Digital Transmission of Subjective Material Appearance

Rodrigo Martín　　　　Michael Weinmann　　　　Matthias B. Hullin
University of Bonn
Institute of Computer Science II
{rodrigo, mw, hullin}@cs.uni-bonn.de

ABSTRACT

The digital recreation of real-world materials has a substantial role in applications such as product design, on-line shopping or video games. Since decisions in design or shopping are often driven by qualities like "softness" or "beautifulness" of a material (rather than its photo-accurate visual depiction), a digital material should not only closely capture the texture and reflectance of the physical sample, but also its subjective feel. Computer graphics research constantly struggles to trade physical accuracy against computational efficiency. However, the connection between measurable properties of a material and its perceived quality is subtle and hard to quantify. Here, we analyze the capability of a state-of-the-art model for digital material appearance (the spatially-varying BRDF) to transport certain subjective qualities through the visual channel. In a psychophysical study, we presented users with measured material SVBRDFs in the form of rendered still images and animations, as well as photographs and physical samples of the original materials. The main insight from this experiment is that photographs reproduce better those qualities associated with the sense of touch, particularly for textile materials. We hypothesized that the abstraction of volumetric materials as opaque and flat textures destroys important visual cues especially in border regions, where fluff and protruding fibers are most prominent. We therefore performed a follow-up experiment where the border regions have been removed from the photographs. The fact that this step greatly reduced the capability of photos to transport important qualities suggests strong directions of future research in applied perception and computer graphics.

Keywords
Material perception; digital material appearance; SVBRDF; visual psychophysics

1 INTRODUCTION

The recent progress in the photo-realistic depiction of digitized materials has led to a paradigm change in important applications where the conventional approach of communicating objects in terms of photos taken by experts is more and more replaced by virtual surrogates. This methodology allows new possibilities such as co-operative product design, product advertising in prototype phase, exhibition of furniture or wearables in specific environments or visualization of cultural heritage objects. The entertainment industry has also drawn a major benefit from advanced digital material models as they allow a more realistic experience of virtual scenarios. While a remarkable reproduction quality has been achieved for virtual/digitized materials, there is still a gap in appearance between them and their physical counterparts which, in application, may distort the

perception of the product. In particular, the accurate reproduction of surface reflectance behavior under varying illuminations and viewing conditions still remains a challenging task. For this reason, many material catalogs [Hal10] still opt for using pictures or even physical material samples to illustrate their collections instead of digitized models, despite the potential benefits that they entail.

In this paper, we aim at investigating this breach in appearance between digitized materials and their physical counterparts by analyzing how perceptual material information is transmitted through different stimuli. For this purpose, we consider the perception of materials by assessing a set of subjective qualities that can be assigned to either the tactile, visual or affective category, depending on the nature of the interaction that best reveals them. We then conducted a psychophysical study to compare the communication of these attributes based on different representations given by real material samples, photographs of these samples as well as static and animated renderings of the digitized materials represented by the spatially-varying BRDFs (SVBRDF) model [Nic77], which is deemed to be a standard representation in research and industry [Ell12]). The materials evaluated belonged to two semantic categories, leathers and fabrics. A key observation obtained from

this experiment is that the SVBRDF model is not capable of preserving important qualities of material appearance, especially the tactile ones. Even a dynamic change of viewpoint does not seem to improve the perception of materials. Thus, the loss of information is presumably not caused by the limited resolution of the digitized samples but due to the abstractions intrinsic to the model. Upon closer inspection of the stimuli we observed that the differences between photos and virtual materials are most prominent at grazing angles, where the SVBRDF model fails to capture the volumetric material structure, intricate light scattering effects and the partial transparency of protruding fibers. Consequently, the perception of material properties such as softness, stiffness or transparency is not accurately recreated in the digitized representation, deviating from the correspondent photos and physical samples. With that in mind, we designed a follow-up study in which the respective border regions were digitally removed from the photos for a subset of relevant materials. Indeed, this step led to a significant deterioration in the transmission of tactile and affective properties, confirming our initial suppositions.

Our main findings are:

- Digitized materials (SVBRDF model) are not capable of adequately transmitting certain perceptual material information, being outperformed by simple photos of material samples. However, there is also an gap between photos and real materials.
- There are no significant differences in material-quality perception between static and animated digitized representations.
- The depictions of digitized materials suffer from a significant loss of information at grazing angles, where the SVBRDF model cannot represent appearance accurately.

To the best of our knowledge, this is the first perceptually-motivated work in evaluating how subjective material appearance is transmitted through digitized models (SVBRDF) in comparison to photos and real material samples. Conclusions from this set of studies are restricted to the given stimuli, but can provide useful insights for future research in developing realistic material representations.

2 RELATED WORK

In this section, we provide a condensed synopsis of research on the perception of subjective material qualities and the evaluation of digital material appearance models for graphics applications.

Perception of Materials and Their Qualities. The interest in unveiling the principles and reasons that determine the visual appearance of materials as well as in how humans visually perceive materials and their

properties has received an increasing attention over the last years. Respective surveys [Ade01, And11] provide a discussion of the main problems and challenges in this area of research including the perception of material surface and properties. A further examination of the challenges in material perception is provided by [Fle14], where the author outlines a new theory of material perception based on 'statistical appearance models'. Among the phenomena that contribute to material appearance, glossiness has received a considerable amount of attention. Several approaches aimed at finding perceptually meaningful reparameterizations of material gloss by exploring the relationships between physical parameters and the perceptual dimensions of glossy appearance [Pel00, Wil09]. The human capability of perceiving material gloss (gloss constancy) under varying motion, disparity and color conditions was investigated by Wendt et al. [Wen10]. In addition to gloss, Ho et al. [Ho06] researched the visual estimation of surface roughness, discovering that its perception is strongly influenced by the illuminant angle.

Motion is another aspect that has an important impact on the appearance of materials' surface. By analyzing the optical flow, Doerschner et al. [Doe11] identified three motion cues, in which the brain could rely in order to identify material shininess. Our investigation further evaluates which additional subjective information (if any) is revealed by motion when compared to still renderings of digital materials. Other than motion, shape and geometry have proven to be critical aspects in the perception of materials [Van07]. The importance of the shape for material categorization is well-known [Ade01] and also can be used as an additional cue for material recognition [DeG16]. In this regard, one of the conclusions of our research is the emphasis on geometry and appearance under grazing angles, as a decisive feature to accurately assess material qualities. Indeed, the tasks of material categorization and material property judgment are closely related as demonstrated by Fleming et al. [Fle13]. Their studies revealed a high degree of consistency between these two assignments, implying that humans access similar information about materials when performing both tasks.

Although the experimental procedure initially involves purely visual stimuli, the participants also rated the same attributes for the real material samples in a sort of interaction that makes use of all senses (multimodal or full-modal interaction). The described approach relate to previous studies in multimodal material perception [Fuj15, Mar15], which highlight the importance of the tactile and auditory channels in the perception of material information. Their work also relies on ratings not only for surface material properties, but also a set of affective attributes.

Perceptual Evaluation of Material Appearance Models. Both material appearance acquisition and modeling have been deeply researched [Hai13, Wei16]. Widely used digital representations of materials exhibiting a spatially varying appearance include Spatially-Varying BRDFs (SVBRDFs) [Nic77] and Bidirectional Texture Functions (BTFs) [Dan97]. Both representations model material appearance depending on the spatial position $\mathbf{x}$ on the object surface, the light direction ω_i and the view direction ω_o. SVBRDFs allow a more compact modeling of surface reflectance behavior than BTFs at the cost of neglecting effects of light exchange at subtle surface structures. As SVBRDFs have become a standard in industry [Ell12], we use this representation to analyze differences in the human perception of real and digitized materials.

Regarding the perceptual evaluation of appearance models, several investigations focused on analyzing the level of realism achieved by a concrete model. In this context, Meseth et al. [Mes06] verified the ability of BTF models to achieve photo-realism in comparison to standard representations (BRDFs) and photographs, at a coarse and fine scale. It was demonstrated that BTF materials entail a significant increase of realism over BRDFs at both scales, albeit being still inferior to the scene photographs. A study from Filip et al. [Fil16] determined and predicted the critical viewing distances at which a certain BTF can be replaced by the correspondent BRDF representation without decreasing the overall visual impression. Additionally, Jarabo et al. [Jar14] examined the effects of approximate filtering on the appearance of BTFs in different domains (spatial, angular and temporal). The authors identified interesting correlations between high-level descriptors and perceptually equivalent levels of filtering as well as with low-level BTF statistics.

3 EXPERIMENT 1: METHODS

The proposed experiment investigates the performance of a well-known appearance model when transmitting subjective material qualities in comparison to equivalent photographs from real materials. In addition, the exercise examines whether the consideration of a higher spatial resolution through motion in digital scenes provides additional cues in the aforementioned task. Throughout this section the stimuli acquisition, selection of material qualities and experimental procedure will be detailed.

3.1 Stimuli

Selection of Materials In the scope of this research, we explore the perception of physical and affective material qualities for two semantic classes (leathers and fabrics). Restricting our selection to these two concrete, well-known categories allows us to keep the study

and its conclusions manageable. Next, we have chosen ten material samples pertaining to these classes, each of them with an approximate size of 120×120 mm^2, with nearly flat geometry to match the requirements of the acquisition device, which is described in the following paragraphs. With this fine selection, we intended to maximize the relative intra-class heterogeneity not only in terms of the physical properties but also the aesthetic characteristics.

Photographs of Materials. In order to make the real and the virtual materials as comparable as possible, both the real and the virtual scene should share comparable geometry and illumination conditions. With that intention, our real scene was composed by a cardboard cylinder (80 mm diameter) to which the sample was attached. Cylindrical geometries have been frequently used in previous perceptual studies [Fil16] because of its well-defined texture mapping and for being one of the most discriminating shapes [Van07]. We covered the uppermost and lower part of the sample with white pieces of cardboard, which gently fixed the material to the cylinder. The height of the visible part of the sample along the vertical axis was approximately 90 mm. A reflecting sphere with a diameter of 50 mm was situated 10 mm right from the cylinder, and the whole setup was placed under natural illumination using a white, uniform piece of cloth as background. This arrangement is not arbitrary, given that during our internal tests we learned that subjects are more adept at this kind of subjective exercises when some context regarding the scene is provided. The complete setup can be observed in Figure 1, where the digital camera (Nikon 1 J5, resolution of 5568×3712 pixels) was situated at a distance of 280 mm in front of the material sample. We took a picture for each specimen while keeping the light and viewing conditions constant. The images were then corrected regarding white-balance, cropped and scaled to match the resolution of the final device (see Section 3.3). Moreover, during the photo session we used a remote-controlled 360° spherical panoramic camera (Ricoh Theta S) to probe the scene illumination. The resulting high-dynamic-range environment map was utilized to illuminate our virtual scenes.

Digitized Materials. The digitization of the material samples was carried out using a commercial scanning device [XR16] that allows the measurement of (flat) material samples. After taking images of the material sample from different viewpoints and under different illumination conditions, a surface normal map is obtained and the reflectance behavior is stored in terms of a Ward-SVBRDF. We refer to the supplementary material for more details on the material digitization process. The output format (AxF) is supported natively by several rendering applications such as Autodesk VRED, which was employed to generate the renderings used

Figure 1: View of the photo setup

Tactile	Visual	Affective
rough–smooth	shiny–matte	expensive–cheap
hard–soft	bright–dark	natural–synthetic
thick–thin	transparent–opaque	beautiful–ugly
stiff–flexible	homogeneous–	unrealistic–believable
	heterogeneous	

Table 1: Opposite-meaning quality pairs

in this study. We approximated the geometry of the described photographic setup in a virtual scene and used VRED's Full Global Illumination algorithm to render it, lighting the scene with the previously calculated environment map. For the animated scene, we rotated the camera 60° back and forth around the cylinder in the Y-axis, and rendered the scene at 60 frames per second to get a clip with a duration of 4 seconds. The resulting photos and renderings are shown in Figure 2.

Real Materials. During the course of the experiment, we handed samples from the actual materials to the participants, hence, allowing a full-modal experience of the individual material qualities. Instead of the samples that were used for the acquisition, we used smaller portions of the same sample (approximately 70×70 mm^2) to avoid damaging the originals due to the interaction and in favor of the scalability of the process.

3.2　Selection of Material Qualities

In an initial step, we focused on finding a meaningful subspace of subjective adjectives that characterizes our selection of materials. The importance of this task was first addressed by Rao and Lohse [Rao96] for the concrete case of visual textures. We collected a list of 42 subjective material qualities organized in 21 opposite-meaning adjectives, which were observed to be the most recurrent ones in related literature regarding material perception [Fle13, Fuj15, Jar14, Mar15]. Such qualities were conceptually separated in three different groups with respect to their tactile, visual or affective nature. In pursuance of getting a smaller subspace of qualities that maximizes the transported information about our particular material collection, we conducted a pilot experiment in which we handed out the 10 original material samples to 7 participants along with a list

of the 42 individual adjectives. The subjects were asked to mark the adjectives that better describe each sample. There was no restriction regarding the number of adjectives to choose. From the results, we selected the most voted attribute pairs in each of the three groups for our experiments, leading to a final assortment of 11 adjectives (see Table 1). Although it was not in our original list, we additionally included the pair 'unrealistic-believable', which provides information about the level of realism portrayed by the virtual materials.

3.3　Experimental Procedure

The user study was conducted using tablet computers (Toshiba Excite Pro 10.1, resolution of 2560×1600 pixels) running a custom Android application. This experimental setup makes our study scalable to larger surveys in addition to representative of contemporary consumer hardware. The procedure was carried out in a quiet, well-illuminated room and organized in sessions with a maximum of 7 participants. An introductory presentation was provided before performing the exercise to explain the procedure and clarify inquiries. Participants were instructed to infer the qualities which were not evidently revealed in a particular representation.

Different techniques were contemplated to perform perceptual quality ratings across our stimuli. Although double stimulus ratings or forced-choice pairwise comparisons may lead to smallest measurement variance, they would also increase the number of trials and, thus, make the whole study more difficult to accomplish. Therefore we decided to employ single stimulus ratings in which, for each stimulus, the subjects had to rate the selected qualities on a 7-point Likert scale characterized by a slider with values ranging from -3 to 3 (see supplementary material). Each of the values in the slider was consistently labeled with a term indicating the intensity of the stimuli in both axes (e.g., very bright, bright, a bit bright, neutral, a bit dark, dark, very dark). The actual procedure consisted of four different presentations or conditions, in which different material images were presented to the participants in randomized order along with the rating questionnaire. In addition, the participants had the chance to examine the real samples, serving the respective ratings as ground truth. The conditions that compose the experiment are illustrated in Figure 3 and listed below:

- Photographs (PH) taken from the real materials.
- Digitized static renderings (DR) from materials using the SVBRDF reflectance model.
- Digitized video renderings (DV) using the same reflectance model, where the camera rotates around the sample in the Y-axis.
- Full-modal condition (FM). Physical material samples were given to the participants so that they could interact with them.
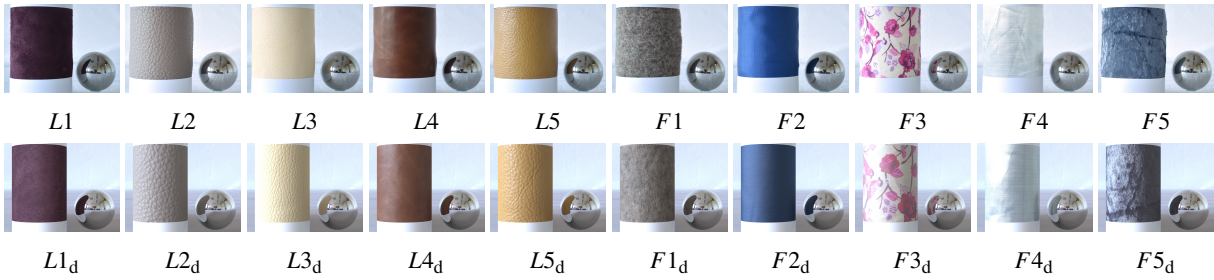
Figure 2: In the upper row, the pictures from the samples utilized in the study. In the lower row, the correspondent digitized material renderings. Larger stimuli images are provided in the supplementary material.
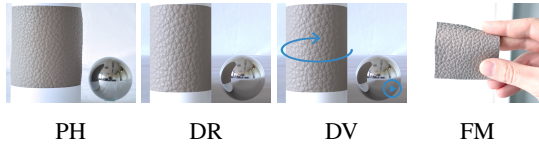

Figure 3: Stimuli presented in the psychophysical experiment corresponding to the four different conditions for an example material *L2*. From left to right, photograph (PH), digitized render (DR), digitized video (DV) and the physical sample (FM).

As the interaction with the real samples may bias the rest of the task, the condition FM was constrained to be the final one, while the order of the remaining conditions was randomized. Additionally, the application was instrumented to identify incorrect realizations of the assignment (e.g. skipping a material), in order to make the data more reliable. A total of 20 subjects (13 females, mean age 27.69; 7 males, mean age 27.00) participated voluntarily in the experiment. They were all naïve to the purpose of the experiment and reported normal or corrected-to normal visual acuity. They also provided informed consent and were compensated economically for their participation. From all the combinations of the conditions, materials, qualities and subjects, we obtained $4 \times 10 \times 12 \times 20 = 9600$ rating responses that are analyzed in the next section.

4 EXPERIMENT 1: RESULTS

To evaluate how the subjective attributes were perceived in the aforementioned material presentations we performed a conjoint analysis of the participants' preferences and non-parametric tests. The participants' rating responses were deemed reliable (Cronbach's $\alpha = .93$). In addition, participants' mean ratings and confidence intervals for each material and quality are included in the supplementary material.

4.1 Conjoint Analysis

With the purpose of gaining a general understanding regarding how material perception differs between the individual conditions, we made use of conjoint analysis techniques [Gre90] on the subjects' ratings. This method has been used extensively in market research to measure the preferences of the customers among multi-attributed products and services. In our experiment, we analyzed the three visual conditions C_i with $i \in \{PH, DR, DV\}$ in the conjoint analysis and considered the following question: 'To what degree does condition C_i transmit the quality $q_k \in Q$ in comparison to the other conditions?'. Due to our experimental procedure based on single ratings, we cannot directly compare two conditions. Instead, we can evaluate them with respect to the full-modal representation (FM). From this, we can infer that a certain condition C_i is more suitable to represent an individual property than another C_j if the participants' ratings better agree to the ones obtained for the full-modal condition (FM). In contrast, if the ratings are distant, the depiction is less realistic and, consequently, less suitable.

In order to carry out this comparisons, we use the weighed voting schema described in Martín et al. [Mar15] to compute the 'utility scores' (or 'part-worth utilities') s_i. For a certain combination of material, quality and subject, r_i and r_j denote the ratings for two particular conditions (C_i and C_j with $i, j \in \{PH, DR, DV\}$) and r_{FM} denotes the ratings for the full-modal task which serves as ground truth. The calculated intermediate utility scores ($s_{i,j}$) are defined according to

$$s_{i,j} = \begin{cases} |r_{FM} - r_i| - |r_{FM} - r_j| & \text{if } |r_{FM} - r_i| > |r_{FM} - r_j| \\ 0 & \text{else} \end{cases}. \tag{1}$$

To compute the final utility scores s_i, and the normalized 'importance scores' $T = (t_i)$ for each condition, we consider the matrix composed of the calculated intermediate scores $S = (s_{i,j})$ with $S \in \mathbb{N}^{N \times N}$ and $s_{i,i} = 0$. Note that, in general, the matrix S is not symmetric. Then T is given by

$$T = \frac{\sum\limits_{i} s_{i,j}}{\sum\limits_{i,j} s_{i,j}} \quad \text{where} \quad i, j \in \{PH, DR, DV\}. \tag{2}$$

The resulting scores, separated by property and material, can be seen in Figure 4. A clear evidence regarding

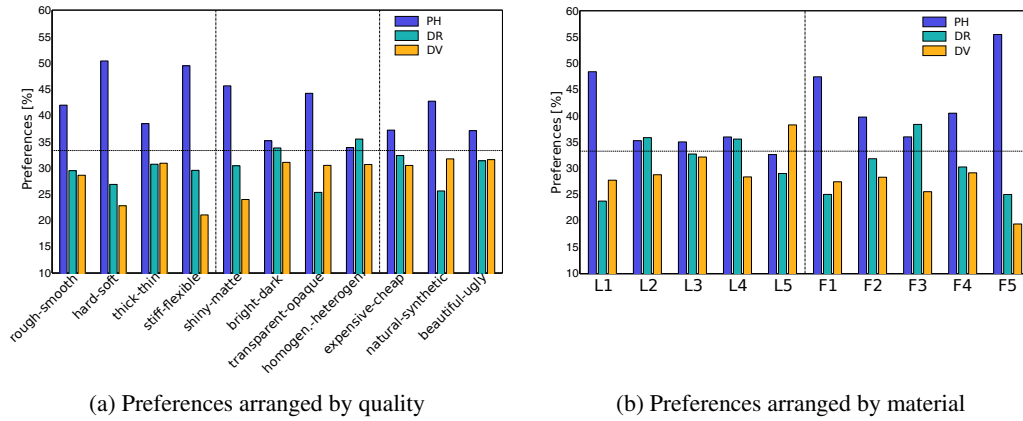(a) Preferences arranged by quality   (b) Preferences arranged by material

Figure 4: Summary of the conjoint analysis revealing participants' preferences for each condition according to our voting schema. The preferences are separated by quality (left figure) and by material (right figure). The PH condition is preferred for almost every quality and performs particularly better for fabric materials.

the preference of a certain condition with respect to the other ones would imply that the respective condition depicts the reality more accurately, for the corresponding material or quality. Indeed, the obtained results indicate a clear predilection towards PH for almost all qualities and materials. This preference is particularly noticeable for the tactile adjective pairs (e.g. 'hard-soft', 'stiff-flexible') but can also be observed for visual properties (e.g. 'shiny-matte', 'transparent-opaque') and affective properties (e.g. 'natural-synthetic'). Considering the preferences organized per material, the condition PH is especially favored for fabric specimens. In fact, if applying conjoint analysis between the two material classes, the scores obtained for digitized leathers ($t_{PH} = 38.09\%$, $t_{DR} = 31.15\%$ and $t_{DV} = 30.76\%$) are higher than the ones for digitized fabrics ($t_{PH} = 45.26\%$, $t_{DR} = 29.23\%$ and $t_{DV} = 25.51\%$). Another interesting finding shows up when comparing the importance scores among static and dynamic renderings (DR and DV). Initially, the video presentation only performs better when transmitting transparency and naturalness. Applying conjoint analysis between DR and DV exclusively led to a more balanced overall preference of $t_{DR} = 52.60\%$ and $t_{DV} = 47.40\%$. The pair 'unrealistic-believable' does not apply to the real material stimuli and hence, it was not considered during the analysis.

This way, conjoint analysis provides insights regarding how well individual qualities are transmitted by the different conditions and, hence, which of the corresponding representations is most suitable. In the next section, we intend to additionally discover if and where significant differences among the ratings of the conditions are manifested.

4.2 Non-Parametric Tests

In addition to compare each condition against the ground truth (FM), we would also like to detect whether, and if so also where, meaningful discrepancies between the individual conditions occur. This

may help us to understand how differently these representations transmit material qualities. A preliminary Shapiro-Wilk normality test determined that, for certain combinations of material and quality, our data do not come from a normally distributed population. This fact, together with the ordinal nature of the Likert scales, discredit analyses based on group means. Thus, we applied non-parametric tests (Friedman and Wilcoxon) in order to detect significant differences between the ratings (dependent variable) of the four conditions (independent variable).

Given our experimental design, we will be able to draw valid conclusions only for a single material-quality pair at a time ($p_i = \{m_j, q_k\}$, given a material $m_j \in M$ and quality $q_k \in Q$), across all subjects. For better understanding, we first consider the pair p_i given by the combination of material $L1$ and the quality 'rough-smooth'. Applying Friedman's test revealed that the effect of the different conditions on the subjects' judgments is significant ($\chi^2(3) = 19.86, p < .05, r = .00$). The post-hoc analysis with the Wilcoxon signed-rank procedure resulted into rejecting the null hypotheses for the comparisons FM $\leftrightarrow$ DR, FM $\leftrightarrow$ DV and DV $\leftrightarrow$ PH, i.e. these representations have a significantly different effect on the participants' ratings. In contrast, the comparisons FM $\leftrightarrow$ PH, DV $\leftrightarrow$ DR and DR $\leftrightarrow$ PH showed no interaction effect on the ratings. In order to extend our findings to the complete collection of M materials and Q qualities, we performed the same analysis for each possible combination of material and quality $p_i \in P$. Then, we summed up the number of occurrences in which, for a particular p_i, we rejected the null hypotheses and, therefore, the ratings among conditions were determined to be significantly different (at least $p < .05$ for all the cases). We refer to this sum hereafter as the "dissimilarity score". Here, the presence of a high dissimilarity score between FM and another condition would outline how good or bad the respective depiction
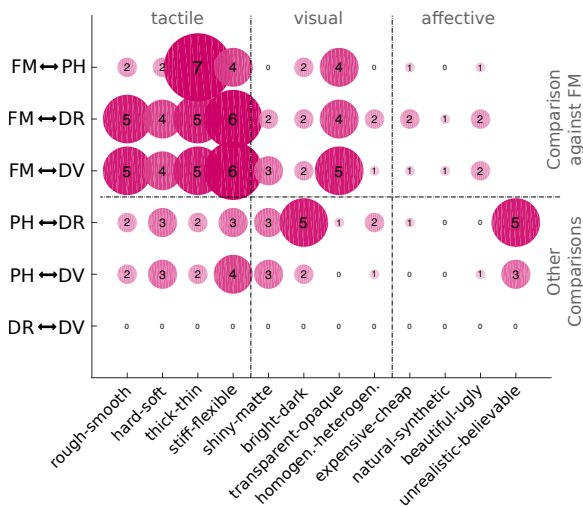
Figure 5: Number of tests with a significant effect on subjects' ratings (at least $p < .05$), summed along all ten materials. Each row compares two conditions while each column represents a quality pair. The largest scores are located in the second row (FM ↔ DR) and third row (FM ↔ DV), especially for tactile qualities. In contrast, there are no significant effects in the lower row (DR ↔ DV).

transmits real world information. In addition, by means of the same evidence for the rest of the scores, we may learn how differently photographs and digitized materials illustrate the individual qualities and if there is any significant impact in the ratings coming from motion. The outcome for all ten materials is separated by quality and shown in Figure 5.

As can be observed, the largest scores are mainly concentrated when the conditions FM ↔ DR and FM ↔ DV are compared, and this is especially appreciable for qualities categorized as tactile (upper-left quadrant). Besides, we can observe high scores between FM and the rest of the conditions for the adjective pairs 'thick-thin', 'stiff-flexible' and 'transparent-opaque'. This fact indicates that none of our representations is able to fully communicate these concrete qualities. Furthermore, the small scores found between the conditions FM ↔ PH in the remaining adjective pairs suggests that photos transmit most of the qualities good enough. In general, these results correlate well with the findings from the conjoint analysis, as they also tend to indicate the predominance of photographs over our digitized materials, especially in the tactile domain. In fact, the differences in the perceived realism ('unrealistic-believable' dimension) between PH and virtual materials confirm this trend. Finally, no significant dissimilarities were discovered in the comparison DR ↔ DV, i.e. the overall perception of material qualities is not affected by motion.

5 EXPERIMENT 2

During the course of the previous experiment, we observed that the samples with padded and fluffy appearance do not transmit appropriately material appearance in the digitized conditions and, hence, were deemed to be more unrealistic (see supplementary material). These features are more salient in the distinctive border regions, which possibly behaved as one of the main sources of information in favor of photographs. Due to the limited resolution of the reconstructed surface geometry, these structures are not accurately captured and the SVBRDF model is not capable of reproducing surface effects like self-occlusions, interreflections or transparency. To better understand how this matter influences the transmission of material appearance and which subjective attributes are most affected, we designed a follow-up study in which the perception of digitized materials was compared to the perception of real materials within photos where the border features have been digitally removed. The description of the experimental procedure and results are provided in the following sections. The supplementary material additionally provides the mean response ratings and confidence intervals for each material and quality.

5.1 Methods

From the materials selected for the previous experiment, we chose a subset of samples whose digitized stimuli were perceived to be particularly different from their correspondent photos in the experimental analysis, failing to transmit many of the considered attributes. According to this, we selected the set $M_2 = \{L1, L5, F1, F5\}$, where material $L5$ was only included to have an equal number of leathers and fabrics in the scope of this study. From the original photographs we removed the visible material borders from the cylindrical geometry to which the sample was attached using Adobe Photoshop, resulting into a flat silhouette shape as shown in Figure 6. Accordingly, we rendered again the digitized materials to match the new resolution from the cropped photographs. Other than that, we also aimed at comparing our visual stimuli against the real materials and we considered the same assortment of perceptual qualities as in the previous experiment. Nevertheless, in this experiment no motion was included, i.e. the considered conditions are:

- Cropped photographs (PH_c) taken from the real materials, where the borders have been removed.
- Digitized static renderings (DR) from materials using the SVBRDF reflectance model.
- Full-modal condition (FM). Physical material samples were given to the participants so that they could interact with them.

Again, the order of the materials and conditions was randomized except for FM, which was constrained to be
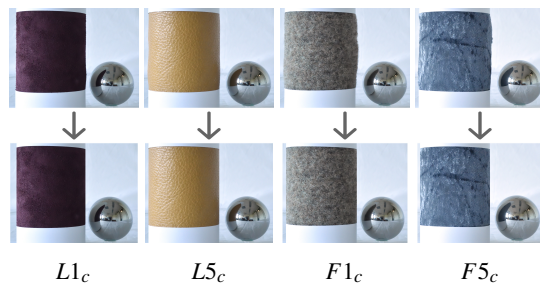
$$L1_c \qquad L5_c \qquad F1_c \qquad F5_c$$

Figure 6: Pictures from the two leathers ($L1$, $L5$) and two fabrics ($F1$, $F5$) selected for the follow-up experiment, before and after crop operation.

the last one. 19 subjects (12 females mean age 27.08; 7 males, mean age 28.57) took part in the experiment under the same conditions as the previous one. The resulting $3 \times 4 \times 12 \times 19 = 2736$ rating responses are evaluated in the next section.

5.2 Results

In the following, we show the outcome of performing conjoint analysis and non-parametric tests on the subjects' ratings and compare them w.r.t. the results of the previous experiment. A Cronbach's alpha value of $\alpha = .89$ confirms the reliability of the ratings.

Conjoint Analysis Similar to Section 4.1 we performed a conjoint analysis in order to reply the question: 'To what degree do the conditions PH_c and DR transmit the quality $q_k \in Q$ in comparison to FM?'. Figure 7a illustrates the importance scores per quality for the current experiment, in which the borders were removed, while Figure 7b shows the scores obtained for Experiment 1, if only the data from the subset M_2 of materials were taken into account. Direct comparison between the scores corresponding to the conditions PH_c (Experiment 2) and PH (Experiment 1) reveals how the preferences for the cropped photographs become significantly smaller for all the tactile and affective attributes so that, for certain cases, these are surpassed by the DR scores. Certainly, the score difference between conditions PH and PH_c in both experiments should be a good indicator regarding which perceptual attributes were most damaged with the border-feature removal. According to this, the most deteriorated pair was 'rough-smooth' (-20.52%), followed by 'natural-synthetic' (-18.63%), 'expensive-cheap' (-15.55%) and 'stiff-flexible' (-15.10%). Contrarily, the pairs 'bright-dark' ($+22.50\%$) and 'homogeneous-heterogeneous' ($+14.30\%$), were surprisingly better communicated without the borders. In this case, the silhouette information present in the photos from Experiment 1 could have acted as a misleading cue to judge homogeneity and brightness. Finally, the importance scores separated by material are shown in Figure 7c. When compared to the scores obtained

in Experiment 1 (Figure 7d), we notice substantial changes as the preferences for PH_c diminish in favor of the ones for the DR condition except for the material $L1$, whose scores remain relatively constant.

Non-Parametric Tests As in our previous study, we perform non-parametric tests (Friedman and Wilcoxon) to detect meaningful differences among the respective ratings (dependent variable) for the three conditions PH_c, DR and FM (independent variable). Anew, we carried out multiple comparison tests between the conditions (applying Wilcoxon signed-rank procedure) and generalized our findings by summing the resulting occurrences of rejected null hypotheses for each material-quality pair $p_i \in P$. The resulting dissimilarity scores are displayed in Figure 8a together with the scores resulting when applying the same test in Experiment 1 (Figure 8b), for the material subset M_2. the DV condition was ignored as the video stimuli were not used in the follow-up study.

From the results depicted in the figures, we can outline three main observations. First, the large dissimilarity scores found in the bottom comparison PH $\leftrightarrow$ DR for Experiment 1 have disappeared when moving to $PH_c \leftrightarrow$ DR in Experiment 2, which suggests that both representations lie much closer in the follow-up study. Second, the middle row comparing FM $\leftrightarrow$ DR only contains subtle changes in the scores obtained for both experiments. This fact is coherent with the stimuli as these conditions have not changed between experiments. Third, the top row comparing FM $\leftrightarrow PH_c$ in Experiment 2 presents, for most of the considered qualities, higher scores as in the original study. This fact suggests that the perception of photographs and real materials differs more significantly when the silhouette-border information is not present. However, the pair 'thick-thin' displays an unexpected opposite tendency. Again, borders may have acted as a misleading cue to judge thickness on these concrete samples.

6 CONCLUSIONS

In the scope of this investigation, we have studied the perceptual differences between stimuli based on standard digital material appearance models in terms of Spatially Varying BRDFs, photos of real materials (leathers and fabrics) and the actual material samples on the task of transmitting a rigorously selected group of subjective qualities. Additionally, we explored the effect of motion on the perception of the stimuli based on digitized material representations. Because of the observation that the appearance of photographed materials and their digitized counterparts differ particularly at the material borders, a second experiment was designed to explore to what degree the appearance of materials under flat viewing angles could cause the loss of information between photographs and renderings.

(a) Prefs. by quality (Exp. 2)   (b) Prefs. by quality (Exp. 1, M_2)   (c) Prefs. by material (Exp. 2)   (d) Prefs. by material (Exp. 1, M_2)
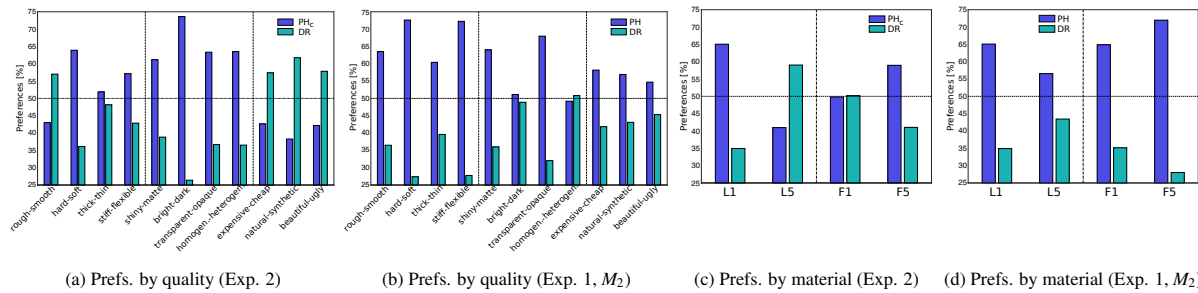
Figure 7: Summary of the conjoint analysis showing the participants' preferences for each condition in Experiment 2, where the material borders were removed from the photos, in contrast to Experiment 1, separated by quality (left figures) and material (right figures). The lower scores for the condition PH_c in comparison to PH show how the transmission of tactile and affective qualities as well as fabric samples deteriorates when the borders are removed.
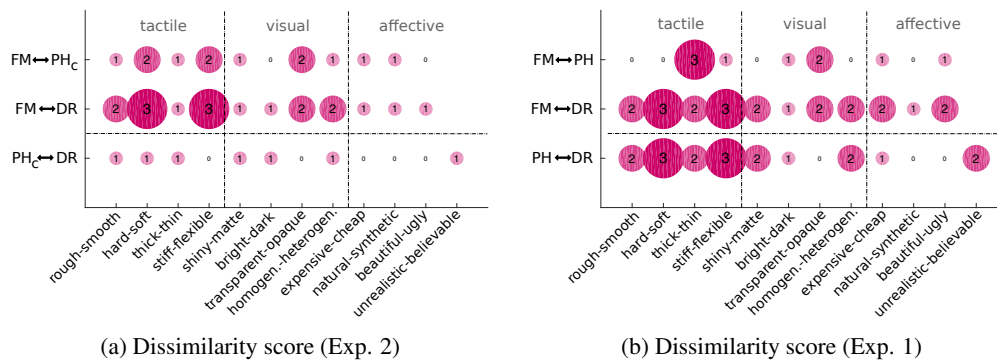

(a) Dissimilarity score (Exp. 2)   (b) Dissimilarity score (Exp. 1)

Figure 8: N° of tests with a significant effect on subjects' ratings (at least $p < .05$), summed along the subset M_2 of four materials. On the left, the dissimilarity scores for Experiment 2, where the borders were removed from the photos. On the right, the respective scores for Experiment 1. Note the high scores in the comparison PH ↔ DR for Experiment 1, whereas they partially move to the first row FM ↔ PH_c in Experiment 2. Meanwhile the middle row presents little variation.

One of the main findings of our investigations is that the considered digitized models are not able to fully transmit basic subjective properties according to the reality. Most of the analyzed perceptual qualities were better perceived in photos of real materials in comparison to renderings, but there is also a perceptual gap between photos and physical materials. This effect has proven to be true especially, but not exclusively, for tactile attributes and the fabric samples. Furthermore, motion information did not affect the perception of digitized materials significantly. The latter is especially relevant for the 'shiny-matte' dimension as they may contradict the documented fact that motion cues can override static ones while judging shininess [Doe11, Wen10]. Nevertheless, their experiments are based in much simpler appearance models (isotropic Ward model and grayscale Phong model respectively) which probably led to a better shininess isolation and recognition. Finally, our investigations indicate that more attention has to be paid to the accurate reconstruction of the distinctive material geometry as well as the acquisition of material appearance under grazing angles. In our measurements, the lowest camera was mounted with a zenith angles of 67.5° and, hence, these particular appearance effects cannot be recovered.

Although our studies provide interesting evidences, they cannot be extrapolated to other material categories (e.g. paper, stone, wood, etc.) for which additional experiments would have to be performed. We also acknowledge certain aspects that could have limited the expressiveness of the digitized materials used in our experiments, including:

- The generation of the virtual scene is approximate, i.e. the virtual camera position and the scene geometry slightly deviate from their physical counterparts.
- Scale differences between virtual and real material sample. Due to restrictions of the acquisition process, the digitized material represents a slightly smaller patch from the original one.
- The environmental light varied during the photo session due to the movement of sun and clouds.
- A color shift between the real and the virtual materials which also comes from the acquisition process.
- Not all the materials presented in this study were suitable to be represented by the SVBRDF appearance model, since it does not account for important surface effects. Consequently, some digitized representations were visibly defective.

By and large, we consider the results presented in this investigation an important step in the immense task of unveiling the perception of digital environments to improve the overall experience. To conclude, we point out the necessity of research in several directions such as the application of more appropriate, material-specific appearance representations. In this regard, BTFs models might help regarding the reproduction of fine effects of light exchange within the digital material representation at the cost of rather long acquisition times. Another interesting avenue of research could be to explore the linkage between perceptual qualities and physical measurable material properties (i.e. stiffness or roughness). Finally, the transmission of material qualities could benefit from a multisensory approach. In particular, the use of sound has proven to be beneficial for the assessment of tactile qualities [Mar15], which were not successfully transmitted using purely visual models.

7   ACKNOWLEDGMENTS

8   REFERENCES

[Ade01] Adelson, E. On seeing stuff: the perception of materials by humans and machines. In Proc. SPIE, 4299:pp. 1–12, 2001.

[And11] Anderson, B. Visual perception of materials and surfaces. In Current Biology, 21(24):pp. R978 – R983, 2011.

[Dan97] Dana, K. J., Nayar, S. K., van Ginneken, B., and Koenderink, J. J. Reflectance and texture of real-world surfaces. In *Procc of the Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 151–157. 1997.

[DeG16] DeGol, J., Golparvar-Fard, M., and Hoiem, D. Geometry-informed material recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1554–1562. IEEE, 2016.

[Doe11] Doerschner, K., Fleming, R., Yilmaz, O., Schrater, P., Hartung, B., and Kersten, D. Visual motion and the perception of surface material. In Current Biology, 21(23):pp. 2010 – 2016, 2011.

[Ell12] Ellens, M. and Lamy, F. From color to appearance in the real world. In *Proc. SPIE*, vol. 8291, pp. 82910B–82910B–8. 2012.

[Fil16] Filip, J., Vávra, R., Havlíček, M., and Krupička, M. Predicting visual perception of material structure in virtual environments. In Computer Graphics Forum, 2016.

[Fle13] Fleming, R., Wiebel, C., and Gegenfurtner, K. Perceptual qualities and material classes. In Journal of Vision, 13(8):p. 9, 2013.

[Fle14] Fleming, R. W. Visual perception of materials and their properties. In Vision Research, 94:pp. 62 – 75, 2014.

[Fuj15] Fujisaki, W., Tokita, M., and Kariya, K. Perception of the material properties of wood based on vision, audition, and touch. In Vision Research, 109:pp. 185 – 200, 2015.

[Gre90] Green, P. and Srinivasan, V. Conjoint analysis in marketing: New developments with implications for research and practice. In Journal of Marketing, 54(4):pp. 3–19, 1990.

[Hai13] Haindl, M. and Filip, J. *Visual texture: Accurate material appearance measurement, representation and modeling*. Springer Science & Business Media, 2013.

[Hal10] Hallett, C. and Johnston, A. *Fabric for Fashion: The Swatch Book*. Laurence King Publishing, 2010.

[Ho06] Ho, Y., Landy, M., and Maloney, L. How direction of illumination affects visually perceived surface roughness. In Journal of Vision, 6(5):p. 8, 2006.

[Jar14] Jarabo, A., Wu, H., Dorsey, J., Rushmeier, H., and Gutierrez, D. Effects of approximate filtering on the appearance of bidirectional texture functions. In IEEE Trans. on Visualization and Computer Graphics, 20(6):pp. 880–892, 2014.

[Mar15] Martín, R., Iseringhausen, J., Weinmann, M., and Hullin, M. Multimodal perception of material properties. In *ACM SIGGRAPH Symposium on Applied Perception*, pp. 33–40. ACM, 2015.

[Mes06] Meseth, J., Müller, G., Klein, R., Röder, F., and Arnold, M. Verification of rendering quality from measured btfs. In *Proc. of the 3rd Symposium on Applied Perception in Graphics and Visualization*, pp. 127–134. ACM, 2006.

[Nic77] Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T. Geometrical considerations and nomenclature for reflectance. National Bureau of Standards Monograph #160, U.S. Dept. of Commerce, 1977.

[Pel00] Pellacini, F., Ferwerda, J., and Greenberg, D. Toward a psychophysically-based light reflection model for image synthesis. In *Proc. of SIGGRAPH*, pp. 55–64. 2000.

[Rao96] Rao, A. and Lohse, G. Towards a texture naming system: Identifying relevant dimensions of texture. In Vision Research, 36(11):pp. 1649 – 1669, 1996.

[Van07] Vangorp, P., Laurijssen, J., and Dutré, P. The influence of shape on the perception of material reflectance. In ACM Trans. Graphics, 26(3), 2007.

[Wei16] Weinmann, M., Langguth, F., Goesele, M., and Klein, R. Advances in geometry and reflectance acquisition. In *Eurographics - Tutorials*. The Eurographics Association, 2016.

[Wen10] Wendt, G., Faul, F., Ekroll, V., and Mausfeld, R. Disparity, motion, and color information improve gloss constancy performance. In Journal of Vision, 10(9):p. 7, 2010.

[Wil09] Wills, J., Agarwal, S., Kriegman, D., and Belongie, S. Toward a perceptual space for gloss. In ACM Trans. Graphics, 28(4):pp. 103:1–103:15, 2009.

[XR16] X-Rite. Tac7 scanner. http://www.xrite.com/categories/Appearance/tac7, 2016. Accessed at 25th April 2017.