of Figure 5 were used in a second iteration of the study described above, in place of the original six. A total of 7 subjects participated in this follow-up, three of whom participated in the original. All subjects were presented with the M^{BRDF} curve function. Subject ages ranged from 18 to 30, with 4 being male and 3 female. The stimuli were presented in the same dark room on the same monitor as before. The distances derived from these measurements are not used in the development of the model presented in Section 5, but are used to evaluate its performance on new spatial patterns in an appearance image.

4 RESULTS OF HUMAN SUBJECTS STUDY

The measured Δ functions of the T_i appearances using the M^{BRDF} and M^{RGB} functions are shown in Figure 6. Each dashed line is a smoothed spline fit to the just noticeable difference thresholds measured at $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8\}$. The solid line in each plot is the average of those subject Δ functions. It is denoted $E(\Delta)$ and summarizes the expected just noticeable difference behavior of the sampled subjects.

The two curve functions produce unique paths through appearance space, with particularly noticeable changes for T_4, T_5 and T_6. These three appearance pairs feature the most significant changes to glossiness across the images, so it is not surprising that image interpolation behaves differently from BRDF interpolation in that circumstance. When the appearance transition is dominated by diffuse albedo (pairs T_1, T_2, and T_3), the two modes are similar.

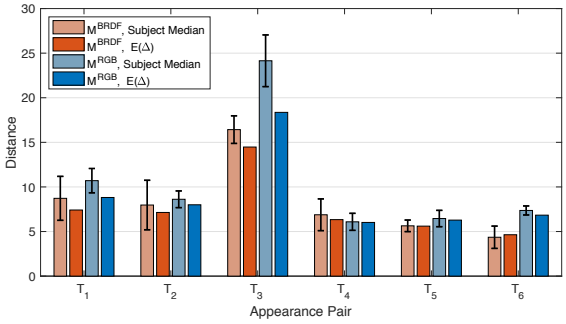

Figure 7: Summary of path lengths calculated from the Δ functions shown in Figure 6. This shows both the median distance for the subjects (with error bars showing standard error), and the distance estimate from the $E(\Delta)$ functions (solid line in Figure 6).

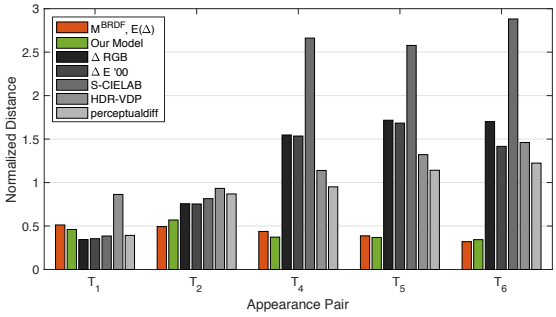

Figure 8: Normalized appearance distances on the $\{\mathcal{U}(T_i)\}$ dataset. The measured distance, estimated by M^{BRDF}'s $E(\Delta)$, is shown in red. Our model's predicted distances are in green and existing metrics are in shades of gray. These are all single estimates, so no error bars are shown.

The qualitative spatial uniformity of the ideal target has a noticeable impact on the shape of the Δ function measured using M^{BRDF}. The smoother the appearance is, the more sensitive humans are to any deviation from that appearance, which is shown by much smaller delta values near $\alpha = 1$. T_4 and T_6 show that this trend continues even when the ideal target features a glossy highlight. T_5 utilized an "ideal" that was itself fairly non-uniform, more so than the others. Not surprisingly, this produces a flatter Δ function.

Figure 7 presents the results of Equation 1 applied to M^{BRDF} and M^{RGB} for each T_i. The low saturation bars show the median path length, calculated with each subject's Δ function. Error bars on those bars represent the standard error of the subject distribution for each appearance pair. The saturated bars show the path length estimate calculated from the expected $E(\Delta)$ function of each appearance pair.

M^{BRDF} and M^{RGB} produce consistent non-uniformity measurements for each T_i, even though they were evaluated with distinct subject sets. The mean Δ functions are also consistent estimates of path length with respect to each subjects' measurement; the one exception being M^{RGB} for T_3, where the mean Δ function substantially underestimates non-uniformity compared to the subject distribution. The correlation coefficient between the path lengths for M^{RGB} and M^{BRDF} is 0.9691 making the image-interpolation path a remarkably accurate estimator even though it had no knowledge of SVBRDF parameters.

5 MODELING APPEARANCE UNIFORMITY

Unlike the work of [Ferwerda et al. 2001; Gkioulekas et al. 2013; Wills et al. 2009], which sought to find relationships between specific physical appearance parameters and perceived distance along an appearance dimension, an image-based metric can operate on images from any source, including photographs. In situations where the SVBRDF or other physical model of the material may be unknown or unavailable, an image-based metric is more powerful and flexible. We evaluated past image difference predictors by how well they match the measured appearance non-uniformities. We develop a neural network-based predictor that performs better.

Before continuing, we provide a brief description of the collections of human uniformity measurements used in our evaluation. We keep the datasets separate because they differ in terms of measurement methodology, lighting conditions for the simulations, and spatial characteristics caused by the printing configuration. To account for differences in scale across the measurement methods and image metrics, all reported distance estimates have been scaled such that each metrics' estimate of $\mathcal{U}(T_3)$ is equal to 1.

- $\{\mathcal{U}(D65)\}$ - The original 144 materials measured by visual search from [Ludwig et al. 2018], lit using a simulated D65 light booth.
- $\{\mathcal{U}(spot)\}$ - The same 144 material pairings as $\{\mathcal{U}(D65)\}$ but lit with a white spotlight to represent a less general lighting condition.
- $\{\mathcal{U}(T_i)\}$ - The six pairs at the top of Figure 5 simulated under a D65 light booth, measured in the main experiment.
- $\{\mathcal{U}(V_i)\}$ - The bottom six pairs of Figure 5, measured in the follow-up validation experiment.

5.1 Existing Image Metrics

The five existing image difference metrics we evaluated are summarized below. These models have a varying degree of complexity in how they handle color differences and how they operate spatially over image regions:

- ΔRGB - The mean per-pixel L_2 norm of the cubed roots of the RGB values [Gkioulekas et al. 2013].
- ΔE *'00* - A modification of ΔRGB that replaces the color difference metric with the more perceptually uniform ΔE '00 [Sharma et al. 2004].
- *S-CIELAB* - A spatially-aware filter for ΔE [Zhang and Wandell 1997].
- *perceptualdiff* - A metric that models the human visual system [Yee et al. 2001].
- *HDR-VDP 2.2.1* - A newer visual difference predictor in the same vein as *perceptualdiff*. [Mantiuk et al. 2011].

Algorithm 1 Calculate a feature vector for a pair of images.

Require: A and B are square images of the same size.

$dim \leftarrow [256, 128, 64, 32, 16]$

Generate standard image pyramids, where C refers to A or B:

$C_i \leftarrow \text{ScaleTo}(C, dim(i))$ ▷ $\forall i \in [1:5]$

$L(C_i) \leftarrow \text{Laplacian}(C_i, C_{i+1})$ ▷ $\forall i \in [4:1]$

Generate blended image pyramids, where C refers to A or B:

$blend_5 \leftarrow (A_5 + B_5)/2$

$blend_i \leftarrow (H(A_i) + H(B_i))/2$ ▷ $\forall i \in [4:1]$

$H(C_i) \leftarrow L(C_i) + \text{ScaleTo}(blend_{i+1}, dim(i))$

Compute mean per-pixel differences for feature vector:

return $[\Delta E(A, B), \Delta E(A_1, B_1), \ldots, \Delta E(A_5, B_5),$
$\Delta E(H(A_1), H(B_1)), \ldots, \Delta E(H(A_4), H(B_4)),$
$||L(A_1) - L(B_1)||_2, \ldots, ||L(A_4) - L(B_4)||_2]$

Figure 8 highlights the poor behavior of existing models on the six appearances in $\{\mathcal{U}(T_i)\}$. Appearance pairs T_4, T_5, and T_6 have noticeable shifts in diffuse albedo, although in these cases it is a uniform shift across the entire surface. They also feature large regions displaying white specular highlights that are mismatched between each pairs' A and $I(A)$ images. As is to be expected, such a low frequency change does not impact human judgment of uniformity since that change is along a different dimension of appearance. The failings of existing approaches to accurately model human behavior necessitated the development of a new appearance metric targeting the 3D printed appearances that were studied.

5.2 Neural Network Uniformity Metric

Given that prior image difference predictors lend too much importance to low frequency changes in appearance, we developed an approach that built upon existing color difference work but provided more flexibility in dealing with differences at various frequency levels. The high-level diagram of the algorithm was shown earlier in Figure 1. The mean per-pixel ΔE '00 color difference is calculated for a hierarchy of resolutions, as well as the mean L_2 differences between the derived Laplacian pyramids [Burt and Adelson 1983]. This vector of color differences encodes how the appearances differ over a range of frequencies and forms the input feature set for a simple neural network. The neural network was then trained to match human behavior.

Algorithm 1 lists the procedure that calculates the neural network features for a pair of input appearance images. While it can hypothetically operate on any two appearances, we trained it under the expectation that one image is a relatively similar appearance representing the ideal target. Initially, it builds partial image pyramids, $[A_i]$ and $[B_i]$, for the appearance A and its uniform target B, using square images at 256, 128, 64, 32, and 16 pixels per side. [2] Laplacian difference pyramids, $[L(A_i)]$ and $[L(B_i)]$, are calculated from the pyramids for each input image. The Laplacian images capture the specific patterns at each resolution.

Lastly, two blended image pyramids, $[H(A_i)]$ and $[H(B_i)]$, are reconstructed from the two Laplacian pyramids by averaging the lower resolution images between the two pyramids, and adding the individual higher resolution images. Each pair of images in the blended pyramid contains the resolution-specific features of each appearance anchored to the average appearance up to that resolution. The mean per-pixel ΔE '00 color difference is calculated between A and B, each A_i and B_i, and each $H(A_i)$ and $H(B_i)$ image pair. The mean per-pixel L_2 difference is calculated between each $L(A_i)$ and (B_i) image pair.

The 14-element feature vector calculated by Algorithm 1 is then input into a fully-connected neural network, implemented using the MATLAB Neural Network Toolbox[Mathworks Inc. 2018]. Different subsets of the full feature vector were evaluated, but they did not achieve the same accuracy or consistency. While a variety of network structures were explored, it was found that a single layer containing four hidden nodes connected to a single output layer worked best. The hidden layer uses the hyperbolic tangent sigmoid as its activation function, and the output node uses a simple linear function. The training process for the weights and biases of the neural network is described below.

Training. Training images were taken from the first three datasets. The last dataset, $\{\mathcal{U}(V_i)\}$, was fully withheld from training to act as a completely novel set of appearances, featuring new spatial patterns, to act as validation for the model. Since each of $\{\mathcal{U}(D65)\}$ and $\{\mathcal{U}(spot)\}$ consist of 144 appearance pairs and their distances, the 6-pair $\{\mathcal{U}(T_i)\}$ is augmented to produce additional training examples.

By evaluating M^{BRDF} at any two points, the curve provides additional realistic, physically-plausible appearance images. The perceived distance between those sampled images can be estimated using Equation 2. The paths were regularly sampled along α intervals in increments of 0.1, e.g. $(\alpha_i, \alpha_j) \in \{(0.0, 0.1), \ldots, (0.9, 1.0)\}$. Each path sampled in this manner formed 55 additional appearance pairs and distance estimates, for a total of 330 elements in the augmented $\{\mathcal{U}(T_i)\}$ dataset.

A fixed, equal number of samples were withheld from each of the three datasets to form test and validation sets. The validation set was used to perform early stopping to prevent over-fitting. The test set was fully withheld from training and provides an estimate at how well the network generalizes to new inputs. The Levenberg-Marquardt optimization algorithm [Hagan et al. 2014] was used to train the weights and biases of the final layer.

Evaluation. Table 1 summarizes the results of our model and existing models on the various appearance uniformity datasets. The top section shows the test RMSE values for the samples that were withheld during training, which are very similar to the overall performance of our model. This indicates that our model should generalize well to appearances similar in nature to those shown in Figure 2, as materials with that range of spatial patterning were used during training.

As is clear, none of the existing techniques perform consistently well on every dataset. The overall error rate and correlation with respect to human measurements for our model showed substantial improvement when compared to all existing metrics that were reviewed. $\{\mathcal{U}(V_i)\}$ was used as additional validation for our model's generality since it contains appearances with novel patterns, but otherwise fits within the domain of uniformity measurement for 3D printed materials.

[2] The remaining levels of the pyramid were found to have little additional information.

Table 1: Top table displays withheld test set accuracy, with training dataset breakdown. Bottom table compares accuracy of existing metrics and our model (italic) for all appearances in three datasets. Top performers are in bold. Absolute error is measured by RMSE; correlations are shown in parentheses as R^2 values.

	Test RMSE (R^2) During Training			
	$\{\mathcal{U}(D65)\}$	$\{\mathcal{U}(spot)\}$	$\{\mathcal{U}(T_i)\}$	Union
Our Model	*0.15 (0.6)*	*0.11 (0.8)*	*0.03 (0.9)*	*0.11 (0.8)*
	Net RMSE (R^2)			
	$\{\mathcal{U}(D65)\}$	$\{\mathcal{U}(spot)\}$	$\{\mathcal{U}(T_i)\}$	$\{\mathcal{U}(V_i)\}$
Δ RGB	0.18 (0.6)	0.20 (0.7)	0.91 (0.2)	0.26 (0.8)
ΔE '00	0.20 (0.6)	**0.12** (0.7)	0.83 (0.1)	0.29 (0.8)
S-CIELAB	0.23 (0.5)	0.41 (0.7)	1.65 (0.3)	0.23 (0.8)
perceptualdiff	0.28 (0.3)	0.38 (0.2)	0.55 (0.0)	0.43 (0.1)
HDR-VDP	0.20 (0.6)	0.36 (0.6)	0.70 (0.3)	0.21 **(0.9)**
Our Model	***0.14 (0.7)***	*0.13* ***(0.8)***	***0.05 (1.0)***	***0.17*** *(0.8)*

The error rate of our model on this novel dataset is comparable to its other performances and continues to out perform existing approaches. While existing image difference predictors could conditionally approach similar error rates, they struggled when the material and its ideal target differed due to more complex appearance changes.

6 CONCLUSION

This work demonstrates a novel approach to characterizing appearance non-uniformities that result from 3D printing. Building off the 3D printed color appearances previously studied using a visual search psychophysical study, we presented a new human subjects study to evaluate the perceptual rate of change between an appearance and an ideal target. These appearances feature spatial color variations, different gloss levels, and mesoscale bump profiles. After analyzing the collected human measurements, we successfully trained a small neural network built around a simple hierarchical image feature set that could accurately match the perceived non-uniformities. The neural network model performed better and more consistently than commonly used image-based difference metrics, including on materials with completely unique spatial patterns.

This model can be used to quantify the uniformity of printed materials and evaluate the quality of improvements to printing devices. Further experiments can be performed to increase its ability to cover other types of 3D printed materials.

We acknowledge that the range of appearances studied in this work is a focused subset of all possible appearances. This was done so that the problem was approachable and so that our solution could immediately be applied to real-world 3D printing issues. The success of the neural network approach to measuring uniformity provides a powerful incentive to continue collecting datasets of human appearance judgments. While uniformity is only one of the many appearance dimensions for spatially-varying materials, it still represents a step beyond past appearance research that constrained the problem to homogeneous materials. The techniques presented here can be repeated for novel appearance images to build a collection large enough to train a deep convolutional neural network that has more expressive power than our current model.

ACKNOWLEDGMENTS

Partially funded by a University of Minnesota Doctoral Dissertation Fellowship. [Muryy et al. 2014] provided the geometry.

REFERENCES

S. P. Arun. 2012. Turning visual search time on its head. *Vision Res.* 74 (Dec. 2012), 86–92.
Alan Brunton, Can Ates Arikan, and Philipp Urban. 2015. Pushing the Limits of 3D Color Printing. *ACM Trans. Graph.* 35, 1 (Dec. 2015), 1–13.
Peter J. Burt and Edward H. Adelson. 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Trans. Commun.* 31, 4 (April 1983), 532–540.
Edul N. Dalal and Kristen M. Natale-Hoffman. 1999. The Effect of Gloss on Color. *Color Res. Appl.* 24, 5 (Oct. 1999), 369–376.
James A. Ferwerda, Fabio Pellacini, and Donald P. Greenberg. 2001. A psychophysically based model of surface gloss perception. In *Proc. SPIE*. SPIE, 1–11.
Roland W. Fleming, Ron O. Dror, and Edward H. Adelson. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision* 3, 5 (June 2003), 347–368.
Andrew Gardner, Chris Tchou, Tim Hawkins, and Paul Debevec. 2003. Linear Light Source Reflectometry. *ACM Trans. Graph.* 22, 3 (July 2003), 749–758.
Ioannis Gkioulekas, Bei Xiao, Shuang Zhao, Edward H. Adelson, Todd Zickler, and Kavita Bala. 2013. Understanding the Role of Phase Function in Translucent Appearance. *ACM Trans. Graph.* 32, 5 (Sept. 2013), 1–18.
Martin T. Hagan, Howard B. Demuth, Mark H. Beale, and Orlando De Jesús. 2014. *Neural Network Design* (2nd ed.).
V. Havran, Jiří Filip, and Karol Myszkowski. 2016. Perceptually Motivated BRDF Comparison using Single Image. *Computer Graphics Forum* 35, 4 (July 2016), 1–12.
Richard S. Hunter and Richard W. Harold. 1987. *The Measurement of Appearance* (2nd ed.). Wiley-Interscience, New York.
Leonid L. Kontsevich and Christopher W. Tyler. 1999. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res.* 39, 16 (Aug. 1999), 2729–2737.
Michael Ludwig, Nathan Moroney, Ingeborg Tastl, Melanie Gottwals, and Gary Meyer. 2018. Perceptual Appearance Uniformity in 3D Printing. In *Proceedings of IS&T Electronic Imaging*. Society for Imaging Science and Technology, 1–12.
David L. MacAdam. 1942. Visual Sensitivities to Color Differences in Daylight. *J. Opt. Soc. Am.* 32, 5 (May 1942), 247–28.
Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2. In *SIGGRAPH*. ACM Press, New York, New York, USA, 1–13.
Phillip J. Marlow, Barton L. Anderson, and Juno Kim. 2012. The Perception and Misperception of Specular Surface Reflectance. *Current Biology* 22, 20 (Oct. 2012), 1909–1913.
Mathworks Inc. 2018. MATLAB and Neural Network Toolbox. (2018).
David Kirk Mcallister. 2002. *A Generalized Surface Appearance Representation for Computer Graphics*. Ph.D. Dissertation.
Alexander A. Muryy, Roland W. Fleming, and Andrew E. Welchman. 2014. Key characteristics of specular stereo. *Journal of Vision* 14, 14 (Dec. 2014), 1–26.
Thiago Pereira and Szymon Rusinkiewicz. 2012. Gamut Mapping Spatially Varying Reflectance with an Improved BRDF Similarity Metric. *Computer Graphics Forum* 31, 4 (July 2012), 1557–1566.
Tim Reiner, Nathan Carr, Radomír Měch, Ondřej Šťava, Carsten Dachsbacher, and Gavin Miller. 2014. Dual-Color Mixing for Fused Deposition Modeling Printers. *Computer Graphics Forum* 33, 2 (May 2014), 479–486.
Erwin Schrödinger. 1920. Grundlinien einer Theorie der Farbenmetrik im Tagessehen. *Annalen der Physik* 368, 21 (1920), 397–426.
Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. 2004. The CIEDE2000 Color-Difference Formula. *Color Res. Appl.* 30, 1 (Dec. 2004), 21–30.
Peter Vangorp, Jurgen Laurijssen, and Philip Dutré. 2007. The Influence of Shape on the Perception of Material Reflectance. *ACM Trans. Graph.* 26, 3 (July 2007), 1–10.
Weiming Wang, Haiyuan Chao, Jing Tong, Zhouwang Yang, Xin Tong, Hang Li, Xiuping Liu, and Ligang Liu. 2015. Saliency-Preserving Slicing Optimization for Effective 3D Printing. *Computer Graphics Forum* 34, 6 (Sept. 2015), 148–160.
Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a Perceptual Space for Gloss. *ACM Trans. Graph.* 28, 4 (Aug. 2009), 1–15.
Gunter Wyszecki and W. S. Stiles. 1982. *Color Science* (2nd ed.). Wiley-Interscience, New York.
Hector Yee, Sumanita Pattanaik, and Donald P. Greenberg. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph.* 20, 1 (Jan. 2001), 39–65.
Xuemei Zhang and Brian A. Wandell. 1997. A spatial extension of CIELAB for digital color-image reproduction. *J. Soc. Inf. Display* 5, 1 (March 1997), 61–63.