Effects of Mesoscale Surface Structure on Perceived Brightness

Michael Ludwig, Gary Meyer; University of Minnesota

Abstract

Surface geometry can play an important role in our ability to understand and interpret material appearance and properties. This property ranges from large-scale shape changes impacting our identification of reflections to visible surface roughness affecting how glossy a material appears. In this work we present a user study that examines numerous surface geometries that are defined at the mesoscale: small enough to be considered indicative of the material and not object geometry, but large enough to be visible from a distance with the naked eye. Subjects matched the perceived brightness of a ray-traced bumpy surface to a flat surface with adjustable intensity. Multiple classes of bumpy surface were generated and presented to subjects so that the effects of surface pattern on perceived brightness could be studied. We show that two predictive models of brightness are only conditionally accurate but that humans have a consistent means of measuring overall brightness.

Introduction

In this work we begin to examine numerous surface geometry patterns and how subjects perceive their brightness from different viewing and lighting directions. The surface structures are generated at the mesoscale, a scale above microscopic roughness profiles that are well described by statistical models used in BRDFs and below macroscopic that fundamentally influences the overall shape of an object. Mesoscale surface patterns are interesting because they provide small, but visible, cues as to lighting and shading while still frequently being identified as part of a material (e.g. the ridges of stucco, coarseness of brick, and grains of certain woods). The patterns may be random, produced naturally, or follow very regular forms based on artistic design decisions. This last point is important because humans can manufacture mesoscale patterns with much more ease than new paints or coatings and are frequently employed in product design to create signature looks or improved haptic feel while maintaining appearance standards. Indeed our work is inspired by our interactions with the automotive industry where they design surface patterns at this scale for the interior of vehicles.

The mesoscale is somewhat subjective and dependent on the distance to the viewer. For example when examined up close, concrete and other building materials can have visible non-smooth surfaces but at a distance it can be accurately described as a plane with a reflectance model such as Oren-Nayar's BRDF [10] due to our visual system's finite resolution. A good rule of thumb that was applied to the surfaces generated in this study is that the mesoscale is just large enough to produce visible patterns from shading but not so large that it drastically changes the silhouette of the object. Unfortunately, the parameter space for surfaces patterns at this scale is nigh infinite. To constrain the scope of this study, we have chosen to evaluate perceived brightness of these surfaces. There are many aspects of appearance, ranging from brightness, color, glossiness, texture, apparent tactile roughness, and material makeup that can be inferred by the visual system [6]. We begin with brightness for its relative simplicity and it will act as a good foundation before moving to more advanced aspects. Additionally, we restrict the surface geometries to a set of parameterizable generators that produce families of distinct patterns.

The preliminary results of our work into this area are presented in this paper. The next section describes background work examining brightness and lightness perception in humans, material perception, and other related work from the psychophysics and computer graphics communities. The next section describes the experimental setup for a user study conducted as well as the process for generating numerous parameterized surfaces at the mesoscale. Our aim is to sparsely sample the space of surface patterns and identify how brightness is interpreted over a spatially varying bumpy material. Analysis and data from two user studies is shown in subsequent section. Lastly, we conclude with a discussion on future directions and implications for this research area.

Background

Brightness has often been studied alongside lightness. We continue the common distinction of brightness being the perceived luminance of an object and lightness being the apparent reflectance of the object. We are only concerned with brightness as subjects are aware that the surfaces under question are made of a uniform material. These properties have been studied in very synthetic scenarios with 2D elements arranged as concentric annuli [16], complex rectangular patterns [1], and designed to evoke depth relations [15]. Additionally, past research has focused on flat or smooth patterns when judging brightness and this work is the first to our knowledge to approach the spatially-varying problem.

We are interested in exploring brightness when the stimuli is a much more physically accurate simulation of a 3D surface. Research has shown that realistic lighting can have an impact on subjects' abilities to identify gloss [7]. Other work has shown that perceived shape and depth can affect our interpretation of color [2]. Given this, it is necessary to include the supporting scenery that assists in identifying lighting and shading from texture.

Elements of this study have been inspired by the computer graphics and psychophysics experiments into the perception of glossiness. This body of work has demonstrated the effectiveness of user studies comparing and matching rendered images [5], and the impact that surface geometry variations can have on the perception of surface properties [11, 8, 13]. Although the discovery that mesoscale surface structure can impact how humans perceive glossiness is significant, the geometry patterns have not been analyzed in a systematic way. Several studies that have approached the subject have chosen arbitrary and distinct structures

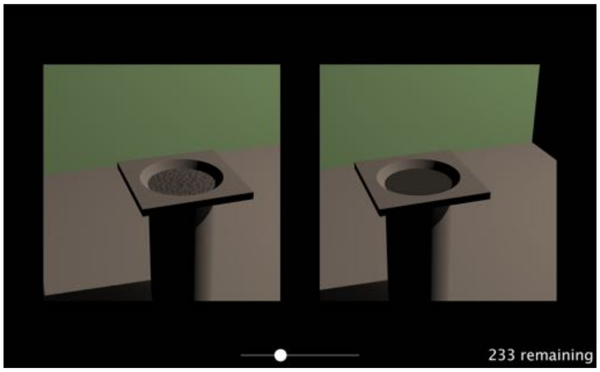


Figure 1. *The experiment user interface presented to subjects. The slider along the bottom is controlled by mouse movement and adjusts the brightness of the right pedestal.*

and patterns (1/f noise or overlapping bubbles) leading to difficulties comparing their results. Additionally one can argue that given the complexity of introducing visible surface variations into a stimuli, a property such as gloss is perhaps too far-reaching before understanding simpler traits.

To that end, we report on a study designed to evaluate how humans interpret the brightness of surfaces with visible roughness or structure that is still small enough to be considered part of appearance and not geometry (the mesoscale). We target brightness as it is one of the most primitive of perceived quantities when viewing a surface. We examine numerous surface patterns to identify commonalities in how humans interpret surface geometry at this scale. A goal is to provide insight into future studies that consider perceived spatially varying appearances, as well as guidelines for interpreting past research involving mesoscale surface patterns.

Experiment

Our experiments were designed to serve two purposes; first, to see how consistently people judge the brightness of a surface with mesoscale patterns of shading and second, to see if there are trends across varieties of surface geometries. During the experiment, a subject is presented with two copies of a scene side-by-side, as shown in Figure 1. On the left presentation, the disc in the middle of the pillar displays a ray-traced surface pattern. The right presentation's disc's brightness is controlled by a slider. Subjects were tasked with adjusting the brightness until it best matched the overall brightness of the complex surface. A matching adjustment task was chosen to avoid the pair-wise explosion that would occur if subjects were to compare patterns. Given the size of surface pattern space considered, a comparison task is intractable.

Once matched, as reported by the subject, the screen was cleared briefly before advancing to another trial with a different scene configuration. The matching task was time-limited to 15 seconds. If this time was exceeded the trial was advanced automatically. This short time period encouraged measuring brightness of the overall pattern. When longer or unlimited periods of time were given during preliminary studies, subjects frequently tried to match the brightness of only the surfaces not in shadow. Periodically subjects were given a short break to relieve fatigue. Prior to the experiment, all subjects were given a demonstration of mesoscale surfaces in the real world using a molded plastic plaque from an automotive company and then trained with the user interface.

Training consisted of performing the same adjustment task, but on surface patterns not included in the actual study. A window of reasonable values was selected by the authors, and visual feedback was provided if the subject's matching attempt fell outside of the window. All subjects consistently fell within the acceptable window by the end of five training trials, many even on their first trial. Several had issues at the very beginning of training while they learned the 15-second time window.

Two user studies were performed; the studies were identical except for the selected stimuli as described in Section "Scene Selection for Subjects". Each study had twelve participants with no subject participating in both studies. Both studies had 8 female and 4 male subjects each. Subjects ranged in age from 18 to mid-60's with normal or corrected-to-normal vision and no reported visual impairments. The study was conducted in a darkened room on a single 24-inch Dell monitor in full screen with a resolution of 1920×1080; the monitor was viewed at a distance of 60cm. The stimuli, as described below, was tonemapped to the display using Reinhard's photographic operator [14].

We next describe our process for producing numerous parameterizable mesoscale surfaces, then a discussion of the overall stimulus presented to subjects, followed by our strategy for sparsely sampling the large number of scenes we produced.

Mesoscale Surface Generation

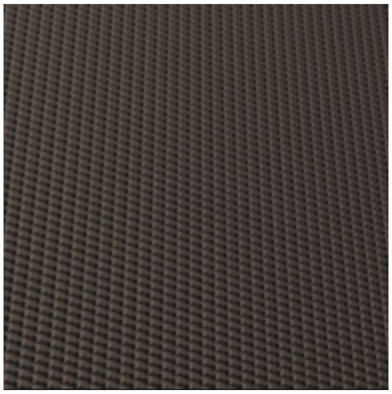
Surface patterns were automatically generated using a variety of processes to achieve a spread of patterns that ranged from completely regular patterns (that might be manufactured) to randomized, more natural patterns. This range includes the classes of surface examined by past studies on the perception of gloss. Each surface generator was parameterized so that many variations could be produced while still having a cohesive structure. The patterns created are described in Figure 2. The patterns shown in Figure 2e and Figure 2f feature irregular noise generated using Perlin noise [12].

Overall, 136 total surface patterns were generated by varying their available parameters to get a range of stipple sizes, elevations, etc. The size of elements within the surfaces ranged from 1mm to 8mm, which given the scene stimuli parameters, covered the characteristic sizes of mesosurfaces.

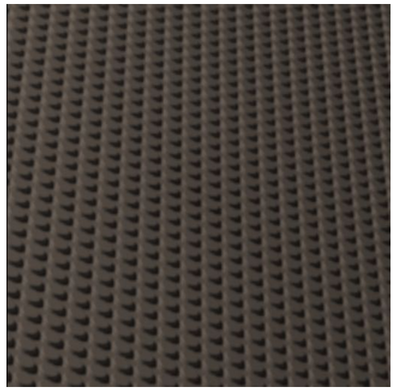
Stimuli Design

Every generated surface pattern was rendered from sixteen view points and sixteen lighting directions, for a total of $256 \times 136 = 34,816$ images. Surface patterns were lit and viewed from multiple directions so that any view or light direction dependence on the perceived brightness could be detected. The height elevation of each surface was applied to a plane. This was chosen over a more complex macroscale geometry to remove any confounding factors caused by the shading gradients of the macro surface. The view and lighting positions were distributed evenly over the hemisphere.

These images are presented in a relatively complex scene to provide improved depth and lighting cues. This helps remove inversions in the interpretation of the bump patterns and misinterpretations of the shadowing pattern as an albedo texture. Fig-



(a) Rectangular stipples, with parameterized size, spacing, and depth.



(b) Dotted stipples, with parameterized radius, spacing, bubbliness, and depth.

(c) Randomized rectangular stipples, as rectangular stipples but with random variation per stipple.

(d) Randomized dotted stipples, as dotted but with random variation per stipple.

(e) Thresholded Perlin noise to evoke semi-natural ridges.

(f) Perlin noise for a natural roughness.

Figure 2. *Surface pattern classes generated as part of this study into mesoscale surface appearance.*

ure 3 displays the scene the surfaces are placed in. The pedestal provides shading gradients and casts a strong shadow to help the subject infer the light direction. The pedestal stands in the center of a room with four differently colored walls, which alleviates the sense of the object floating in space and potentially helps the subject track where they are viewing from in each trial. The walls and floor are modeled with a perfectly diffuse material, while the mesoscale surface is a plastic material modeled with the GGX distribution [17] and parameters chosen to be similar to plastic sample plaques we have studied from industrial designers.

The light within the room is a 5500K temperature sphere approximately the size of a light bulb and is placed according to the trial's lighting condition. This relatively simple lighting scenario allows changing the direction of the light to have a meaningful impact on the surface appearance while remaining a reasonable real-world configuration (room with bare light bulb). Although there is evidence that real-world, complex environments help perceive glossiness of a material, because the chosen material of our sample is not significantly specular this is less critical. By using a simpler light, we are able to measure a good baseline before advancing to more complex lighting scenarios in future work.

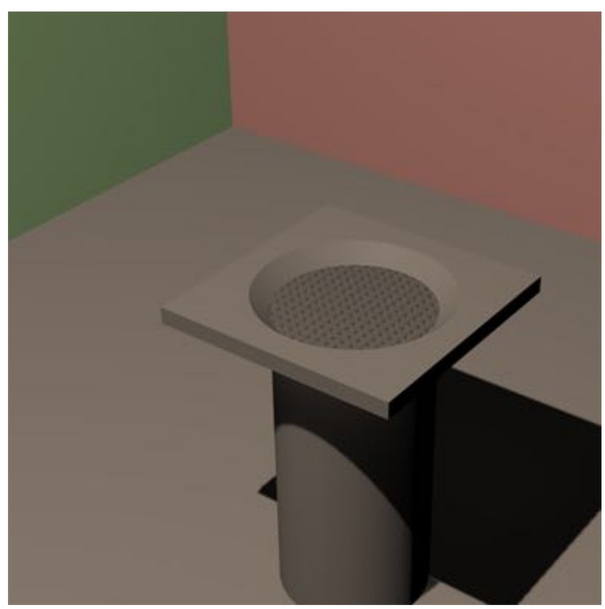


Figure 3. *Closeup of stimuli scene presented to subject, demonstrating shading, strong cast shadows, and walls for context.*

Scene Selection for Subjects

Even with the constraints imposed on pattern generation and limiting the scenes to sixteen views and lighting conditions, our dataset of rendered mesoscale surfaces consists of 34,816 images. This is far too many to present to subjects in a reasonable time frame. Instead we opt to do a semi-randomized sparse sampling of the surface patterns to maximize the number of geometries seen while ensuring reasonable repetition across subjects. To that end, the two user studies conducted used different scene selection criteria while otherwise following the exact same experimental procedure described previously.

In the first study, each subject was assigned a random subset of the generated surface patterns. For each assigned pattern a random view or light was chosen and the sixteen images matching that condition are included in the trial set for the subject. Additionally, a noise-patterned surface (Figure 2f) was evaluated in the fixed-view and fixed-lighting conditions by every subject, for an additional 32 trials per subject. The noise pattern was chosen for viewing by every subject because it has frequently been used in glossiness perception studies. All selected trials for a subject are shuffled together to avoid ordering adaptation. The shuffled block of trials was repeated three times to get repeated measures to test whether subjects were significantly changing their responses over time, and to form a better estimate of their matched brightness. This first study captures data a single surface viewed by many subjects, as well as a sparse sampling of other surfaces viewed by a single subject, all from multiple view and lighting directions.

The second study's selection criteria was designed to complement the data acquired from the first study. Half of the trials considered by a subject in the second study were chosen from conditions previously seen by only a single subject. These conditions were drawn randomly but were weighted towards view and light poses that had a higher variance across all occurrences. Preliminary analysis of the data showed that this higher variance within and between subjects' measured brightness occurred when the light was oriented away from the normal of the stimuli plane and when the viewing direction approached specular. The second half of trials for a subject relied on the same variance-based sampling to choose a view and light pose but the surface pattern was drawn from the set of patterns not previously seen in the first study. Like before all trials were repeated three times and shuffled. Unlike the first study, each selected set of trials was presented to multiple subjects. This second study provides additional data to validate the responses from the first study's subjects and broadens the number of viewed surface patterns.

In the next section we present and discuss the data gathered from these two studies.

Results

The goal of this experiment was to determine if the mesoscale surface pattern has a significant impact on our brightness judgments of the object. A very related goal is developing models that can predict any identified differences between the surface patterns. Two proposed models will be described below. In some regards, finding that there are no substantive differences implies that pattern has little effect on perceived brightness; such a conclusion would mean work on more complex perceptual attributes could justifiably use a limited range of mesoscale patterns. However, if differences are found then it substantiates the doubts we brought up in the background section on the broader applicability of past mesoscale perceptual research.

Throughout this section we will at times consider two models of how subjects are estimating overall brightness. Here a model refers to a function that predicts the perceived brightness given the stimuli and its scene parameters. The first model estimates brightness by assuming luminance can be calculated as if the surface were perfectly flat. This is a reasonable hypothesis for mesoscale surface patterns that feature many small flat elements broken up by other structures; if the flat regions dominate a subject's perception it will produce a value similar to when the whole surface is smooth. The second model estimates brightness by averaging the luminance of the rendered stimuli image. This is also reasonable because it approximately models what subjects would see if they were viewing the stimuli from too far away to make out mesoscale details.

Before we can analyze these two models, the collected data must be processed to determine the reliability and consistency of the subjects. The studies were designed to have redundancy within a subject's responses and across multiple subjects. The next section presents our analysis confirming that subjects consistently measured brightness over multiple presentations. The subsequent section analyzes the variability between subjects responses. After subject reliability is evaluated we will present our approach to dimensionality reduction that allows us to make broader conclusions about mesoscale surfaces and perceived brightness.

The following analysis will consider collected data in three broad categories. The first subset are the responses to the trials shared by all subjects in the first study. The surface pattern shown in these responses was a rough noise surface of the class described by Figure 2f. Subjects evaluated the surface for all sixteen light directions (with a single fixed view) and all sixteen viewing directions (with a single fixed light) for a total of 31 poses[1]. This subset of the data is evaluated by the largest group of distinct subjects. The second subset contains the first and are the responses to all trials presented to at least three subjects. The last subset contains all other responses from subjects.

Within Subjects

To test whether or not subjects' responses changed over time from the repeated measures, an rANOVA–a one-way ANOVA grouped by the repeated measures–was performed for the 31 shared scenarios viewed by 12 subjects. The minimum p-value is 0.104 and the maximum value is 0.997. After completing the second user study, we extended the rANOVA analysis to all trials that had been seen by at least three subjects (from the pool of 24 subjects). This amounted to 369 unique surface pattern and light/view combinations. Only 16 of these had rANOVA p-values less than 0.05 but since they were viewed by only a few subjects it is likely noise from outliers. The median p-value over these 369 scenarios is 0.4833. Given this, we cannot reject the null hypothesis that subjects' responses are unchanged over time. In the remaining analysis we average the repeated responses for each subject to get a better estimate of their perceived brightness.

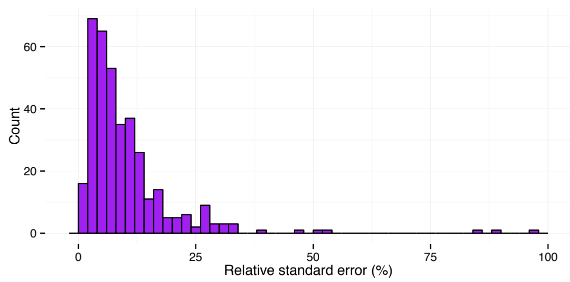


Figure 4. *Histogram of the relative standard error between subject responses.*

Between Subjects

Relative standard error (as a percentage) is used to quantify the consistency between subjects. Only the 369 trial scenarios viewed by at least three subjects were considered. Relative standard error is used so that errors can be compared across the different scenarios. The error was calculated over each scenario's subject responses, after averaging over each subject's repeated measures. The distribution of error is shown in Figure 4 and is heavily skewed towards the lower end, with a peak around 5%. This is a strong indication that subjects evaluate brightness in a consistent manner. Anecdotally it was reported that more trouble was had evaluating surfaces that presented a mixture of very bright highlights combined with dark shadows. This is verified by the increased variance in responses for scenes at specular with a glancing light angle.

Relative Light and View Poses

During scene generation, each surface pattern was rendered under 256 light and view poses. However, if the light and viewing direction are described by $N\dot{H}$ and $N\dot{L}$ terms there are far fewer poses to consider. This idea is motivated by the frequent appearance of $N\dot{L}$ and $N\dot{H}$ in reflectance distributions, where N refers to the geometric normal (stimuli plane in this case), L refers to the direction to the light, and H is the half vector between L and the view direction. We will refer to this as the *relative pose* description.

For each surface, the scenes and subject response distributions can be grouped into equal relative pose bins. Within each bin it may not necessarily be the case that subjects perceive brightness the same. This would be because of the orientation of the surface geometry with respect to L or V that is lost when using the relative pose coordinates.

To confirm if this is the case, the Kolmogorov-Smirnov test [9] was used for each pair of subject brightness distributions within a bin. The KS test can be used to determine if two distributions are distinct for low p-values. Our approach forms a matrix of p-values for the pair-wise comparisons, with $p = 1$ along the diagonal. If any element of this matrix has $p < 0.05$ we consider that relative pose coordinate of the scene to be inconsistent. Less than 2% of these bins exhibited inconsistencies and there was no trend amongst those for a particular surface pattern or scene, so they are likely due to outliers. Thus, it is reasonable to consider this relative pose as a reduced coordinate space for evaluating brightness; the absolute orientation of surface geometry does not significantly impact brightness judgments after factoring out relative orientation effects.

[1] One light+view pose was present in both blocks.

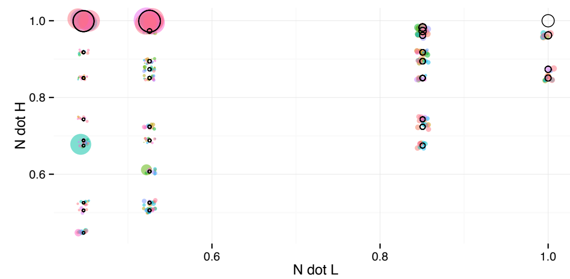


Figure 5. *Perceived brightness for each surface in relative pose space. Colors distinguish surface geometry. The black circle is the luminance predicted by the flat surface model. Note that the positions of each data point have been jittered to improve readability.*

Figure 5 shows all scenes arranged by their relative pose. The size of the points corresponds to the average perceived brightness measured for that surface and relative pose. The color of the points corresponds to the surface geometry of the scene. The data is also compared to the flat surface model, drawn as a black circle at the relative pose coordinates. From this figure it is clear that all surface patterns follow certain trends such as increased brightness at specular and are reasonably consistent with the flat surface model. However, it is hard to identify trends, similarities, and differences between surface geometries from this visualization. The next section will approach this problem.

Surface Geometry Dimensionality Reduction

The subject responses for each relative pose provide a sparse sampling of the perceived brightness in the relative pose coordinate space. We model correlation between surface patterns as the dot product between vectors containing the perceived brightness for each relative pose of the surface. This works well for surfaces that have mismatched samplings as we can effectively disregard unmatched poses from the similarity evaluation.

Evaluating this distance function for each pair of surfaces creates a distance matrix that can be used with non-metric multidimensional scaling (MDS) [4], which is a powerful way of reducing this complex space into two dimensions. The projected surface locations will respect, to the best degree possible, their distances or similarities in the higher dimensional space. Figure 6 shows the results of applying multidimensional scaling to the set of surfaces seen by at least three subjects. The six hue blocks correspond to the six pattern classes from Figure 2. The flat surface model is included in the distance matrix, where its brightness is evaluated for all necessary relative pose coordinates, and is drawn as a black point. Figure 7 shows a similar figure except the inter-surface distances were computed for all surfaces that had any subject response. In this case surfaces that had been viewed by a single subject are based solely on their perceived brightness.

In examining the two MDS plots, many similarities arise between the two. Overall, Figure 7 resembles a more ragged version of Figure 6, indicative of the added potential outliers. The flat surface model resides in a reasonably central location and the inner cluster of surface points are noise-based surface pat-

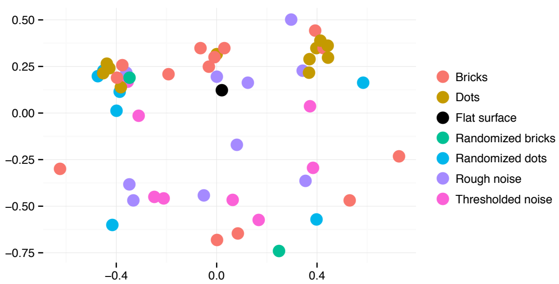


Figure 6. *Results of multidimensional scaling using data evaluated by at least three subjects.*

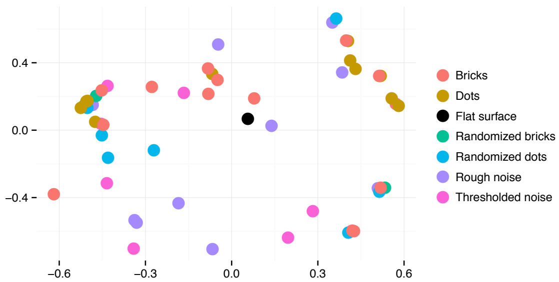


Figure 7. *Results of multidimensional scaling using all data.*

terns. These noise-based classes (both threshold and rough) are still fairly spread out compared to other class clusters, suggesting that these pattern types have more variety in their possible brightness profiles. Several clusters of brick and dot patterns are located near the flat surface coordinate. These patterns are dominated by flat regions parallel with the overall stimuli plane broken apart by vertical gaps. When these gaps do not create unreasonable shadowing, e.g. they are small compared to the size of the brick or dot, the surface approaches the idealized flat surface. This relationship is captured in the MDS plots, providing justification for the intuition by the flat surface model mentioned previously.

Model Comparison

The two models introduced earlier in the paper will be evaluated using the first data subset containing responses on a single surface from the first 12 subjects. The responses to the 31 poses are shown in Figure 8 where the light and view pose is arranged arbitrarily along the x-axis. The 36 separate responses (12 subjects $\times$ 3 repeated measures) are shown as transparent black points. Each subject's repeated measures are averaged and color-coded per-user across all poses. The average over every subject is shown as a purple trend line, alongside the two models: the average luminance of the image, and the luminance of a flat surface.

The accuracy of the models can be measured by the probability of their prediction being the population mean of perceived brightnesses. Performing the Student's t-test for each scenario shows that depending on the viewing and lighting condition, subject responses are significantly different from either model although the difference between them is relatively small. The flat surface model t-test produced p-values ranging from 2.853×10^{-12} to 0.882 and the surface mean p-values range from 1.464×10^{-5} to 0.818. Given that these distributions were based on only twelve subjects it is hard to rule out the models based on this alone. In general the p-values for the surface mean model were higher and more frequently above a significance test of 0.05. Thus for certain poses, the surface mean model could represent the population's perceived brightness. However, the at-specular scenarios in Figure 8 show a distinct separation of the flat surface model from both the subject average and surface luminance average. Interestingly though, the subject responses are frequently brighter than the surface mean model when at specular even if they are not as bright as the flat surface model predicts. When off specular, both models regularly fall amongst the subject distribution for the scene.

It is possible that a specular-dependent effect occurs in our perception of brightness over a complex mesosurface. Additional models will need to be considered, such as an interpolation between the two proposed here, or one based on the histograph of luminance of the rendered stimuli.

Discussion

The results collected so far are both promising and challenging. It is a good indication that there is significant agreement between subjects and that brightness judgments are consistent over time. Additionally the subject responses were robust enough to identify clusters of similar surface patterns with distinct brightness profiles. An important question that remains is how significant the variations between these clusters are and if they impact the results of previous research. It is likely that particular families of mesoscale surface patterns will scale up or down the effects on glossiness previously identified.

A potential future direction for this work is to explore if these clusters of materials agree with subjects when asked if two patterns match. Work on achromatic texture matching [3] suggests that even if surface's apparent brightnesses are the same their distribution of shadows and highlights will create an overall mismatch. Another avenue is further exploration of the predictive models for perceived brightness. Two models were proposed and evaluated in this work, each proving successful for certain classes of pattern or scene configuration. However, a more universal model must still be developed.

This work is valuable as a foundation for pursuing more complex questions regarding mesosurfaces. One such option is a massive simulation-driven approach to surface appearance that can be validated against this collected data. Additionally, the overall perceived brightness of a surface can be valuable for concrete applications. Real-time rendering that's required to approximate appearances and downsample far-away geometries can take advantage of this work to better filter mesoscale surfaces as their details lose resolution. In many cases the standard averaging will prove sufficient but accuracy will be lost unless it is brightened at specular directions.

In summary, we have presented a user study designed with a more complex and natural stimuli for the subject and have evaluated many different surface patterns. Our experimental design has allowed us to consider a wide variety of surface patterns. By densely sampling a small subset and then progressively sparser sets, the surface pattern domain was evaluated more thoroughly than what we've encountered in the past. Additionally the experimental time for each subject was kept to a minimum. We have also developed an approach for measuring similarity be-

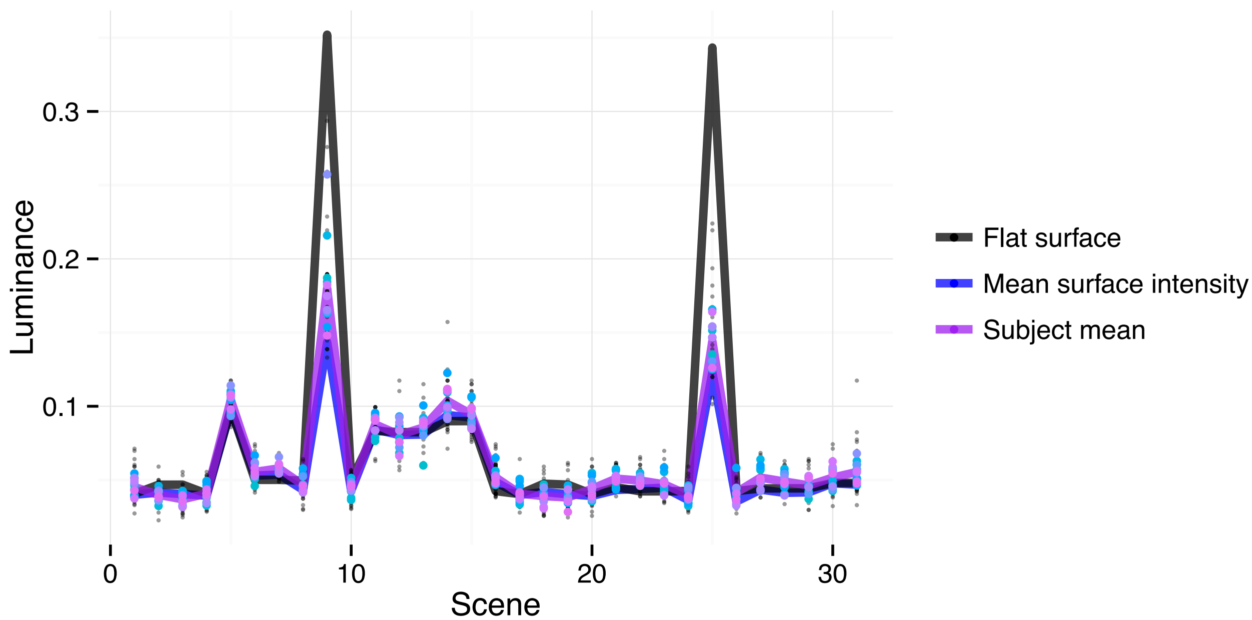


Figure 8. *Perceived brightness for the rough-noise surface presented to all subjects in the first study. The y-axis is linear luminance.*

tween spatially-varying surfaces under very sparse sampling. This approach is able to identify expected clusters given the generated surface geometries. The view and light dependent behavior of unique surface patterns is detectable from sparse brightness measurements although they may not be sufficient for appearance matching. Analysis of subject responses shows that brightness is consistent over time and that there is little variation between subjects, although it increases along the specular direction.

References

[1] Lawrence E. Arend and Branka Spehar. Lightness, brightness, and brightness contrast. *Percept. Psychophys.*, 54(4):446–456, 1993.

[2] M G Bloj, D Kersten, and a C Hurlbert. Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, 402(6764):877–879, 1999.

[3] Charles Chubb, Michael S. Landy, and John Econopouly. A visual mechanism tuned to black. *Vision Res.*, 44:3223–3232, 2004.

[4] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, 1994.

[5] James A. Ferwerda, Fabio Pellacini, and Donald P. Greenberg. A psychophysically-based model of surface gloss perception. In *Proc. SPIE Hum. Vis. Electron. Imaging*, pages 291–301. SPIE Press, 2001.

[6] Roland W. Fleming. Visual perception of materials and their properties. *Vision Res.*, 94:62–75, 2014.

[7] Roland W. Fleming, Ron O. Dror, and Edward H. Adelson. Real-world illumination and the perception of surface reflectance properties. *J. Vis.*, 3(5):347–368, 2003.

[8] Phillip J. Marlow, Juno Kim, and Barton L. Anderson. The perception and misperception of specular surface reflectance. *Curr. Biol.*, 22(20):1909–1913, 2012.

[9] George Marsaglia, Wai Wan Tsang, and Jingbo Wang. Evaluating Kolmogorov's distribution. *J. Stat. Softw.*, 8(18), 2003.

[10] Shree K Nayar and Michael Oren. Visual appearance of matte surfaces. *Science*, 267(5201):1153–1156, 1995.

[11] Stefano Padilla, Ondrej Drbohlav, Patrick R. Green, Andy Spence, and Mike J. Chantler. Perceived roughness of 1/f?? noise surfaces. *Vision Res.*, 48(17):1791–1797, 2008.

[12] Ken Perlin. An image synthesizer. *ACM SIGGRAPH Comput. Graph.*, 19(3):287–296, 1985.

[13] Lin Qi, Mike J. Chantler, J. Paul Siebert, and Junyu Dong. How Mesoscale and Microscale Roughness Affect Perceived Gloss. In *Proc. 3rd Int. Conf. Appear.*, pages 48–51, 2012.

[14] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic Tone Reproduction for Digital Images. *ACM Trans. Graph.*, 21(3):267–276, 2002.

[15] James Schirillo, Adam Reeves, and Lawrence E. Arend. Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information. *Percept. Psychophys.*, 48(1):82–90, 1990.

[16] Robert Shapley and R. Clay Reid. Contrast and assimilation in the perception of brightness. *Proc. Natl. Acad. Sci. U. S. A.*, 82(17):5983–5986, 1985.

[17] Bruce Walter, SR Marschner, Hongsong Li, and KE Torrance. Microfacet models for refraction through rough surfaces. In *Proc. 18th Eurographics Conf. Render. Tech.*, number 1992, pages 195–206. Eurographics Association, 2007.

Author Biography

Michael Ludwig received his BS in computer science, with an emphasis on human-computer interaction and graphics, from the University of Minnesota in 2010. He worked with the GroupLens HCI Lab after graduation until he joined Professor Gary Meyer's color and perception lab in 2012 as a PhD student. His work focuses on material representations and real-time rendering of complex materials, as well as BRDF scanning and fitting with applications to cultural heritage and computer vision.