

Naturalness Perception of 3D Prints with Highlighted Features

Raed Hlayhel¹, Simon Hviid Del Pin¹, Takahiko Horiuchi², Sony George¹, Jon Yngve Hardeberg¹, Aditya Suneel Sole¹

¹Colourlab, Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway;

²Graduate School of Informatics, Chiba University, Chiba, Japan

E-mail: raed.hlayhel@ntnu.no

Abstract. Naturalness is a complex appearance attribute that is dependent on multiple visual appearance attributes like color, gloss, roughness, and their interaction. It impacts the perceived quality of an object and should therefore be reproduced correctly. In recent years, the use of color 3D printing technology has seen considerable growth in different fields like cultural heritage, medical, entertainment, and fashion for producing 3D objects with the correct appearance. This paper investigates the reproduction of naturalness attribute using a color 3D printing technology and the naturalness perception of the 3D printed objects. Results indicate that naturalness perception of 3D printed objects is highly subjective but is found to be objectively dependent mainly on a printed object's surface elevation and roughness.

Keywords: naturalness, 3D printing, material appearance, subjective analysis

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

[DOI: 10.2352/J.Percept.Imaging.2025.8.000501]

1. INTRODUCTION

Naturalness is a key attribute that is used to assess the appearance of an object [15]. It is a complex attribute that involves the interaction of a multitude of other appearance attributes, mainly color, roughness, and gloss [15]. This complexity is evident when defining naturalness, which can be approached in two ways:

- (1) Objects that are not manipulated by humans, that is, possessing the quality of being in accordance with nature; for example, natural wood versus varnished wood [19].
- (2) Close matching between the understanding of a scene in the observer's mind—including (but not limited to) materials comprising the scene and scene depth—and the memory of such scenes and materials in an observer's memory [12].

The main difference between these definitions is that the first is more restrictive compared to the second. While the former only considers objects as being natural if they are

in their “raw” form, the latter relates naturalness more to the realism of the scene presented. These definitions reveal a layer of complexity when understanding the naturalness attribute, and how naturalness perception is dependent on an observer's preferences and familiarity with the scenes and materials involved.

But despite its complexity, naturalness remains an important attribute to achieve correctly for an accurate and high-quality reproduction [19, 34]. Studies have shown that observers react more favorably to natural-looking samples as opposed to fake-looking replicas. This was the case in the study by Overvliet and Soto-Faraco [19], where observers deemed natural-looking wood samples more valuable than fake-looking samples, as well as in the study by Reinhard et al. [23], where their “natural look” was essential for the acceptance of eye prostheses by patients. Interestingly, naturalness judgment is also considered when evaluating “unnatural” objects, such as 3D printed processed foods. Groot [10] has shown that when 3D printed food was considered more natural looking, its acceptance by observers was higher. And still, naturalness is highly sought after in many other applications such as cultural heritage conservation/reproduction, interior decor/design, and art [15].

3D printing technology is a process where three-dimensional objects are constructed layer by layer from digital design files. It offers many advantages over traditional manufacturing techniques, such as enabling the production of complex and customized objects and cheaper manufacturing [6]. 3D printing is an umbrella term for many different techniques that use different materials, and are employed for different applications ranging from industrial manufacturing to artistic reproduction all the way to building construction [18]. Of these techniques, PolyJet 3D printing allows for color 3D printing. A PolyJet 3D printer uses UV-curable colored ink droplets that are jetted from inkheads and cured (solidified) using UV-light. The positioning of the ink droplets is determined by an error-diffusion color halftoning step that is intrinsic to the printing pipeline [38]. This technique is useful for a certain approach called graphical 3D printing: the reproduction of an object's appearance using 3D printing [31]. This involves matching the appearance attributes of the reproduced model with the original so that the reproduction

looks and feels the same as the original object. As naturalness is a key attribute in appearance reproduction, it significantly contributes to the perceived quality of an object.

Although progress has been made in understanding how humans perceive the naturalness of objects for 2D and 2.5D applications, research on naturalness perception for 3D printing applications is still in its early stages. This is also true for studies investigating the influence of algorithmically generated surface textures, printed at different elevations, on the perceived naturalness of 3D printed objects. Previous research has explored naturalness perception in various contexts, including 2D images [34] and 2.5D printed reliefs [14, 35]. These studies have identified appearance attributes such as color, gloss, and roughness as important contributors to naturalness. These attributes were linked to technical variables that allow controlling an object's perceived naturalness like its roughness, texture elevation [15, 16], and surface texture profile [35]. However, the challenges of achieving naturalness in 3D printed objects, specifically the complex interactions between elevation levels, texture roughness, and surface texture profiles, remain unexplored. An investigation into the inherent subjectivity of naturalness perception by observers, by using adequate statistical models, is also missing from previous research.

This paper aims to fill in these gaps in understanding and controlling naturalness perception using color 3D printing by investigating the influence and interaction between algorithmically generated surface textures and their elevation on perceived naturalness. This study also employs advanced statistical methods to model observer subjectivity in naturalness perception. We hypothesize that image processing algorithms can be used to produce more meaningful surface texture profiles and to control surface features, such as roughness, as demonstrated by Wang et al. [35]. These generated surface texture profiles would produce more natural-looking 3D printed samples, but that also depends on the texture elevation levels applied. Based on the findings of Kadyrova et al. [15], we hypothesize that lower elevation levels enhance perceived naturalness. Finally, we consider Bayesian analysis to reveal individual differences in naturalness perception that are not captured by the statistical methods already used in the previous literature.

We investigate this hypothesis by applying novel texture extraction algorithms on reference 2D images to produce semantically meaningful 3D surface texture reproduction with different elevations. We then conduct a subjective experiment to qualify the perceived naturalness of the 3D printed samples. Instead of relying on simple Mean Opinion Scores or z -scores, we use a Bayesian approach that accounts for individual observer differences and preferences, providing deeper insights into naturalness perception.

The main contributions of this work are as follows:

- (1) a novel application of texture extraction algorithms for generating displacement maps in color 3D printing;
- (2) a rigorous subjective evaluation of the impact of these algorithms and varying elevation levels on perceived naturalness;
- (3) the use of a Bayesian statistical approach to analyze ordinal naturalness ratings, providing a more nuanced understanding of observer preferences;
- (4) identification of key texture features that significantly influence naturalness perception in 3D printed objects.

This article is organized as follows. A brief literature review of similar works is presented in Section 2. Section 3 details the methodology, explaining the generation of displacement maps, the design and printing of 3D models, and the assessment procedure. Results are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes the paper.

2. RELATED WORK

We are interested in the reproduction and assessment of naturalness using PolyJet 3D printing. As discussed in Section 1, the naturalness of a sample helps in defining its worth and quality in a standard observer's mind. However, naturalness is affected by a multitude of variables and is therefore a complex attribute to reproduce. Nevertheless, the literature identifies several key factors influencing naturalness reproduction.

2.1 2.5D and 3D Printing

Since our work involves 3D forms, we focus primarily on findings related to samples with relief, including both 2.5D and 3D methods. Also called relief printing, 2.5D printing uses a similar layer-by-layer approach to that of 3D printing. However, unlike 3D printing, it cannot build true 3D objects and is only used to print texture variations or slightly raised surfaces [1]. As 3D printing is also relevant to such applications, a primary distinction between 3D and 2.5D printing is the use of support material: support is necessary for complex geometries in 3D printing but generally not required in 2.5D printing due to its limited depth [20].

But although 2.5D and 3D technologies are considered different, they do share some similarities. Indeed, both 2.5D and 3D technologies can use the same techniques and materials during printing [20]. This is true for color 2.5D printing and PolyJet 3D printing for example, where both techniques use inkjet UV-curable CMYKW ink in their printing process. This means that findings from studies using one technology or the other can be generalized interchangeably, as the printed samples share the same manufacturing process and the features/biases related to it. Accordingly, we include findings from studies conducted on naturalness of 2.5D printed samples as well. In the present investigation, we used a 3D printer, and so we refer to our samples as 3D printed.

2.2 Naturalness Reproduction Variables

Naturalness assessment for objects with relief is not a straightforward process for observers, and it depends on a multitude of factors [14]. An investigation into naturalness perception of 2.5D samples depicting different materials such as wood, stone, and so on was conducted by Kadyrova et al. [15]. They found that surface texture, characterized by surface elevation and roughness, is a prominent factor

in judging a 2.5D object's naturalness. The aforementioned work was continued in another study [16] only for wood samples, where reliefs of lower elevations were printed. Together, the studies found that observers had a preference for midlevel elevations at 0.5 mm of maximum relief height to be the most natural looking. Van Hoey et al. [32] found that visual distortions caused by surface texture modifications on relief samples decreased the perceived quality among observers. But aside from the visual aspect, surface texture is also defined by its tactile attribute, roughness. Tymms et al. [30] demonstrated that the shape and size of the facets making up the surface of a 3D printed object widely varies the tactile perception and understanding of the said object. They also showed that surface texture depends not only on elevation but also on the positioning of surface facets.

2.3 Texture Extraction and Implementation

Since surface texture depends on the positioning of the facets, added relief should be meaningful to improve tactile perception as shown by another study by Tymms et al. [29]. When based on 2D images, relief was added using simple grayscale images of these colored 2D images [15]. In contrast, Wang et al. [35] investigated better methods to add relief to printed samples. They introduced a methodology that combines texture analysis and semantic understanding to enable the reproduction of textures in 2.5D prints. This involves identifying texture features using filters in the frequency domain [22], such as a high-pass filter or Gabor filter. Wang et al. [35] concluded that incorporating semantic understanding into the printing process produces more realistic and perceptually rich 2.5D prints that closely resemble the original textures. Furthermore, Texture-Aware Error Diffusion Halftoning (TAED), like the algorithm developed by Li et al. [17], may also produce semantically relevant relief placement. In their investigation [17], Li et al. demonstrated that TAED outperformed traditional error-diffusion methods (Floyd–Steinberg [8] and Shiau [24]) both objectively (using 2D image quality metrics, such as the the Sparse Feature Fidelity, Peak Signal-to-Noise Ratio, Mean-Structural Similarity, and computation time) and subjectively (via psychometric experiments). Therefore, TAED offers an innovative approach to digital halftoning that aims to enhance image quality, particularly in preserving texture details.

3. METHODOLOGY

Our work in this study is a multistep process that involves producing samples with varying appearances using 3D modeling, 3D printing of these models, and then subjectively assessing their naturalness through observer categorization and analysis. First, displacement maps are generated from reference images to produce different surface profiles and textures for 3D models. Then the 3D models are printed using PolyJet 3D printing and assessed via a categorical psychometric experiment where observers categorize the naturalness of each sample. Finally, the results of the said experiment are analyzed to determine the prevalent

appearance attributes in naturalness perception of 3D samples.

3.1 Generation of Displacement Maps

Displacement maps, hereafter referred to as DM(s) in the paper, were generated using RGB reference images (images obtained from Pixabay [21], an open-source image database). Considering familiarity and everyday use, these reference images were selected from four different material categories: Wood, Stone, Fruit, and Cultural Heritage Object (two from each category), and are 591×591 pixels wide due to the technical limitations of the 3D printer (refer to Section 3.2 for more details). The cultural heritage images were chosen for their complex surface ornamentation, which presents a challenging scenario for texture reproduction, allowing us to assess the capabilities and limitations of the various algorithms under demanding conditions. The intricate details of the Alhambra ornamentation, for example, characterized by fine lines, sharp edges, and subtle variations in depth, provide a rigorous test for the algorithms' ability to capture and reproduce fine details that are crucial to naturalness perception. However, we acknowledge a potential limitation associated with the use of the Alhambra image. The naturalness judgments in our study are based on the participants' perception of how well the 3D printed samples match their internal representation of the depicted objects or materials. Some participants may be unfamiliar with the Alhambra and its specific ornamentation. This unfamiliarity could potentially influence their judgments, as they may not have a clear mental image of the original structure for comparison. Although this could introduce some variability into the naturalness ratings for the Alhambra samples, the complexity of the image still provides valuable insights into the performance of the DM algorithms.

DMs are grayscale images used in 3D design software to physically alter a 3D object's surface [27]. Physical alteration of a 3D object's surface generates more realistic micro-surfaces, enhancing its appearance through improved light-surface interaction, resulting in better shading and light reflection in the 3D model.

DMs are generated from the eight reference images shown in Figure 1. First, a pixel-wise transformation is applied to calculate a grayscale DM (Gray DM) from the 2D RGB image [26]. Thus, for an image with $N \times M$ pixels, the grayscale value at the pixel (i, j) is calculated using Eq. (1).

$$Y_{i=1\dots N, j=1\dots M} = 0.2989 \times R_{i,j} + 0.5870 \times G_{i,j} + 0.1140 \times B_{i,j}. \quad (1)$$

In Eq. (1), Y represents the grayscale level and R , G , and B are the RGB values of the corresponding pixel in the 2D image. This is a standard RGB-to-grayscale transformation that represents grayscale in a manner aligned with human vision. Consequently, it assigns a higher coefficient to green, followed by red, and then blue based on the eye's sensitivity to each color [26]. The Gray DM, therefore, serves as a starting point for texture extraction and DM generation. Six DMs were generated in MATLAB (version 2023b) using the

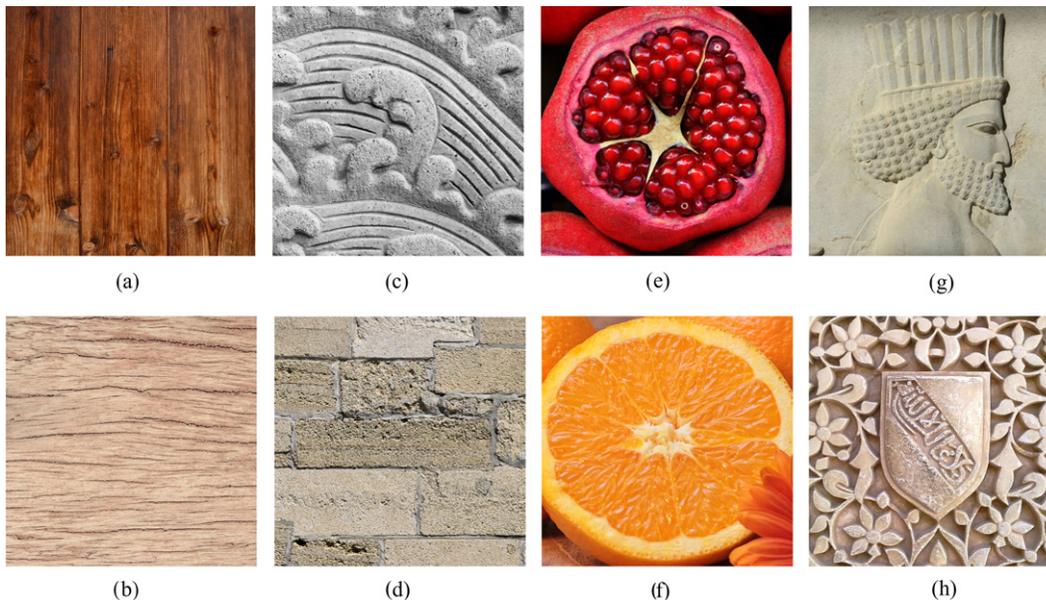


Figure 1. RGB reference images in different categories showcasing different material types. From left to right, the categories are Wood, Stone, Fruit, and Cultural Heritage Object. Individually, the samples will be referred to later on as follows: (a) Reddish Wood; (b) Natural Wood; (c) Stone Fresco; (d) Textured Stone Wall; (e) Pomegranate; (f) Orange; (g) Iranian; (h) Alhambra. For the remainder of the paper, as the editing and printing processes are the same for all images, we illustrate the process for only one image, the Alhambra image.

Gray DM where different appearance attributes like color, roughness, and gloss were considered. Table I presents the description of the Gray DM and the six DMs generated. The DMs were generated using a specific set of algorithms. These DM-generating algorithms are not the only ones we could have used but were chosen because they allow us to have DMs that are related to the content of the image and therefore reproduce texture in a meaningful way. The choice of algorithms is part of the novelty of this work because as far as we know, they have not been used in 3D texture printing before. Therefore, the choice of algorithms is not based on a ranking of performance, it is rather based on the surface texture features we want to control: roughness, foreground highlighting, and edge highlighting. Consequently, we employ algorithms that enable us to control these features either individually or in combination. Furthermore, these algorithms are based on standard image processing functions that are mostly in-built in widely used image processing software like MATLAB and Python. For instance, we used binarization algorithms like *imbinarize* for image forefront highlighting. We also employed halftoning algorithms like Minimized Average Error Diffusion (MAE) and TAED (see Table I) for pixel-wise roughness addition. However, we had to limit the number of algorithms due to practical constraints: a large selection would have resulted in longer printing times and more complex and time-consuming subjective experiments for observers. However, although these algorithms are useful for feature extraction, they also have disadvantages. These range from information loss due to halftoning to a reliance on grayscale images, making the output highly sensitive to shading and gloss variations in the reference images. Hence, this is by no means a definitive

list of algorithms to always use, and future research might uncover novel and better ways to extract and control surface texture from reference 2D images. Thus, the selection of algorithms for this study was primarily based on the variables we wanted to control that are specific to this study.

As an example, Figure 2 illustrates the DMs generated for the Alhambra image.

3.2 Generation of 3D Models and 3D Printing

The 2D images were converted to 3D models by applying the DMs using the *Displace* modifier in the rendering software Blender (version 3.0) [2]. The midlevel was set to zero, ensuring that all micro-facets were added onto the surface rather than being etched into the model. The strength was adjusted to ensure that the 3D models had three different elevation levels: 0.75 mm, 1 mm, and 1.5 mm. The models were exported as .obj files and loaded as such into the 3D printer software. Figure 3 illustrates the transformation of the DM into different elevations depending on the grayscale values according to Eq. (2).

$$d_{i,j} = Y_{i,j} \times E_{i,j}. \quad (2)$$

In Eq. (2), d is the elevation height, Y is the grayscale value of the DM at a given pixel (i,j) , and E is the maximum elevation to be printed (for $Y = 1$) in the 3D sample. Figure 4 illustrates the simulation of a model generated using a DM with three different elevation levels.

To limit the variables affecting naturalness reproduction to *elevation*, *roughness*, and *surface texture*, the color information (in the RGB color space) from the 2D images was UV-mapped (see Fig. 1) onto the 3D model's surface before 3D printing. As we are working with cubes that are

Table I. Displacement maps and their description.

DM No.	Name	Description
1	Gray DM	Generated using the standard RGB-to-grayscale transformation presented in Eq. (1). It also served as a starting point for generating the rest of the DMs.
2	ForeFront DM	A Gaussian adaptive image threshold is applied. It separates the image into two sets of pixels based on a local threshold. Practically, this separates the image into a binary grouping: the foreground (pixel value = 255) and the background (pixel value = 0). The binary image is then multiplied by the original grayscale values to preserve the 8-bit range of pixels (instead of having a binary 1-bit output). Unlike general image threshold techniques that apply the same threshold all over the image, adaptive techniques calculate the thresholds locally in a set neighborhood of pixels, which leads to a better separation of pixels [3]. ForeFront enables control over relief height, thereby adding more sharpness to the printed sample.
3	1bit_TAED DM	This DM is a halftoned ($2^1 = 2$ gray levels, black and white) image, where the texture is considered in the dot placement. We are interested in applying this DM because it allows us to add roughness to our samples because of its halftoned output.
4	2bit_TAED DM	Similar to the 1_bit DM but with multilevel halftoning. This means that the DM is not simply black and white but also contains two intermediate grayscale levels ($2^2 = 4$ gray levels). This allows for a smoother 2.5D surface variation.
5	MAE DM [13]	Similar to the 1_bit halftoned DM. Unlike TAED, there is no intermediary texture detection step. MAE is a standard error-diffusion method and serves as a benchmark for evaluating the performance of the TAED algorithm in 3D.
6	ForeFront-MAE DM	Generated in two steps: (1) the foreground is separated from the background using the Gaussian adaptive image threshold presented earlier; (2) the separated output is halftoned using the MAE halftoning method. This gives us a forefront-separated, halftoned DM. This, in theory, allows us to control the elevation and add roughness to the 3D samples.
7	Canny Edge Detection DM [7]	Generated using the Canny Edge detection method, which produces a binary image where the edges of structures in reference images are highlighted. This method is included to assess whether edge highlighting sufficiently induces the perception of appropriate physical texture in 3D samples.

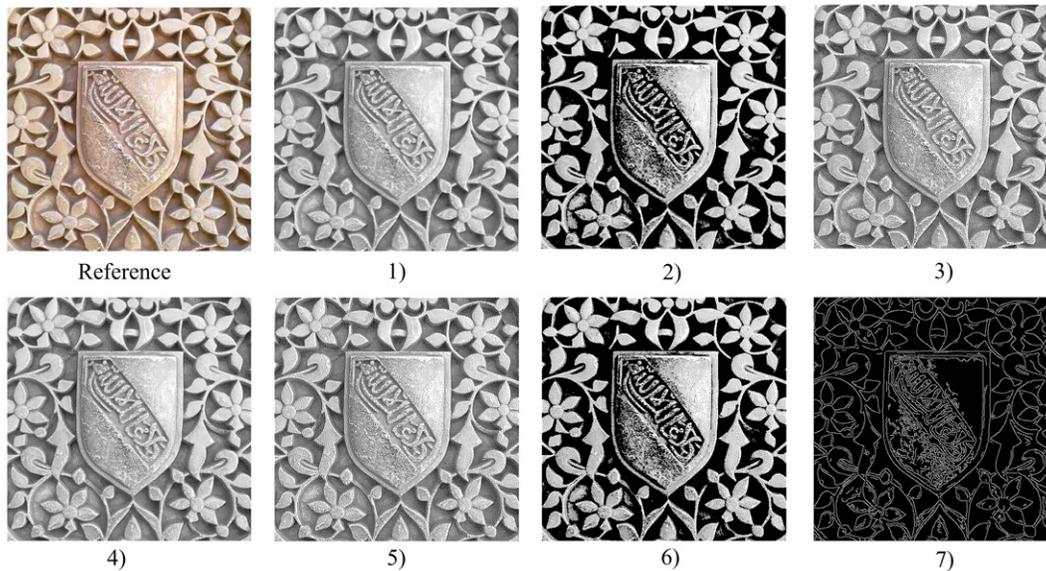


Figure 2. DMs generated for the Alhambra image.

rather simple 3D shapes, no complex UV-unwrapping was needed. We separated the UV-facets of the cube manually. Then, we identified the UV-coordinates corresponding to each face of the cube and added the color information to the

corresponding cube face by UV-mapping the reference 2D image onto the upper face of the cube.

With seven DMs, eight reference images, and three elevation levels, a total of 168 ($7 \times 8 \times 3$) samples were

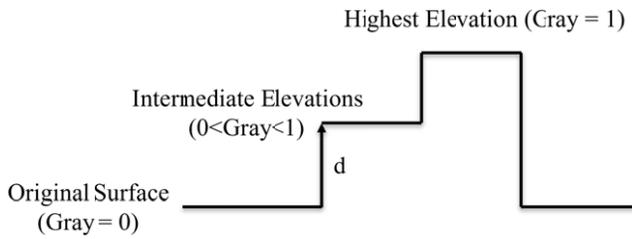


Figure 3. Elevation levels corresponding to different grayscale values.

3D-printed. For printing, we used the Stratasys J55 multimaterial PolyJet 3D printer [25], hereafter referred to as J55 in the paper, to print the 3D models. The ink used is CMYKW (cyan–magenta–yellow–key (black)–white) ink belonging to the Vero family of UV-curable ink [33]. This ink family is characterized by its translucent, specular, and scattering optical properties. The J55 can print with a resolution of 300 dpi (dots per inch) in the X and Y directions, and a resolution of 1354 dpi in the Z direction. This corresponds to a voxel of $0.08 \times 0.08 \times 0.018$ mm in size. We used the GrabCAD Print 3D printing software for 3D model orientation, slicing, and printing. We printed the samples in high-quality mode, with a print time of 4.5 h per print project, with each print project containing 20 samples. To balance material use, printing time, and the need for sufficient detail clarity, the 3D samples were designed with dimensions of $50 \text{ mm} \times 50 \text{ mm} \times 5 \text{ mm}$. This size ensured that the printed features were discernible while minimizing material waste and avoiding excessively long printing times. Furthermore, we used translucent materials on the J55 printer, making it necessary to print thicker samples than usual to obtain an opaque object. The 3D printed sample dimensions and the DM pixel dimensions were kept equal to avoid any automatic stretching or compression of the DM pixels. This was achieved by maintaining a fixed pixel ratio of the 2D image before generating its corresponding DM. That is, for a maximum resolution of 300 dpi for an object of $50 \text{ mm} \times 50 \text{ mm}$ in dimension, the corresponding number of pixels would be 591×591 . After printing, the samples were verified against the expected outcome through visual inspection. We also verified that the printed elevations match the theoretical values by 3D-scanning some of the samples and comparing the sizes of the 3D models and the scanned models. Figure 5 shows the resulting printed samples of the Alhambra image.

3.3 Evaluating the Naturalness Attribute of 3D Printed Samples

A categorical judgment psychometric experiment was conducted to evaluate the perceived *naturalness* of the 3D printed samples. Fifteen observers (7 female, 8 male, average age = 29 years, standard deviation = 8 years, age range [22, 52]) participated in the experiment. None of the participants is an expert in 3D printing although all but two had prior knowledge of color science. The experiment adhered to the ethical guidelines for the “Protection of Research Subjects” set by the Norwegian National Committee for Research

Ethics in Science and Technology [9]. Participants provided signed consent before commencing the experiment, and their anonymity was ensured. Although the number of participants in this experiment is in the acceptable range [11], we acknowledge that we are at the lower end of that range. This should be considered a potential source of bias in our study, and future work needs to heed this limitation.

For this experiment *naturalness* was defined as a matching between the memory of an object in an observer’s mind and the scene portrayed in each sample in this experiment [15]. Participants were shown all of the 168 3D printed samples divided into 8 sets of 21 samples each. Each set corresponds to the samples derived from each reference image shown in Fig. 1. The categorization took place in a Gretag Macbeth Judge II viewing booth under a CIE D50 illuminant. Ambient lighting conditions were set to darkness. Five categories, as listed below, were used to categorize the 3D printed samples.

- (1) Very Un-natural
- (2) Un-natural
- (3) Adequate
- (4) Somewhat Natural
- (5) Natural

Participants were not provided with any reference images to aid them in categorizing the samples. Participants relied on visual assessment of the samples and could also touch the samples (without lifting the sample from the booth floor) to feel the surface texture. Each observer spent approximately 75 minutes completing the experiment. Figure 6 shows the experimental setup.

To understand how participants categorized the naturalness of samples, it would be important to understand their thought process when judging a sample’s appearance both in terms of the naturalness definition used and the attributes considered. Therefore, each session was audio-recorded in its entirety with the participant’s consent. Participants were encouraged to describe all the steps they were following and to describe their thought process when ranking the naturalness of 3D samples along with a few statements when they were done categorizing each set of samples. The audio recordings were then transcribed, and a word frequency analysis was conducted to determine which attributes the observers relied on during their assessment.

Raw individual scores were recorded during the experiment and analyzed. With data collected being categorical and the difference between the categories being variable, we consider our data as ordinal [37]. This type of data should not be misinterpreted as interval data, as this can lead to conceptual and statistical inaccuracies. For instance, it is not necessarily true that a rating of 4 (Somewhat Natural) is twice as good as a rating of 2 (Un-natural) or that the improvement from 1 (Very Un-natural) to 2 (Un-natural) is equivalent to the improvement from 4 (Somewhat Natural) to 5 (Natural). Therefore, it is crucial to handle ordinal data appropriately to avoid misconceptions and incorrect interpretations. Cumulative Link Mixed Models (CLMMs)

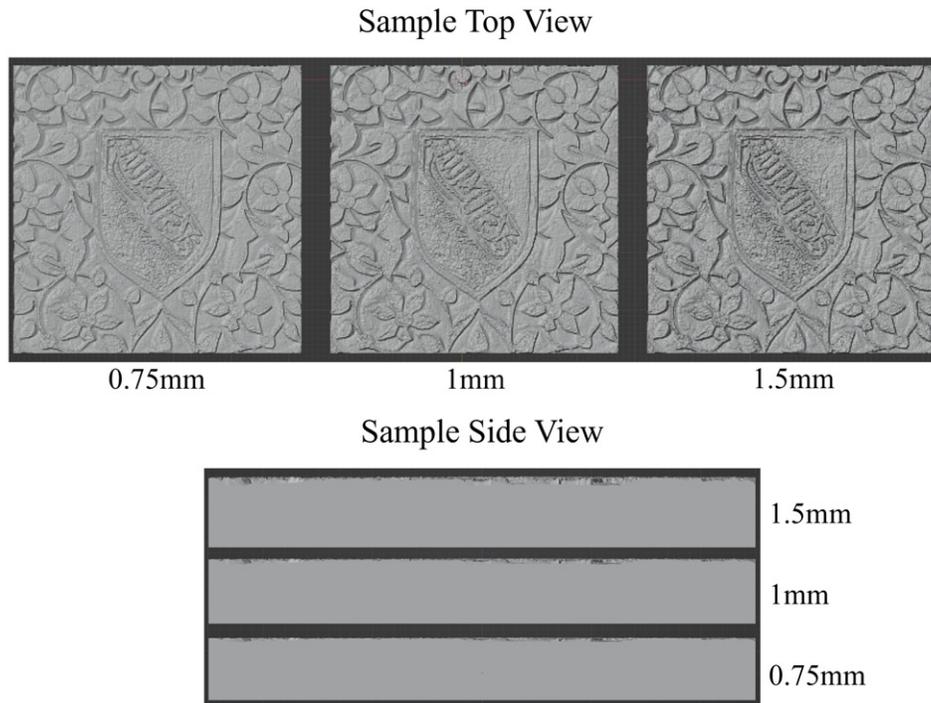


Figure 4. Illustration of three elevations using the ForeFront DM. It can be observed that with an increase in maximum elevation, the sample surface and the difference between peaks and valleys become more pronounced.



Figure 5. Twenty-one printed samples of the Alhambra set. The change in appearance attributes (such as sharpness, elevation, roughness, and color) resulting from variations in surface structure is visible.

are a fitting method for analyzing ordinal data. They operate under the assumption that the observed ordinal variable is a categorization of an underlying continuous variable [5]. This enables CLMMs to provide the probability for each rating rather than a single prediction. The Bayesian approach, which can be implemented using the *brms* package in R, is used to fit these models [4]. This approach offers several benefits over traditional frequentist statistics, including the ability to incorporate prior knowledge, quantify uncertainty, and estimate complex models. This makes it a powerful tool for analyzing ordinal data. In this paper, the Bayesian approach was deemed more appropriate as naturalness

perception and sample categorization can be highly subjective, that is, incurring random effects [28]. Therefore, we use it to estimate the likelihood of sample categorization.

Using *Observer* (referred to as *ID*), *Algorithm*, *Elevation Level*, and *Original Image* as variables, seven statistical models were generated using different variable combinations. We include these variables because we know from the literature that elevation [16] and algorithm [35] affect the naturalness and reproduction quality of 3D printed samples. We also include the participants and images as variables because, as shown in the literature [28], these variables incur random effects that arise from the participants' understanding of



Figure 6. An observer taking part in the subjective experiment.

naturalness and image content shown. We begin with simpler statistical models that include only a few variables. We then formulate the said models and test their predictive performance. After testing, model complexity is increased by adding more variables or by incorporating interactions between existing variables. Through this iterative process and with our selection of variables already determined, we construct and evaluate various statistical models based on their predictive power. Once we build a model with strong predictive power, that is, model prediction values match the observed data, we use the said model for result analysis. All variables in this model are expected to influence the naturalness rating. The effect of each variable on the model can then be examined to test our initial hypotheses.

We start with models that assume that variables are independent at first, but we do not neglect the probability of

interaction between the different variables in the naturalness rating given during the subjective experiment. Therefore, we also include models that take into consideration this variable interaction between a select number of variables. Table II gives a detailed explanation of the statistical models used.

4. RESULTS

These models were evaluated by calculating the Widely Applicable Information Criterion (WAIC) score. WAIC is a Bayesian extension of the Akaike Information Criterion, where a lower WAIC value indicates better performance of the statistical model [36].

Table III presents the performance of the seven models based on the WAIC score. The lower the WAIC score, the better the model fits the data while penalizing for model complexity by adding more predictors. The ImageNest and IDNest models, despite their complexity, demonstrate good balance between the fit and parsimony. Given the assumptions underlying WAIC, these models offer the most insightful understanding of the factors influencing the rating.

Figures 7–11 show the results for the best-performing models: ImageNest and IDNest. Figs. 7 and 8 illustrate the model’s prediction performance while Figs. 9–11 illustrate the change in naturalness perception both globally (across all observers) and individually across the different variables in this study.

5. DISCUSSION

Looking at Table III, we see that model performance varies considerably depending on the choice of included variables. Models with a subset of variables included perform poorly compared to those where all variables are accounted for. This validates our initial choice of variables, meaning that all chosen variables have an effect on naturalness rating,

Table II. Statistical models and their description.

Model No.	Name	Description
1	Observer-Image	Includes <i>ID</i> and <i>Image</i> as independent variables. Suggests that the rating is influenced by the individual observer and the image being assessed, where each variable can lead to higher/lower ratings.
2	Observer-Elevation	Includes <i>ID</i> and <i>Elevation</i> as independent variables. Implies that the rating is influenced by the individual observer and the elevation level of the image.
3	Algorithm-Elevation	Includes <i>Algorithm</i> and <i>Elevation</i> as independent variables. Suggests that the rating is influenced by the algorithm used and the elevation level of the image.
4	Combined	Includes all variables <i>ID</i> , <i>Elevation</i> , <i>Image</i> , and <i>Algorithm</i> independently. Implies that the rating is influenced by all these factors with no interdependence between them.
5	Interaction	Includes an interaction term between <i>Algorithm</i> and <i>Elevation</i> in addition to the variables <i>Image</i> and <i>ID</i> . It suggests that the effect of the algorithm on the rating is linked to the elevation level and vice versa.
6	ImageNest	Includes <i>Algorithm</i> , <i>Image</i> , <i>ID</i> , and <i>Elevation</i> with a nested structure where <i>Algorithm</i> varies within <i>Image</i> . It suggests that the effect of the algorithm on the rating may differ for different images, implying there is a dependence between algorithm performance and image content.
7	IDNest	Similar to ImageNest but with <i>Algorithm</i> varying within <i>ID</i> . It suggests that the effect of the algorithm on the rating may differ for different observers, implying there is an individual preference for certain algorithms between observers.

Table III. WAIC scores of statistical models. Models ImageNest and IDNest show the lowest WAIC scores among the seven models.

No.	Model name	Variables	WAIC
1	Observer-Image	ID + Image	7890.9
2	Observer-Elevation	ID + Elevation	7700.9
3	Algorithm-Elevation	Algorithm + Elevation	6405.8
4	Combined	ID + Elevation + Image + Algorithm	6169.4
5	Interaction	Interaction (Algorithm*Elevation) + Image + ID	6131.6
6	ImageNest	Algorithm \ Image + ID + Elevation	6018.2
7	IDNest	Algorithm \ ID + Image + Elevation	6017.3

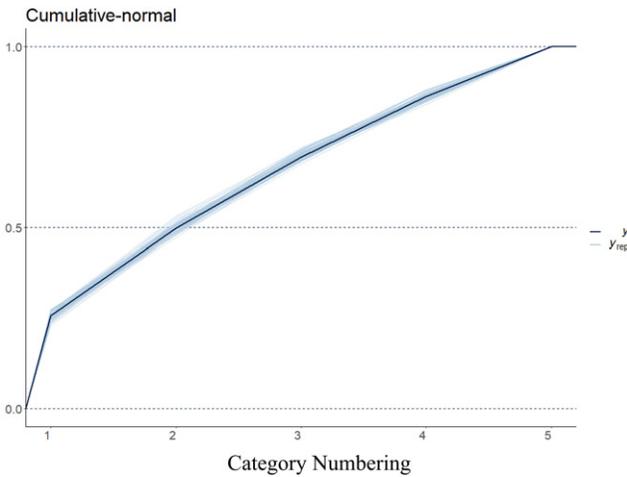


Figure 7. Model prediction (y) versus observed data (y_{rep}). Model performance comparison based on the cumulative normal distribution of predicted naturalness ratings. The plot displays the cumulative probability of each naturalness rating category (1: Very Un-natural, 2: Un-natural, 3: Adequate, 4: Somewhat Natural, 5: Natural) as predicted by the best-performing model (IDNest). The x-axis represents the naturalness rating categories while the y-axis represents the cumulative probability. The curve illustrates the overall distribution of predicted ratings, with a “right-leaning” shape indicating a higher probability for lower naturalness ratings. This aligns with the general observation that a considerable portion of the 3D printed samples were perceived as relatively unnatural. The shaded area around the curve represents the 95% credible interval, reflecting the uncertainty in the model’s predictions.

albeit to varying extents. Furthermore, the variables should not be treated independently as evidenced by the Interaction model’s superior performance compared to the Combined model. With the nested interaction among *Algorithm*, *Image*, and *ID* giving the highest predictive performance, we also validate our starting hypothesis of observer preference and image content effect on naturalness perception. With this initial verification of our choice of variables and hypotheses, we proceed to extract data from the best-performing prediction model.

As shown in Figs. 7 and 8, both models perform well, and the results obtained from the subjective psychometric experiments can be used for further evaluation. We can see that the predicted values (y) are close to the observed

data (y_{rep}) both generally and per algorithm. This shows that our model performs well and that the results from the said models can be trusted and used for further analysis of subjective data. The cumulative normal curve (refer to Fig. 7) being a “right-leaning” curve, the subjective experiment revealed that lower categories had more counts than higher categories, indicating that over half of the 3D samples appeared unnatural to observers. The model estimates are in accordance with the observed data in general although with some inconsistencies in the predictions. This may be due to the ratings being inherently noisy or that there are still some factors missing from the model. Nonetheless, it is still possible to rely on our model for further analysis. These prediction results confirm our first hypothesis concerning the ability to use more complex statistical tools for more in-depth analysis of naturalness perception.

To evaluate general trends and observer-specific preferences from the subjective experiment responses, we look at Figs. 9–11. From Fig. 9 we can see that observers as a group perceived the Gray DM output as the most natural and the Canny Edge output as the least natural output from the seven DM algorithms. It is also apparent that observers preferred the lowest elevation (0.75 mm) for all algorithms except for Gray DM, where the 1 mm elevation was slightly more favorable. Furthermore, the decrease in observer response scores between elevations was not consistent across all algorithms. For example, the ForeFront algorithm shows the steepest drop in scoring, going from an average score of 3.59/5 at 0.75 mm to 2.49/5 at 1.5 mm, representing a 1.1-point drop in average score. However, this sharp decline was less pronounced for some algorithms, notably the Gray DM (mean score difference = 0.35), and 2bit_TAED (mean score difference = 0.7). So overall, there is a clear preference for Gray DMs and lower elevations for naturalness reproduction. These findings confirm our initial hypothesis that lower elevations are preferred for a more natural output. Furthermore, the choice of algorithm used has a significant effect on the naturalness rating and consequently, its perception. Therefore, we validate our starting hypotheses from the findings in Figs. 7–9.

These results represent the aggregate responses of all observers. However, the statistical models reveal that analyzing scoring at the individual level provides a deeper understanding of the scoring process. To test our final hypothesis related to the subjectivity of naturalness perception, we look at the findings from nested models (see Table II): ImageNest, which accounts for the effect of image content on naturalness ratings, and IDNest, which accounts for observer preferences. Figs. 10 and 11 illustrate the rating shift per algorithm for the reference images used and for observers, respectively. The rating shift represents the difference between an individual’s score and the average score for the group. The group average is indicated by the dashed black line; thus, the rating shift depends on the positioning of an observer’s scoring curve relative to the dashed black line. If an observer’s curve for a given algorithm is above the group average, it indicates that the observer

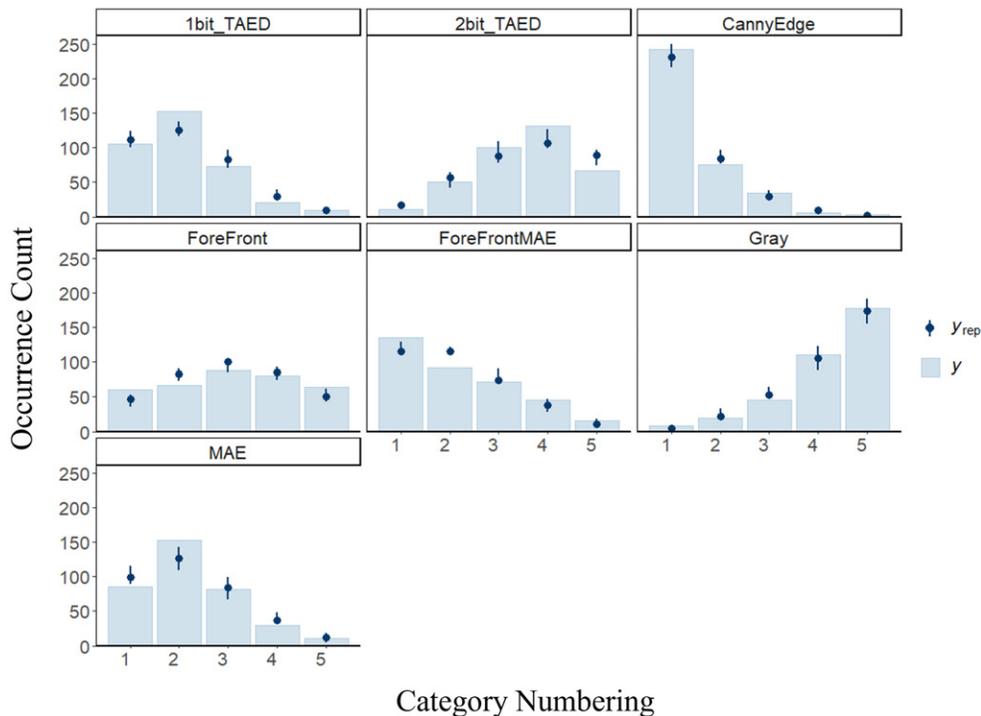


Figure 8. Comparison of model predictions (y) against observed naturalness ratings (y_{rep}) for each of the seven DM algorithms. Each panel displays a scatter plot where the x -axis represents the observed ratings and the y -axis represents the model’s predicted ratings. The closer the data points cluster around the diagonal line ($y = x$), the better the agreement between the model’s predictions and the observed data. Although some deviations are noticeable, attributable to the inherent variability in subjective ratings, the general trend indicates that the model adequately captures the overall patterns in the data. The model shows particularly good agreement for the Gray DM algorithm, which was generally perceived as the most natural, while larger deviations are observed for the Canny Edge algorithm, which was generally perceived as the least natural.

rated that specific algorithm higher than average and vice versa. For instance, ID3 is an observer that generally rates higher than the group (as their rating curve consistently lies above the group average) while ID10 generally assigned lower ratings than the group. The same type of analysis can be conducted to see if there is a preference for an image as well.

Knowing this, we can see from Fig. 10 that there is not one trend that describes all of the observers’ naturalness ratings, as all rating shifts are highly different from each other and from the average group rating. This demonstrates the impact of subjectivity on naturalness perception as defined in Section 1. It suggests that observers perceived the naturalness of the samples differently, likely relating naturalness to their individual preconceived notions of the materials and scenes depicted in the samples.

The same can be said for images as well, as we can see in Fig. 11 that there are multiple curve structures across the different sets, indicating variability in image preferences as well. This difference in rating shifts does not appear to be separated by material category because for example, the Reddish Wood set was rated higher than average across most algorithms, but the Natural Wood set was rated lower than average. Consequently, the behavior of these rating shifts for both observers and images appears to be too subjective to be explained by a single trend across all algorithms. But we can still see one defining factor: the Gray DM algorithm is

consistently rated the highest and the Canny Edge is consistently rated the lowest, with the ForeFront and 2bit_TAED algorithms switching between the second and the third most preferred algorithm for naturalness reproduction. Despite the inherent subjectivity of human ratings, this trend suggests a general inclination to favor the best-performing and worst-performing algorithms across observers and images.

Therefore, to understand why observers preferred an algorithm or an image over others, we rely on the audio recordings of sessions to understand each observer’s thought process when judging a sample’s naturalness and which appearance attributes influenced their categorization the most.

Analysis of the audio recordings revealed that participants used different definitions of naturalness, equating it to the beauty of the sample or to realism. This might be a point of contention because some observers would be basically rating a different attribute compared to others because realism, beauty, and naturalness are subjective concepts. Although observers defined naturalness differently, they primarily used the same descriptors, such as roughness (sometimes referred to as graininess), sharpness, and color to assess a sample’s naturalness. This indicates a common baseline among observers when rating the samples even if expressed using different terminologies.

With this, we proceeded to identify the relevant appearance attributes using the description given by the

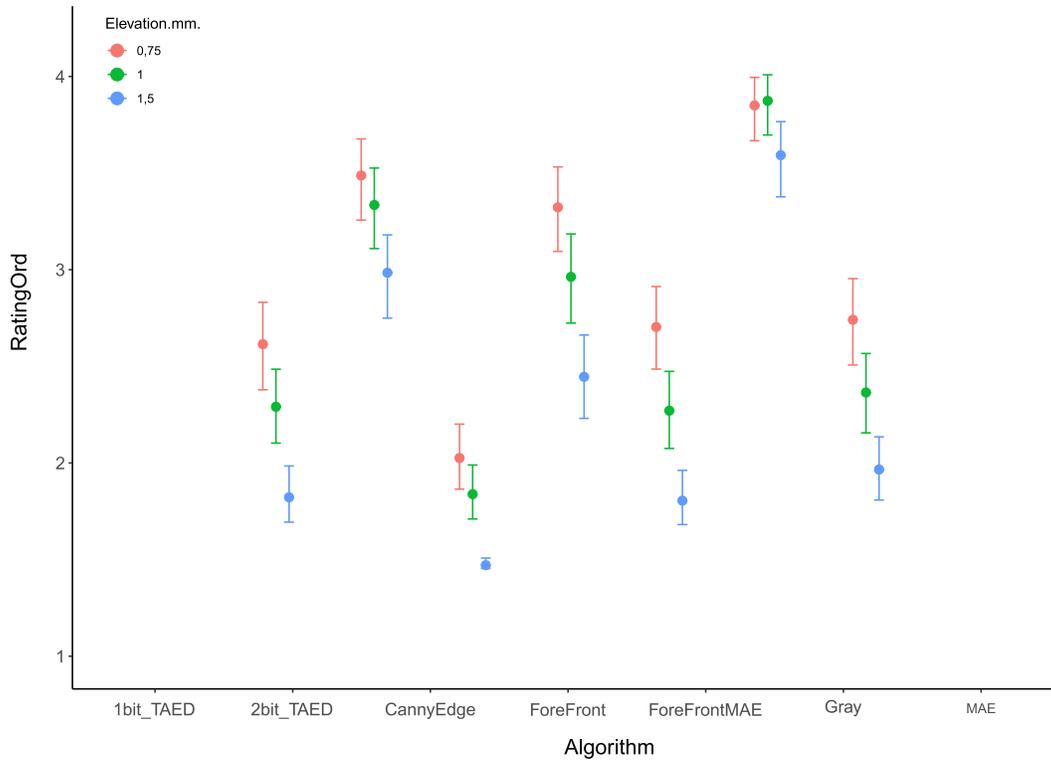


Figure 9. Grouped subjective naturalness ratings (RatingOrd) for all observers across different DM algorithms and elevation levels. The bars represent the posterior means of the naturalness ratings derived from the best-performing CLMM (IDNest). The error bars represent the 95% credible intervals, indicating the uncertainty associated with the estimates. The figure illustrates the overall preference for certain algorithms and elevations. Notably, the Gray DM algorithm consistently received the highest naturalness ratings while the Canny Edge algorithm received the lowest. Lower elevations (0.75 mm) were generally preferred across most algorithms except for the Gray DM, where the 1 mm elevation was slightly favored. The figure also reveals that the effect of elevation on perceived naturalness varied across algorithms. For instance, the ForeFront algorithm exhibited a more pronounced decrease in naturalness ratings with increasing elevation compared to the Gray DM or 2bit_TAED.

observers for each sample. Color was the first attribute assessed by nearly all the observers (similar to the findings of Kadyrova et al. [15]). This was deduced from the description of the sample as “looking normal” or “looking regular” or “having nothing weird” when first receiving the samples. It is important to note that color is theoretically not a variable during the sample creation process because all samples of the same set were produced with the same color input and with the surface profile as the only variable. Consequently, we anticipated that color may not be the primary determinant in naturalness perception. Indeed, for the majority of observers (12 out of 15), the defining factor was the surface finish (the physical texture reproduction) and its effect on the visual texture (color + gloss) as illustrated in Fig. 5. However, we can observe that even with the same color input for all samples, we get varying color outputs for each sample, which explains the reliance of the observers on color to rate a sample’s naturalness.

Overall, observers deemed natural surface finish as follows: smooth, with minimal roughness, having clear and smooth details and edges, having physical texture matching or enhancing the visual texture (primarily gloss), and having the correct color (based on material comprehension) with some color uniformity preferably. In simpler terms, smooth surfaces and clear physical textures that align with the

scene’s visual texture were deemed natural. Conversely, what was deemed unnatural was the opposite: rough (also described as noisy/granular) samples where the roughness overshadowed the visual texture either by creating unrealistic light reflections and/or by altering the color tint of the sample. For example, in Fig. 5, we can observe that rough samples have a yellowish tint that was deemed unnatural by most observers. Another significant factor contributing to the unnatural appearance of samples was edge and color sharpness. In essence, samples were deemed unnatural due to roughness, color, and edge sharpness and a discrepancy between visual and physical textures. However, this is a generalized view over all the observers, and there might be some preferences for attributes over others, thus explaining the individual differences seen in Fig. 10. For example, some observers might prefer a rougher or a sharper sample, leading them to rate ForeFront-MAE higher than other observers. But generally, these individual preferences tend to be overridden when taking the whole group into account.

With this understanding of appearance preference, we can evaluate the statistical model to interpret the results obtained for each of the algorithms. Referring to Figs. 2 and 9, the analysis follows.

DM1 has performed the best. This is because it produces a full range grayscale image ($2^8 = 255$ levels), and the surface

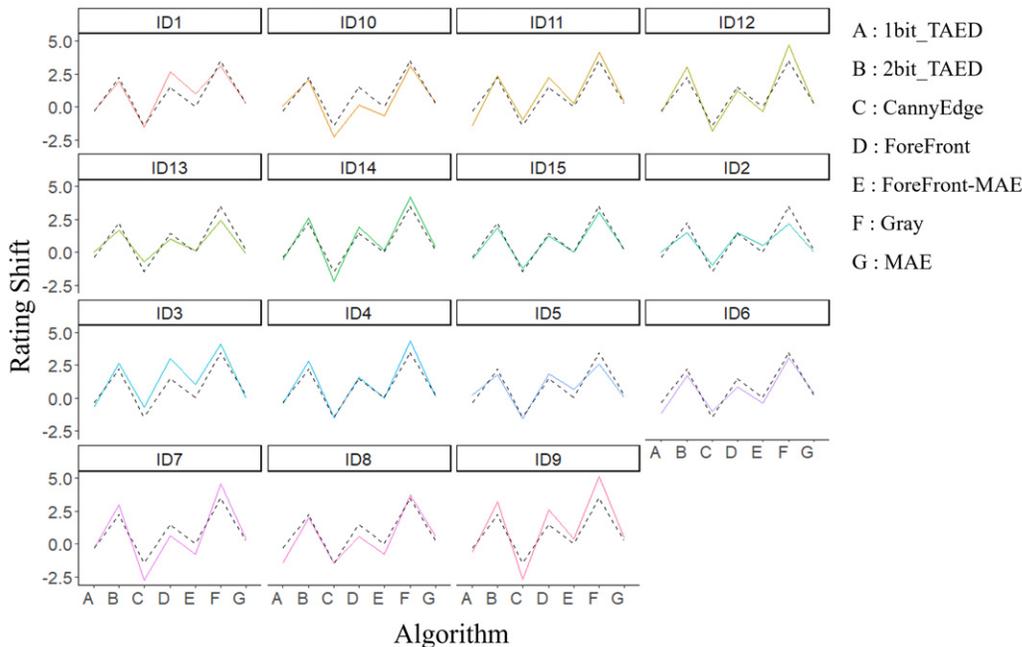


Figure 10. Individual observer rating shifts for each DM algorithm derived from the IDNest model. Each panel represents a single observer (ID1–ID15), and the dashed black line indicates the group average rating for each algorithm. The y-axis represents the rating shift, which is the difference between an individual observer’s rating for a given algorithm and the group average rating for that algorithm. A curve above the dashed line indicates that the observer tended to rate samples made with that algorithm higher than the average while a curve below the line indicates lower ratings. The figure reveals substantial interobserver variability in rating patterns, highlighting the subjective nature of naturalness perception. For example, Observer ID3 consistently rated most algorithms higher than the group average while Observer ID10 generally rated them lower. These individual differences underscore the importance of considering observer variability in subjective image quality assessment.

change between neighboring pixels is smooth resulting in smooth surfaces and edges and a physical texture that is coherent with the visual texture. Although DM1 was evaluated as the best in this particular study, we may have many cases where it would not be the best option to reproduce the surface. As Eq. (1) shows, grayscale values depend on the color of the RGB reference image. Therefore, in cases where the forefront of an image is dark and the background is bright, an RGB-to-grayscale transformation alone would not suffice because the placement would be inverted, thus requiring an additional step to correct the placement of objects in the image. However, the surface finish should be smooth enough to be perceived as natural regardless of the placement of the foreground and background. It would be an interesting idea to test out this particularity in future work.

DM7 has performed the worst. This is because it outputs a binary DM that when applied produces very rough and granular surfaces. This excessive roughness hides the details of the samples, makes them unpleasant to touch and to look at, and creates incoherent physical and visual textures.

DM2 and **DM5** perform relatively poorly compared to other algorithms with increase in elevation. This is because ForeFront Algorithms exhibit a drastic change in pixel value in certain parts of the DM, resulting in sharper samples produced compared to the others. That is also why they perform much worse at higher elevations because the higher the elevation, the sharper the edges become. Consequently, ForeFront DMs appear unnatural at higher

elevations. ForeFront-MAE has the added drawback of being rough, which explains its poorer performance compared to regular ForeFront DMs. However, these DMs reproduce physical texture that is consistent with visual texture, which is why they are rated higher than Canny Edge DMs.

DM3 and **DM4** perform very poorly overall, particularly at high elevations. Similar to DM7, they generate very granular surfaces that obscure details and are generally perceived as unpleasant. However, when we go from 1bit_TAED to 2bit_TAED, where the grayscale range expands from $2^1 = 2$ to $2^2 = 4$, the average score goes from 2.11 to 3.54, which is an increase of 1.43 score points. This corresponds to a shift in category from “unnatural” to “adequate.” This highlights the crucial role of sample surface and edge smoothness in determining perceived naturalness.

This assessment of appearance preferences also explains the lack of preference of any image set by the observers. By relating the rating shift to the ease of reproducing smooth edges and surfaces from the reference images, we can likely explain how observers rated the naturalness of the different image sets. For instance, comparing the Reddish Wood reference image with the Textured Stone Wall image (see Fig. 1), we can see that Reddish Wood has a very uniform surface and color while it is the opposite for Textured Stone Wall. Therefore, it is much easier to achieve smoother surfaces and color uniformity in the case of Reddish Wood than with Textured Stone Wall, where the edges between sections of the image are already highly pronounced. As can

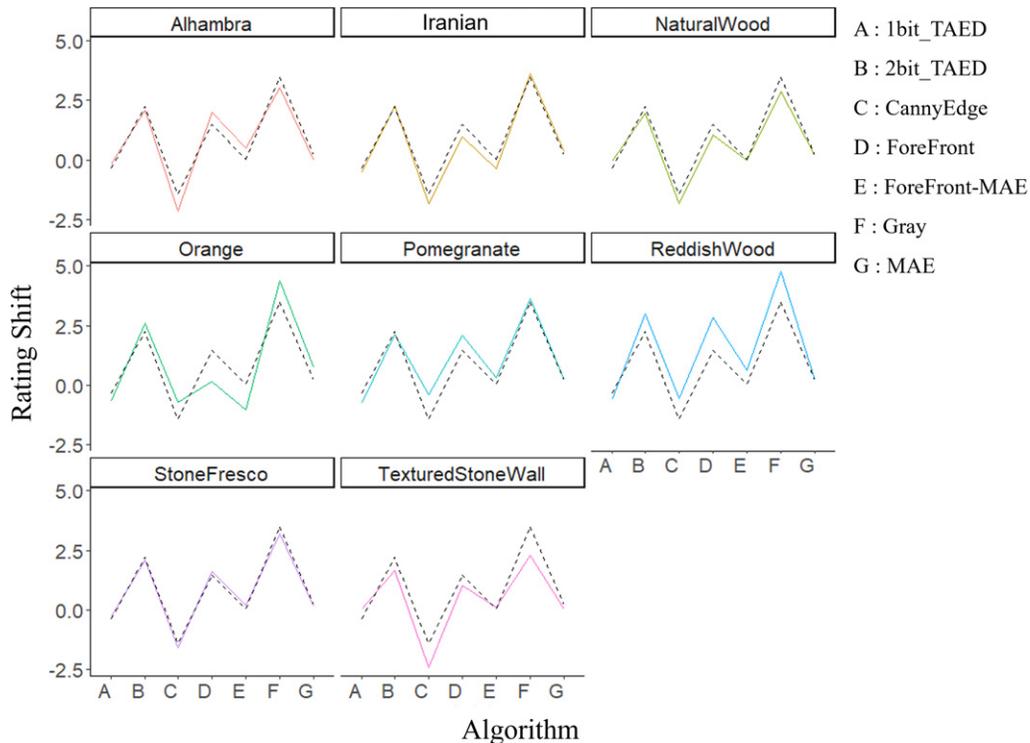


Figure 11. Image-specific rating shifts for each DM algorithm derived from the ImageNest model. Each panel represents a different reference image, and the dashed black line indicates the group average rating for each algorithm. The y-axis represents the rating shift, which is the difference between the average rating for a specific image and a given algorithm and the group average rating for that algorithm. A curve above the dashed line indicates that the image tended to be rated higher than the average for that algorithm while a curve below the line indicates lower ratings. The figure reveals that the perceived naturalness of the 3D printed samples was also influenced by the specific image being reproduced. For instance, the Reddish Wood image was generally rated higher than average across most algorithms while the Natural Wood image was rated lower. These image-specific variations suggest that certain image characteristics might interact with the DM algorithms to affect perceived naturalness. The diverse patterns observed across images highlight the complexity of naturalness perception and the need to consider both individual and image-specific factors.

be seen in Fig. 11, Reddish Wood samples are therefore rated more natural overall while Textured Stone Wall samples tend to be rated as unnatural.

6. CONCLUSION

Eight images were 3D-printed using seven different DMs at three different evaluation levels to assess the naturalness perception of 3D printed samples. Each DM highlighted a different feature of the scene, thereby affecting the naturalness perception of the 3D printed samples. The naturalness of 3D prints was evaluated subjectively via a psychometric experiment. Results from this experiment were analyzed using a Bayesian statistical model to determine the general and individual preferences of the observers. Observers evaluated the 3D printed samples as natural when the said samples had low elevation, smooth surfaces, and edges. Conversely, samples were deemed unnatural when exhibiting high elevation, granularity, and sharp edges. These findings confirm our original hypotheses concerning the impact of elevation and texture extraction algorithms on naturalness perception as well as the subjective nature of this perception. However, the consensus among observers when their results were aggregated indicates that despite the subjective nature of naturalness perception, observers

generally agree on the key characteristics that contribute to the natural appearance of a 3D printed sample.

To further understand naturalness perception, finding a relationship between subjective naturalness scores and objective image quality metrics is an interesting research direction. Additionally, exploring different texture extraction algorithms and applying them at varying elevation levels is an interesting avenue for future research.

ACKNOWLEDGMENT

S.H.D.P. is supported by the project “Quality and Content: Understanding the Influence of Content on Subjective and Objective Image Quality Assessment” (Grant Number 324663; approved on October 1, 2021) from the Research Council of Norway.

REFERENCES

- 1 T. Baar, S. Samadzadegan, H. Brettel, P. Urban, and M. V. O. Segovia, “Printing gloss effects in a 2.5 D system,” *Proc. SPIE* **9018**, 160–167 (2014).
- 2 *Blender 3D Rendering Software*, <https://www.blender.org/>. Accessed: 2023-12-29.
- 3 D. Bradley and G. Roth, “Adaptive thresholding using the integral image,” *J. Graph. Tools* **12**, 13–21 (2007).

- ⁴ P.-C. Bürkner, “BRMS: An R package for Bayesian multilevel models using Stan,” *J. Statist. Software* **80**, 1–28 (2017).
- ⁵ P. Bürkner and M. Vuorre, “Ordinal regression models in psychology: A tutorial,” *Adv. Methods Pract. Psychol. Sci.* **2**, 77–101 (2019).
- ⁶ T. Campbell, C. Williams, O. Ivanova, and B. Garrett, “Could 3D printing change the world,” *Technologies, Potential, and Implications of Additive Manufacturing* (Atlantic Council, Washington, DC, 2011), Vol. 3, p. 18.
- ⁷ J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* **8**, 679–698 (1986).
- ⁸ R. W. Floyd and L. Steinberg, “An adaptive algorithm for spatial grey-scale,” *Proc. Soc. Inform. Display* **17**, 75–77 (1976).
- ⁹ H. Fossheim, Å. Aakra, M. Aschan, F. E. Benth, S. G. Carson, S. Gismervik, A. Goksøyr, H. P. Gulbrandsen, G. I. Johansen, J. Paus, S. B. Rønning, J. Tørresen, E. S. Vigtel, and T. Østerhaug, *Guidelines for Research Ethics in Science and Technology*, 3rd ed. (National Research Ethics Committees, 2024), ISBN: 978-82-7682-118-5.
- ¹⁰ S. Groot, “Effect of perceived risks, naturalness, usefulness and ease of use on the consumer acceptance of 3D food printing,” (Wageningen UR, 2018).
- ¹¹ International Telecommunication Union, “Subjective video quality assessment methods for multimedia applications,” *Recommendation ITU-T 910* (2023).
- ¹² T. J. W. M. Janssen and F. J. J. Blommaert, “Predicting the usefulness and naturalness of color reproductions,” *J. Imag. Sci. Technol.* **44**, 93–104 (2000).
- ¹³ J. F. Jarvis, C. N. Judice, and W. H. Ninke, “A survey of techniques for the display of continuous tone pictures on bilevel displays,” *Comput. Graph. Image Process.* **5**, 13–40 (1976).
- ¹⁴ A. Kadyrova, V. Kitanovski, and M. Pedersen, “A study on attributes for 2.5D print quality assessment,” *Proc. IS&T Twenty-Eighth Color and Imaging Conference* (IS&T, Springfield, VA, 2020), Vol. 2020, pp. 19–24.
- ¹⁵ A. Kadyrova, M. Pedersen, and S. Westland, “Effect of elevation and surface roughness on naturalness perception of 2.5D decor prints,” *Materials* **15** (2022) ISSN: 1996-1944.
- ¹⁶ A. Kadyrova, M. Pedersen, and S. Westland, “What elevation makes 2.5D prints perceptually natural?,” *Materials* **15** (2022) ISSN: 1996-1944.
- ¹⁷ D. Li, T. Kiyotomo, T. Horiuchi, M. Tanaka, and K. Shigeta, “Texture-aware error diffusion algorithm for multi-level digital halftoning,” *J. Imag. Sci. Technol.* **64**, 50410–50411 (2020).
- ¹⁸ T. D. Ngo, A. Kashani, G. Imbalzano, K. T. Nguyen, and D. Hui, “Additive manufacturing (3D printing): A review of materials, methods, applications and challenges,” *Composit. Part B: Eng.* **143**, 172–196 (2018).
- ¹⁹ K. E. Overvliet and S. Soto-Faraco, “I can’t believe this isn’t wood! An investigation in the perception of naturalness,” *Acta Psychol.* **136**, 95–111 (2011) ISSN: 0001-6918.
- ²⁰ C. Parraman and M. V. O. Segovia, *2.5 D Printing: Bridging the Gap between 2D and 3D Applications* (John Wiley & Sons, Hoboken, NJ, 2018).
- ²¹ “Pixabay free internet image database,” <https://pixabay.com/>. Accessed: 2023-12-29.
- ²² J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int. J. Comput. Vis.* **40**, 49–70 (2000).
- ²³ J. Reinhard, P. Urban, S. Bell, D. Carpenter, and M. Sagoo, “Automatic data-driven design and 3D printing of custom ocular prostheses,” *Nat. Commun.* **15** (2024).
- ²⁴ J.-N. Shiau and Z. Fan, “Set of easily implementable coefficients in error diffusion with reduced worm artifacts,” *Proc. SPIE* **2658**, 222–225 (1996).
- ²⁵ “Stratasys J55 Color Poly-Jet Printer,” <https://www.stratasys.com/en/3d-printers/printer-catalog/polyjet/j55-prime/>. Accessed: 2024-04-22.
- ²⁶ “Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios,” in *Int’l. Radio Consultative Committee* (International Telecommunication Union (ITU-R)), Recommendation BT.601-7, Switzerland, CCIR Rep (03/2011).
- ²⁷ L. Szirmay-Kalos and T. Umenhoffer, “Displacement mapping on the GPU—State of the art,” *Comput. Graph. Forum* **27**, 1567–1592 (2008) Wiley Online Library.
- ²⁸ J. E. Taylor, G. A. Rousselet, C. Scheepers, and S. C. Sereno, “Rating norms should be calculated from cumulative link mixed effects models,” *Behav. Res. Methods* **55**, 2175–2196 (2023).
- ²⁹ C. Tymms, S. Wang, and D. Zorin, “Appearance-preserving tactile optimization,” *ACM Trans. Graph. (TOG)* **39**, 1–16 (2020).
- ³⁰ C. Tymms, D. Zorin, and E. P. Gardner, “Tactile perception of the roughness of 3D-printed textures,” *J. Neurophysiol.* **119**, 862–876 (2018).
- ³¹ P. Urban, “Graphical 3D printing: Challenges, solutions and applications,” *Proc. IS&T CIC28: Twenty-Eighth Color and Imaging Conf.* (IS&T, Springfield, VA, 2020), Vol. 2020, pp. 87–90.
- ³² K. Vanhoey, B. Sauvage, P. Kraemer, and G. Lavoué, “Visual quality assessment of 3D models: On the influence of light-material interaction,” *ACM Trans. Appl. Percept. (TAP)* **15**, 1–18 (2017).
- ³³ “VeroFamily J55 PolyJet Printer Ink,” <https://support.stratasys.com/en/Materials/PolyJet/Vero-Family>. Accessed: 2024-04-22.
- ³⁴ T. Virtanen, M. Nuutinen, and J. Häkkinen, “Image quality wheel,” *J. Electron. Imag.* **28**, 013015 (2019).
- ³⁵ J. Wang, J. Allebach, and M. Ortiz-Segovia, “Semantically-based 2.5D texture printing,” *2014 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2014), pp. 2634–2638.
- ³⁶ S. Watanabe and M. Opper, “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory,” *J. Mach. Learn. Res.* **11** (2010).
- ³⁷ K. B. Watson, “Categorical data analysis,” in *Encyclopedia of Quality of Life and Well-Being Research*, edited by A. C. Michalos (Springer Netherlands, Dordrecht, 2014), pp. 601–604, ISBN: 978-94-007-0753-5.
- ³⁸ J. Yuan, G. Chen, H. Li, H. Prautzsch, and K. Xiao, “Accurate and computational: A review of color reproduction in full-color 3D printing,” *Mater. Des.* **209**, 109943 (2021).