Supplemental Material: An intuitive control space for material appearance

Ana Serrano[1] Diego Gutierrez[1] Karol Myszkowski[2] Hans-Peter Seidel[2] Belen Masia[1]
[1] Universidad de Zaragoza, I3A [2] MPI Informatik

Contents

A Additional details on experiments

A.1 Experiment 0: Principal components space

To make sure that choosing only five principal components does not affect the perception of appearance, we have run a pilot test. We use as stimuli 20 BRDFs from the MERL database, manually selected to cover a wide range of appearances (see Fig. S.1). The test follows the two-alternative forced choice (2AFC) methodology. Two images are presented to the user: the original BRDF, next to the same BRDF represented with only the first five components. Each comparison asks about a particular attribute from our list. The user has to choose which of the two depictions of the material better conveys such attribute (for instance, *Which of the two images looks more metallic?*). The order and relative location of each version was randomized. Each subject was shown 25 BRDFs, and a total of 47 subjects took part in the experiment. On average, the BRDF represented with five components was chosen 49% of the times. We ran χ^2 tests per attribute to confirm whether users where actually picking randomly between both options, results are presented in Table 1. For all attributes we obtained a very high p-value, which speaks highly in favor of a random selection by subjects.

We conclude that a five dimensional space suffices for our subsequent tests. Additionally, limiting the space to five dimensions has an additional advantage: When sampling the space to create a larger database of BRDFs (Sec. 3.2 in the main paper), the reduction of the space to 5D improves the sampling process by avoiding the placement of samples in regions of the space with little impact on appearance.

A.2 Experiment 1: Building the space of attributes

Finding a parameter space providing an intuitive representation of material appearance is a long-standing problem, for which no definite answer [Eugene 2008; Choudhury 2014] nor methodology [Schwartz and Nishino 2013] exist, whereas usually naming depends on the field [Adelson 2001]. The parameter space must be reduced enough to be manageable, but also comprehensive enough to allow for rich yet intuitive appearance edits, even for inexperienced users. For this first test, we rendered a large

Table 1: *Results of the χ^2 test for the principal component space experiment.*

	χ^2	Df	p-value
Plastic-like	0.0169	1	0.8965
Rubber-like	0.3137	1	05754
Metallic-like	0.0041	1	0.9489
Fabric-like	0.6792	1	0.4098
Ceramic-like	0.4010	1	0.5265
Matte	0.0051	1	0.9430
Glossy	0.0727	1	0.7874
Bright	0.6545	1	0.4185
Rough	0.2426	1	0.6223
Strength of reflections	0.7059	1	0.4008
Sharpness of reflections	0.4175	1	0.5181

number of stimuli depicting different materials, built an extensive initial list of candidate appearance descriptors, and then relied on a user study to reduce them to a suitable size. We included in our list attributes ranging from high level class descriptors (e.g. ceramic-like) to low level appearance descriptors (e.g., strength of reflections). Relying on Fleming's work [2013], where he states that *we can also make many judgments about the perceived qualities of different materials irrespective of their class membership*, we do not make any restrictions about the type of descriptors in our list.

Stimuli Inspired by recent works on material perception and design (e.g., [Ngan et al. 2006; Kerr and Pellacini 2010; Jarabo et al. 2014]), our stimuli consist of spheres of 60 different materials from the MERL database [Matusik et al. 2003], chosen to span a wide range of different appearances. The spheres are lit by direct illumination. We render them using PBRT, and the *St. Peter's* environment map from the Light Probe Image Gallery [Debevec 1998], since real-world illumination, and that environment map in particular, facilitates material perception in single images [Fleming et al. 2003].

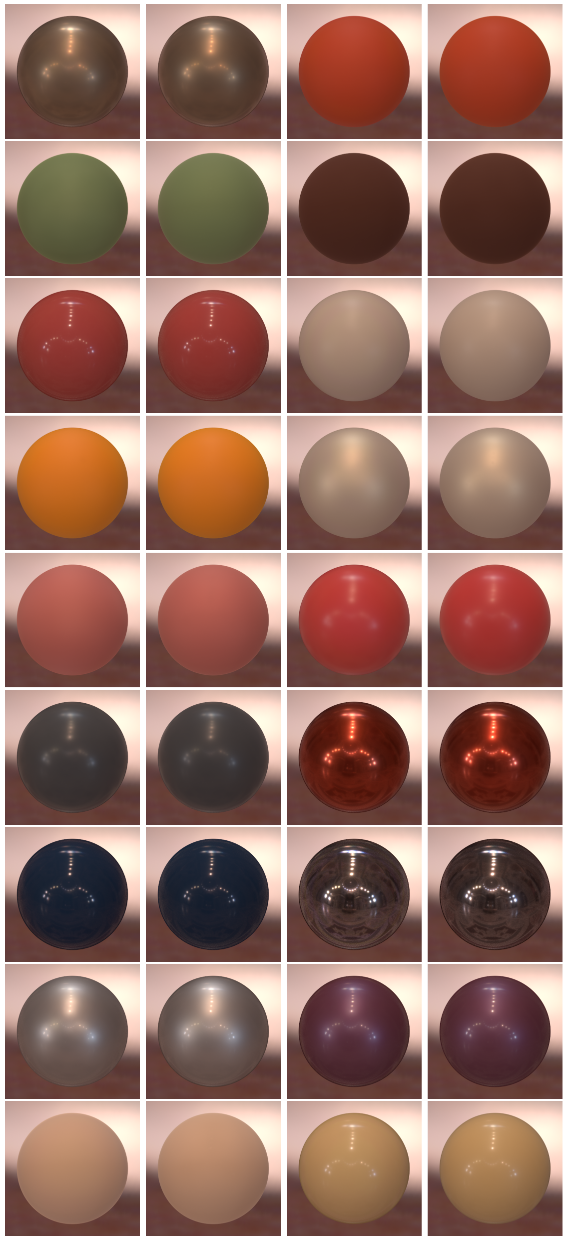
Initial list of attributes We compiled an extensive list of appearance attributes from previous works in industry and academia [Hunter and Harold 1987; Westlund and Meyer 2001; Burley 2012; Wills et al. 2009]. Additionally, seven subjects were asked to provide, for each of our 60 stimuli, at least four attributes that described its appearance, using their own words; this yielded a second initial list of attributes. We ensured that each stimulus was seen by at least two people. We then joined the two lists and reduced the number of entries by clustering semantically equivalent attributes; from this we obtained our initial list of 28 appearance attributes (see Sec. H).

Participants Twenty-six paid subjects took part in our experiment, under controlled conditions in our lab. They all had self-reported normal or corrected-to-normal vision, and had no graphics background.

Procedure We seek to further reduce the initial 28 attributes, keeping only those meaningful and understandable even by inexperienced users, and reasonably well represented in our database. To do this, we devised an experiment in which subjects had to establish, for each stimulus shown, whether each of the attributes applied to the material or not. Each subject was randomly shown 12 stimuli on a calibrated display, and there was no time limit (on average the complete tests took around 20 minutes per subject). Among the stimuli, a specific BRDF (the same for all subjects) was shown twice throughout the experiment, and served as a control stimulus for outlier rejection. This experiment would tell us: First, for which attributes there is a high agreement between users, and therefore they are clearly understood; and second, which attributes systematically received negative answers and thus are not representative of material appearance in our database.

Analysis and main findings We first computed agreement, as the percentage of responses coincident with the majority answer. Additionally, we computed Hamming distances between answers for different attributes, as an indicator of correlation between them, and confirmed these correlations using Pearson's chi-square test [Pearson 1900; Fisher 1922], which analyzes whether there is a relationship between each pair of attributes, as well as the strength of this relationship. Attributes were then removed according to three conditions: a chi-square value above 65, an agreement below 0.8, or a Hamming distance below 0.2. The final list consists of fourteen attributes, covering both high- and mid-level features: *plastic-like, rubber-like, metallic-like, fabric-like, ceramic-like, soft, hard, matte, glossy, bright, rough, tint of reflections, strength of reflections*, and *sharpness of reflections*.

Figure S.1: *Stimuli for the principal component space experiment. For each pair the BRDF on the left is the original, and the BRDF in the right is represented only with the first five components of the PCA space.*



A.3 Data pre-processing: outlier rejection

In Experiment 2 we gather up to 56,000 responses from 400 subjects via Amazon Mechanical Turk. Since these responses are perceptual ratings which we will use to derive our mappings (*attribute-PC space*), we need an effective outlier rejection prior to using the gathered data to fit the RBFNs.

We use the BRDF shown in Fig. S.2 as a control question to reject outliers. We discard full subjects that do not have a reasonable answer to very clear attributes regarding our control image which are:

- Glossy = 4 or 5
- Metallic = 4 or 5
- Strength of reflections = 4 or 5

- Sharpness of reflections = 4 or 5

We also discard BRDFs from our experiment that are confusing for most of the users. We do this by calculating the difference between the 3rd and the 1st percentile of the observations. If this difference is greater than two for more than four attributes of the BRDF, we consider this BRDF as confusing for the users.

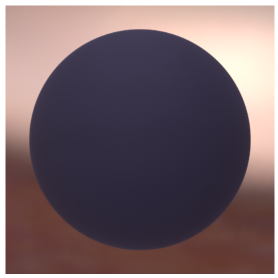
Finally, we discard outliers regarding observations for each attribute in each brdf. We do this if the observations fulfills any of the following conditions:

$Observation < (P_1 - K_d * P_d)$

$Observation > (P_3 + K_d * P_d)$

with $P_d = P_3 - P_1$ and $Kd = 1.5$

Figure S.2: *Control image used for outlier rejection in our experiments*



A.4 User study interface

We show in Fig. S.3 the web-based interface used for the Experiment 0 (2AFC), and in Fig. S.4 the web-based interface used for the Experiments 1 and 2 (Likert rating).

Figure S.3: *Web-based interface used for the experiment 0*

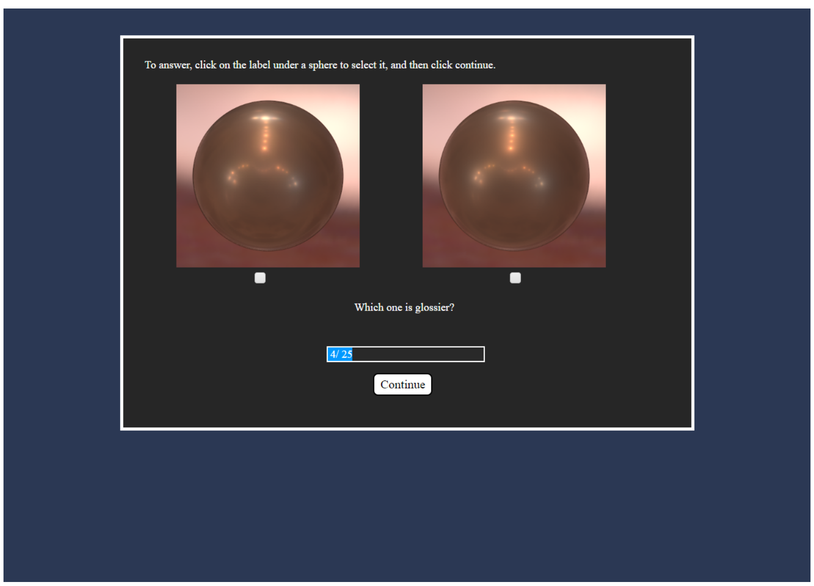
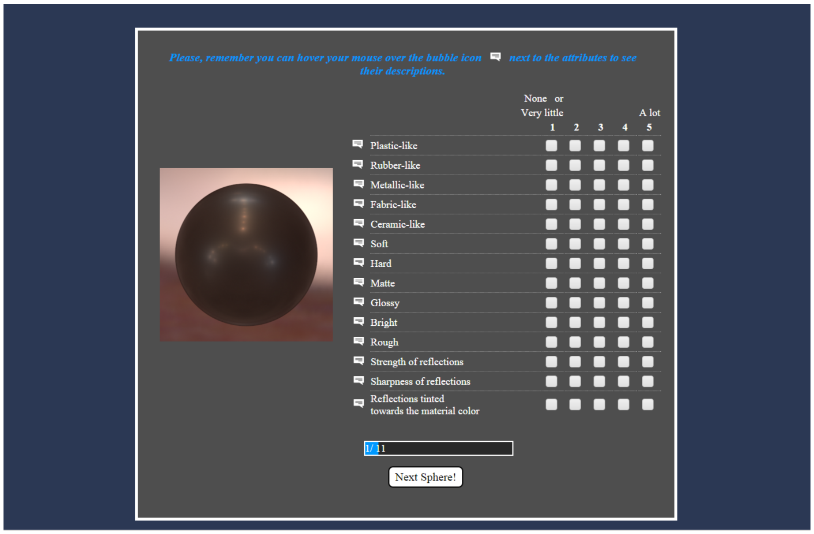


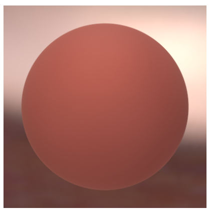
Figure S.4: *Web-based interface used for the experiments 1 and 2*

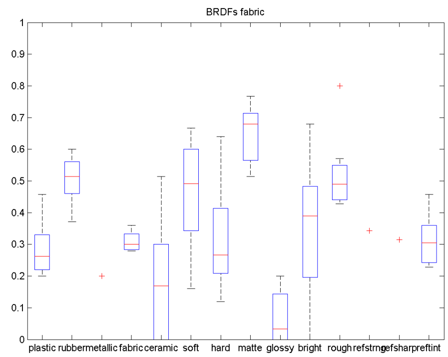


B Per cluster analysis

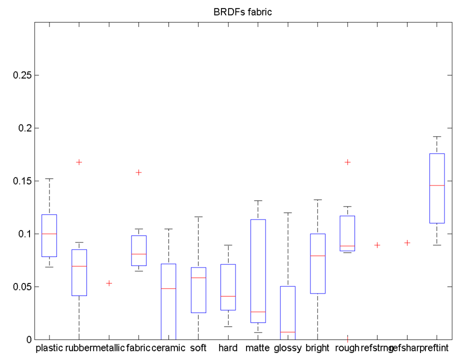
In Figs. S.5 and S.6 we show a per-cluster analysis of the mean and variance. Please refer to the main paper (Sec. 6.2) for cues on how to interpret these plots.

Figure S.5: *Means and variances for different types of BRDFs (I).*

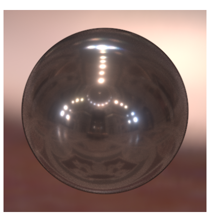


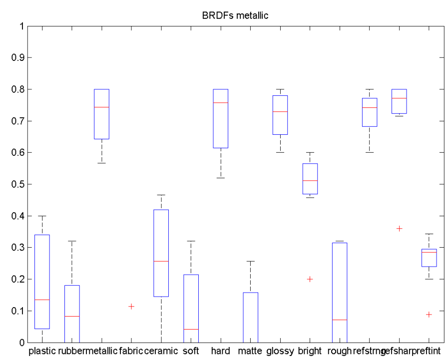


Mean values per attribute for fabric *BRDFs*

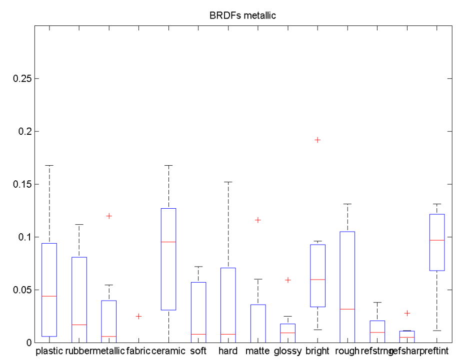


Agreement per attribute for fabric *BRDFs*

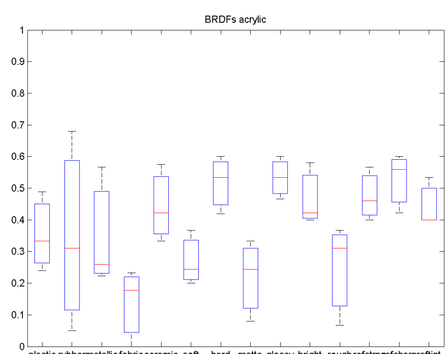




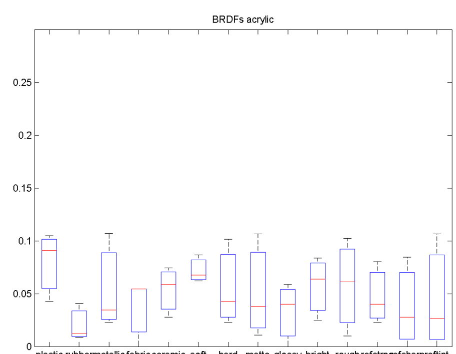
Mean values per attribute for metallic *BRDFs*



Agreement per attribute for metallic *BRDFs*

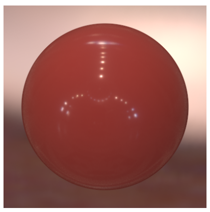


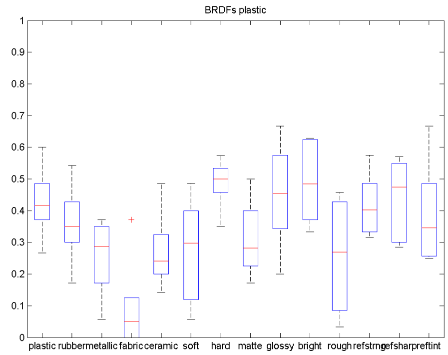
Mean values per attribute for acrylic *BRDFs*



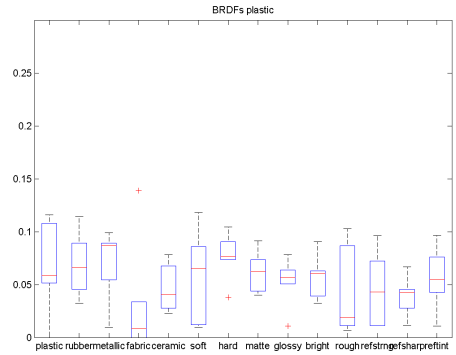
Agreement per attribute for acrylic *BRDFs*

Figure S.6: *Means and variances for different types of BRDFs (II).*

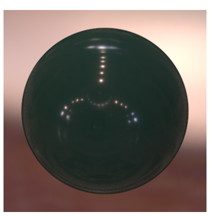


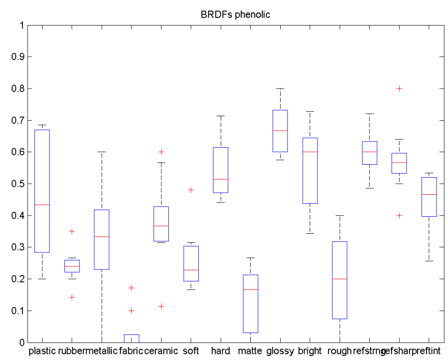


Mean values per attribute for plastic *BRDFs*

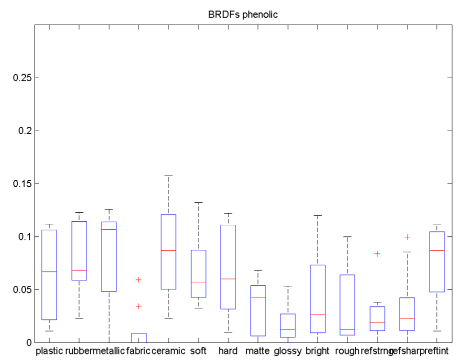


Agreement per attribute for plastic *BRDFs*

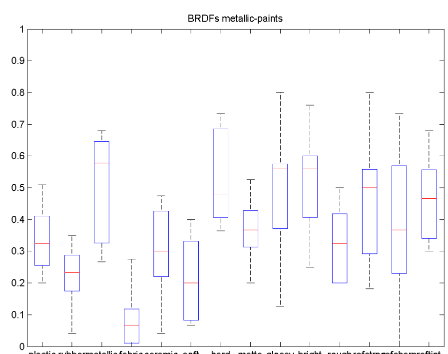




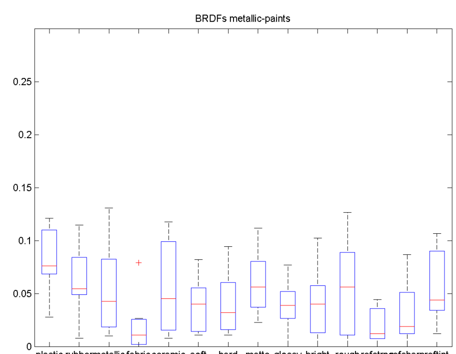
Mean values per attribute for phenolic *BRDFs*



Agreement per attribute for phenolic *BRDFs*



Mean values per attribute for metallic-paint *BRDFs*

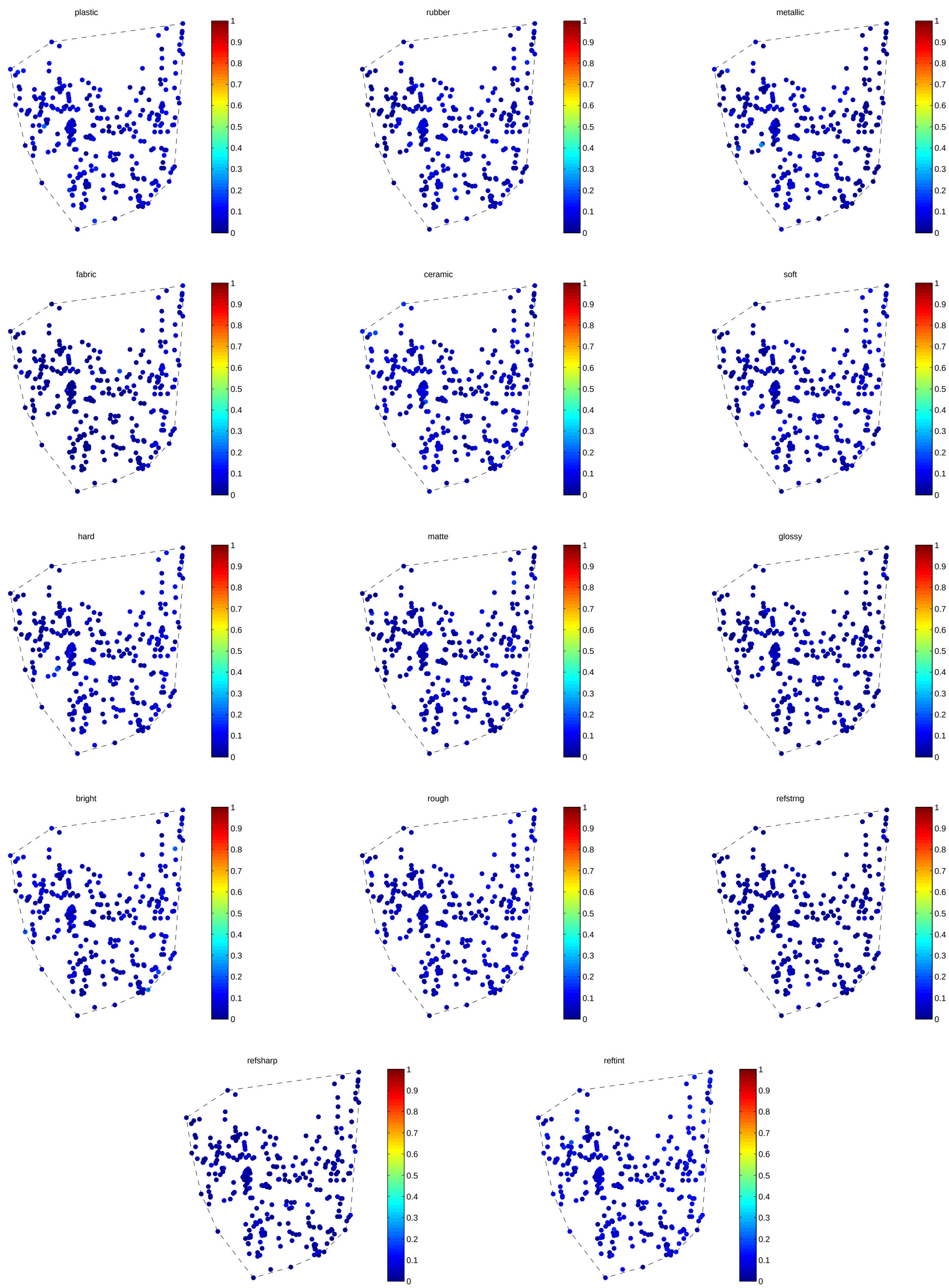


Agreement per attribute for metallic-paint *BRDFs*

C Goodness-of-fit

In this section we show the goodness-of-fit maps derived for all our attributes as explained in Sec. 5.1 in the main paper. We evaluate the goodness-of-fit of the RBFs by calculating for each attribute, and for all the BRDFs in our database, the mean distance between the values predicted by our functionals and the answers actually given by each particular user.

Figure S.7: *Goodness-of-fit maps derived for all our attributes*



D Correlations

In this section we present additional correlation analysis of our attributes. In Sec. 6.3 in the main paper we show the Pearson correlation analysis, we complete this analysis in Fig. S.8 providing the results of the Spearman correlation.

Figure S.8: *Spearman correlation analysis between our attributes*

	Plastic-like	Rubber-like	Metallic-like	Fabric-like	Ceramic-like	Soft	Hard	Matte	Glossy	Bright	Rough	Str. of refl.	Tint of refl.
Rubber-like	0.17	1.00											
Metallic-like	-0.24	-0.71	1.00										
Fabric-like	-0.01	0.52	-0.43	1.00									
Ceramic-like	0.10	-0.08	0.03	0.04	1.00								
Soft	0.10	0.68	-0.61	0.63	0.01	1.00							
Hard	-0.14	-0.65	0.66	-0.57	0.06	-0.85	1.00						
Matte	0.01	0.75	-0.69	0.63	-0.03	0.70	-0.69	1.00					
Glossy	0.00	-0.76	0.75	-0.59	0.12	-0.69	0.73	-0.92	1.00				
Bright	0.06	-0.19	0.12	-0.10	0.20	-0.07	0.10	-0.19	0.29	1.00			
Rough	0.00	0.63	-0.54	0.49	-0.01	0.60	-0.56	0.77	-0.76	-0.12	1.00		
Str. of refl.	-0.05	-0.77	0.79	-0.58	0.08	-0.73	0.75	-0.88	0.92	0.23	-0.73	1.00	
Sharp. of refl.	-0.05	-0.75	0.74	-0.58	0.11	-0.69	0.73	-0.90	0.91	0.21	-0.75	0.94	1.00
Tint of refl.	0.16	0.01	0.03	-0.04	0.14	0.04	-0.04	0.02	0.04	0.20	0.09	0.05	0.05

E Proof of concept with novice users

We provide in Figs. S.9, S.10, and S.11 additional results of the proof of concept test described in Sec. 8 in the main paper.

Figure S.9: *Results from editing the BRDFs* Pair #1. *The task was performed by three different novice users and consisted on finding a BRDF of intermediate appearance given an initial and a final appearance, with 3ds Max (bottom row) and our prototype (top row). Our prototype yields more similar results across users, and allows them to achieve better results in less time.*

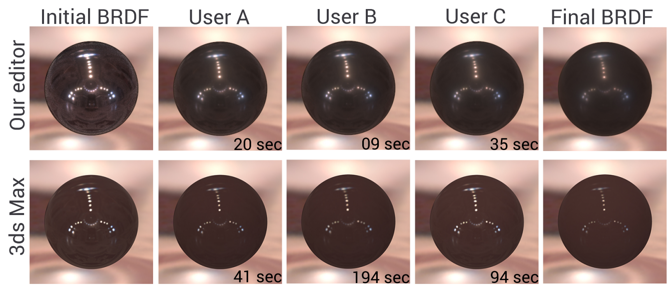


Figure S.10: *Results from editing the BRDFs* Pair #2. *The task was performed by three different novice users and consisted on finding a BRDF of intermediate appearance given an initial and a final appearance, with 3ds Max (bottom row) and our prototype (top row). Our prototype yields more similar results across users, and allows them to achieve better results in less time.*

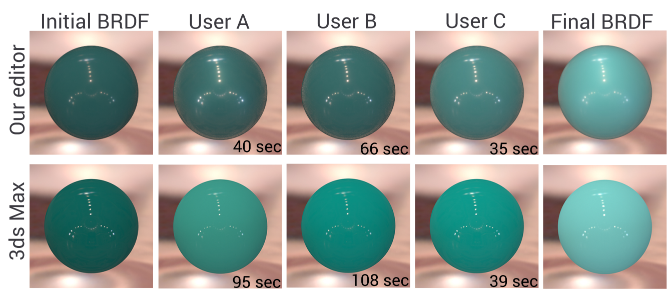
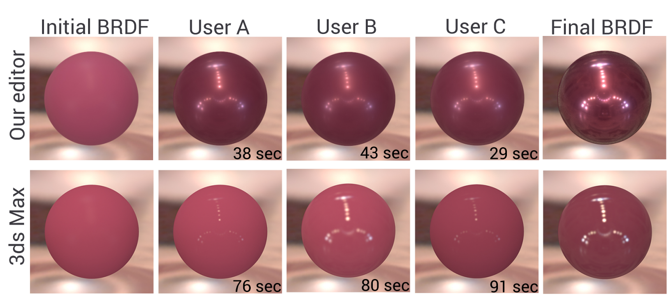


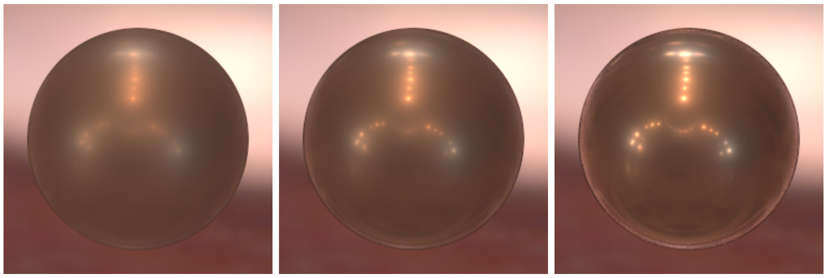
Figure S.11: *Results from editing the BRDFs* Pair #3. *The task was performed by three different novice users and consisted on finding a BRDF of intermediate appearance given an initial and a final appearance, with 3ds Max (bottom row) and our prototype (top row). Our prototype yields more similar results across users, and allows them to achieve better results in less time.*



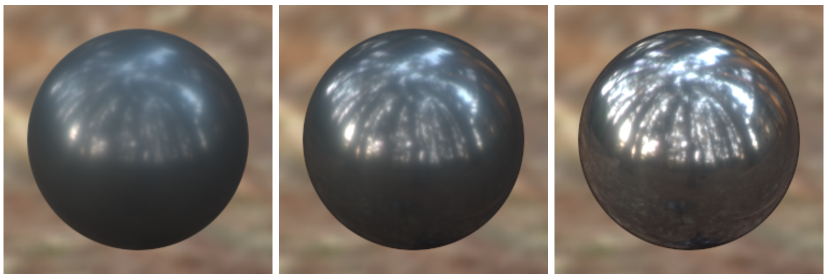
F Additional editing results

In this section we show more examples of BRDFs obtained by modifying attribute values. We indicate the name of the original BRDF and which attribute is modified in each of the examples.

Figure S.12: *Edits of different attributes for a variety of BRDFs ordered alphabetically.*

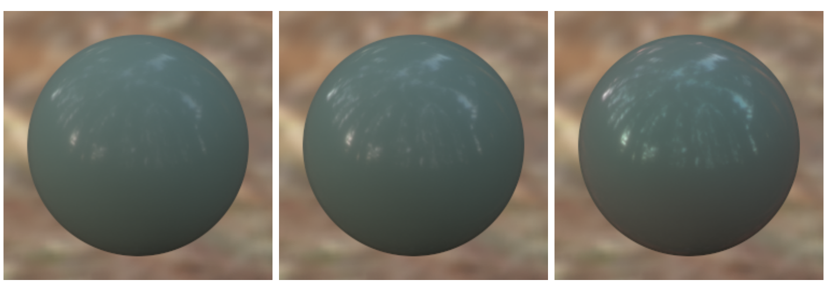


Edits for the brdf alum-bronze of the attribute sharpness of reflections



Edits for the brdf aluminium of the attribute sharpness of reflections

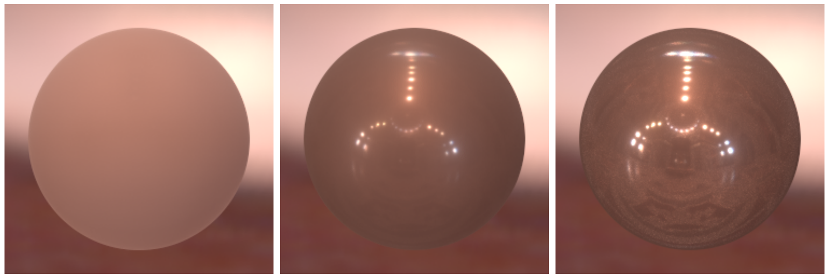
Edits for the brdf aventur-nine of the attribute rubber



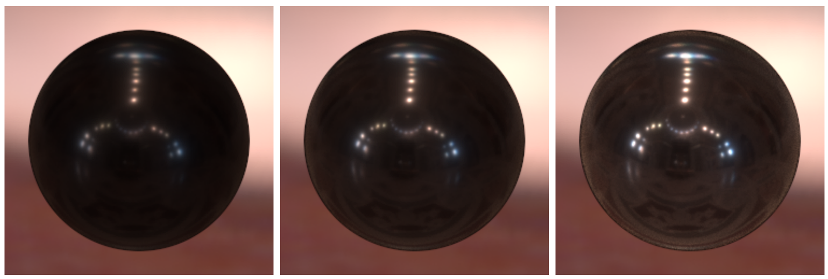
Edits for the brdf aventur-nine of the attribute sharp-ness of reflections

Edits for the brdf beige-fabric of the attribute matte

Figure S.13: *Edits of different attributes for a variety of BRDFs ordered alphabetically.*

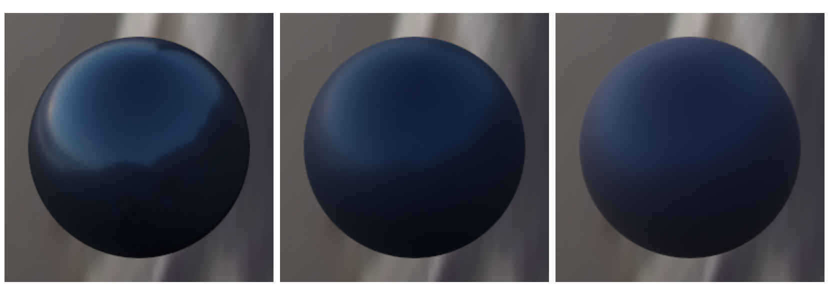


Edits for the brdf beige-fabric of the attribute glossiness

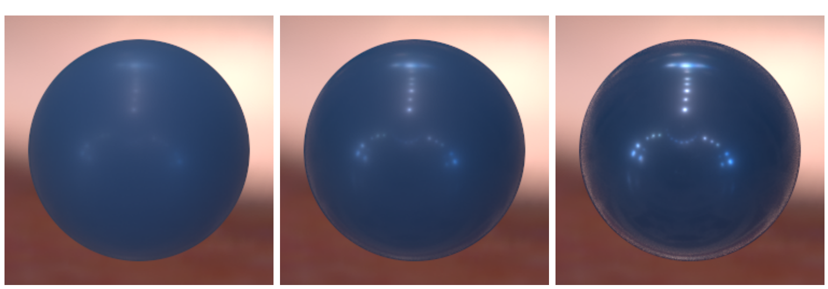


Edits for the brdf black-phenolic of the attribute strength of reflections

Edits for the brdf blue-acrylic of the attribute roughness

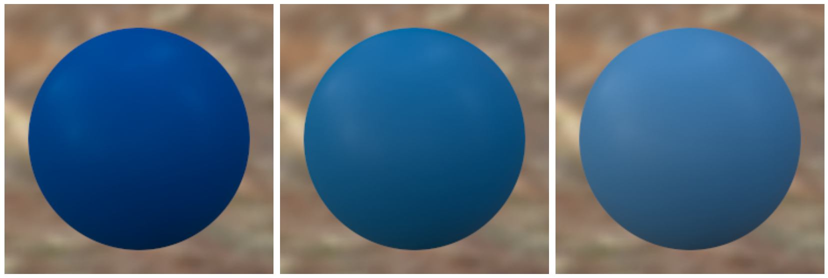


Edits for the brdf blue-metallic-paint2 of the attribute matte

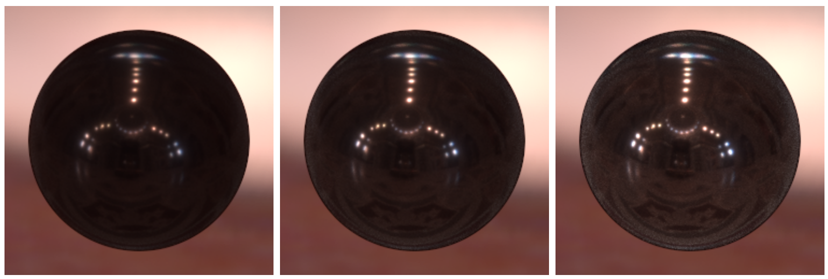


Edits for the brdf blue-rubber of the attribute glossiness

Figure S.14: *Edits of different attributes for a variety of BRDFs ordered alphabetically.*



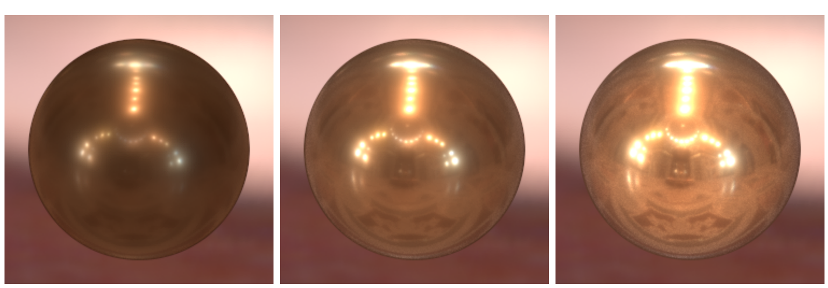
Edits for the brdf blue-rubber of the attribute brightness



Edits for the brdf chrome of the attribute strength of reflections

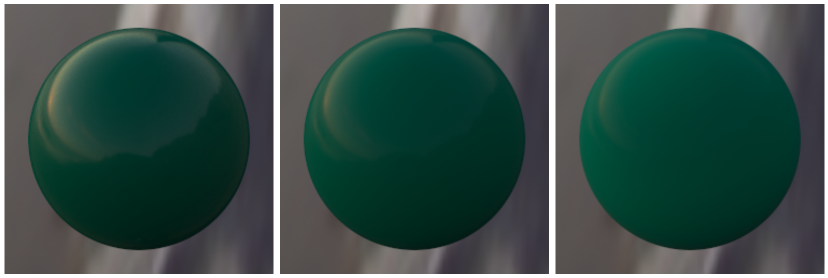
Edits for the brdf chrome-steel of the attribute strength of reflections

Edits for the brdf delrin of the attribute glossiness

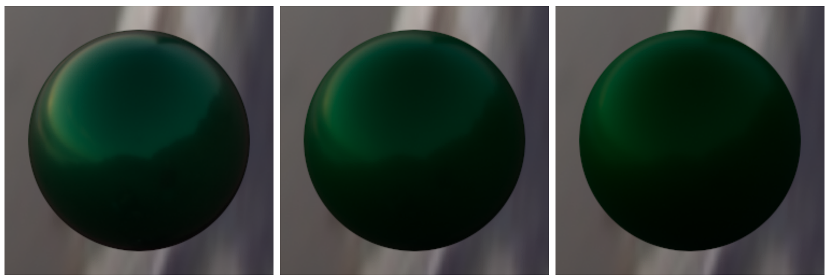


Edits for the brdf gold-metallic-paint3 of the attribute brightness

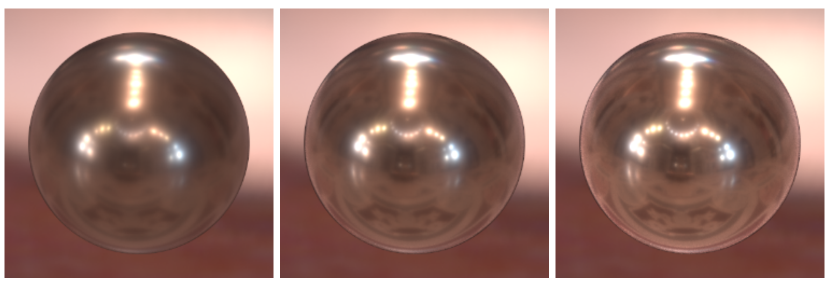
Figure S.15: *Edits of different attributes for a variety of BRDFs ordered alphabetically.*



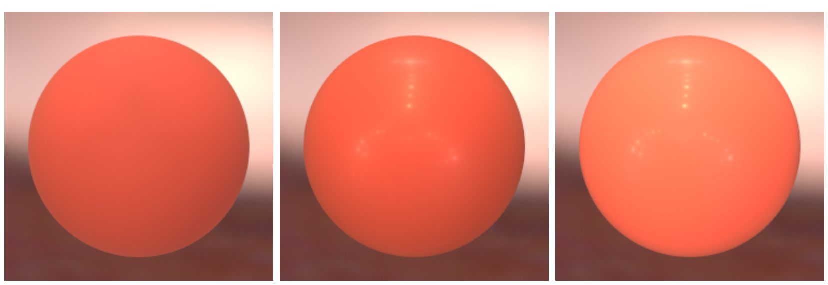
Edits for the brdf green-acrylic of the attribute rubber



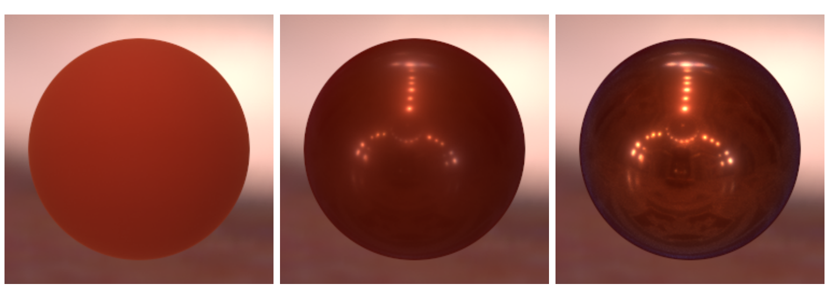
Edits for the brdf green-metallic-paint2 of the attribute rubber



Edits for the brdf nickel of the attribute sharpness of reflections

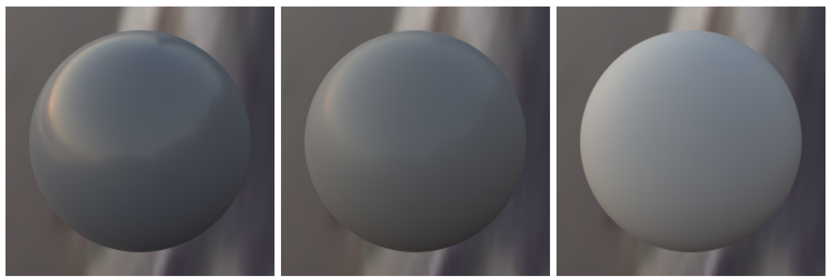


Edits for the brdf pink-plastic of the attribute ceramic

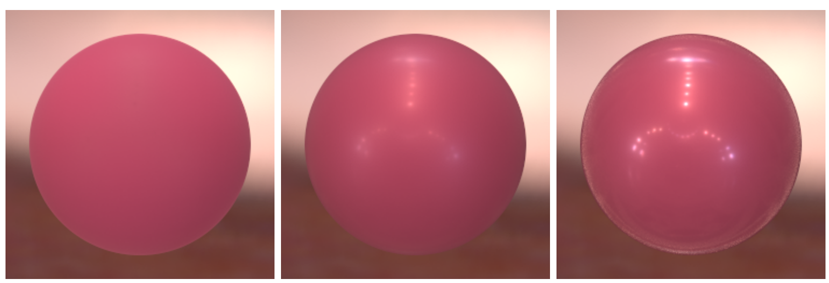


Edits for the brdf red-fabric2 of the attribute metalllic

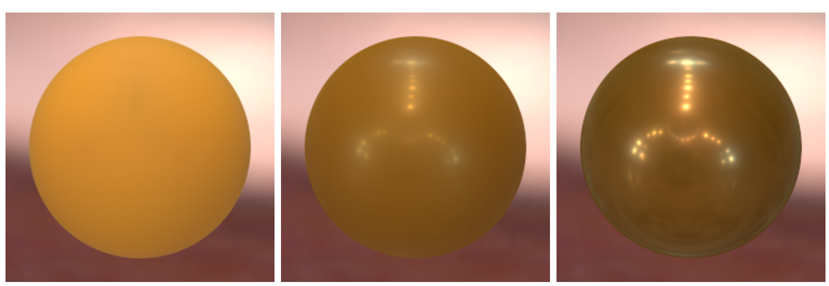
Figure S.16: *Edits of different attributes for a variety of BRDFs ordered alphabetically.*



Edits for the brdf teflon of the attribute matte



Edits for the brdf violet-rubber of the attribute plastic



Edits for the brdf yellow-paint of the attribute metalllic

G Stimuli

We show here the full BRDF database we use, which consists of 94 BRDFs from the MERL database [Matusik et al. 2003] plus 306 new BRDFs which we synthesize as explained in Sec. 3 of the main paper.

Figure S.17: *Stimuli of our experiments including 94 brdfs from the MERL database and 306 new generated brdfs.*

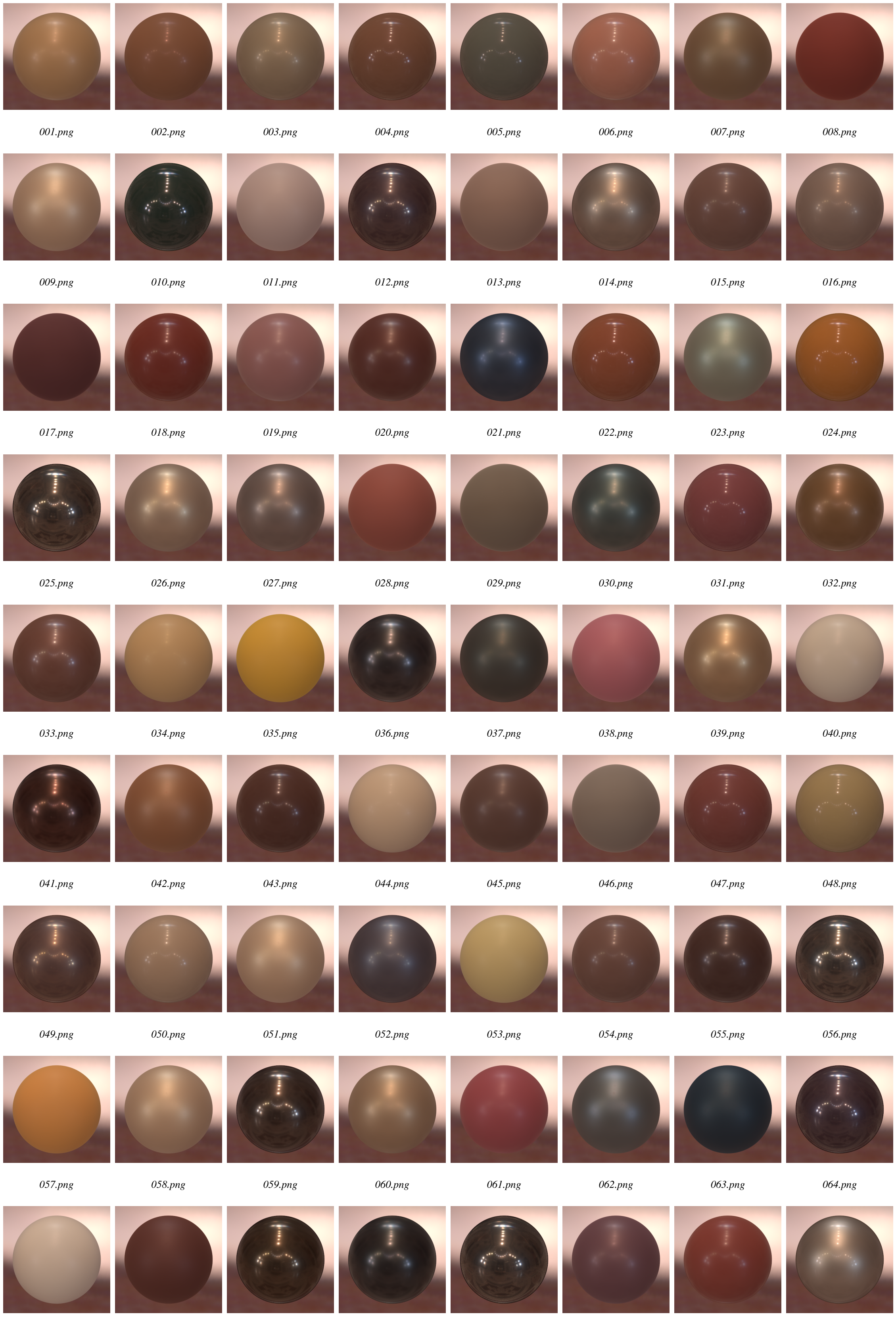


Figure S.18: *Stimuli of our experiments including 94 brdfs from the MERL database and 306 new generated brdfs.*

Figure S.19: *Stimuli of our experiments including 94 brdfs from the MERL database and 306 new generated brdfs.*

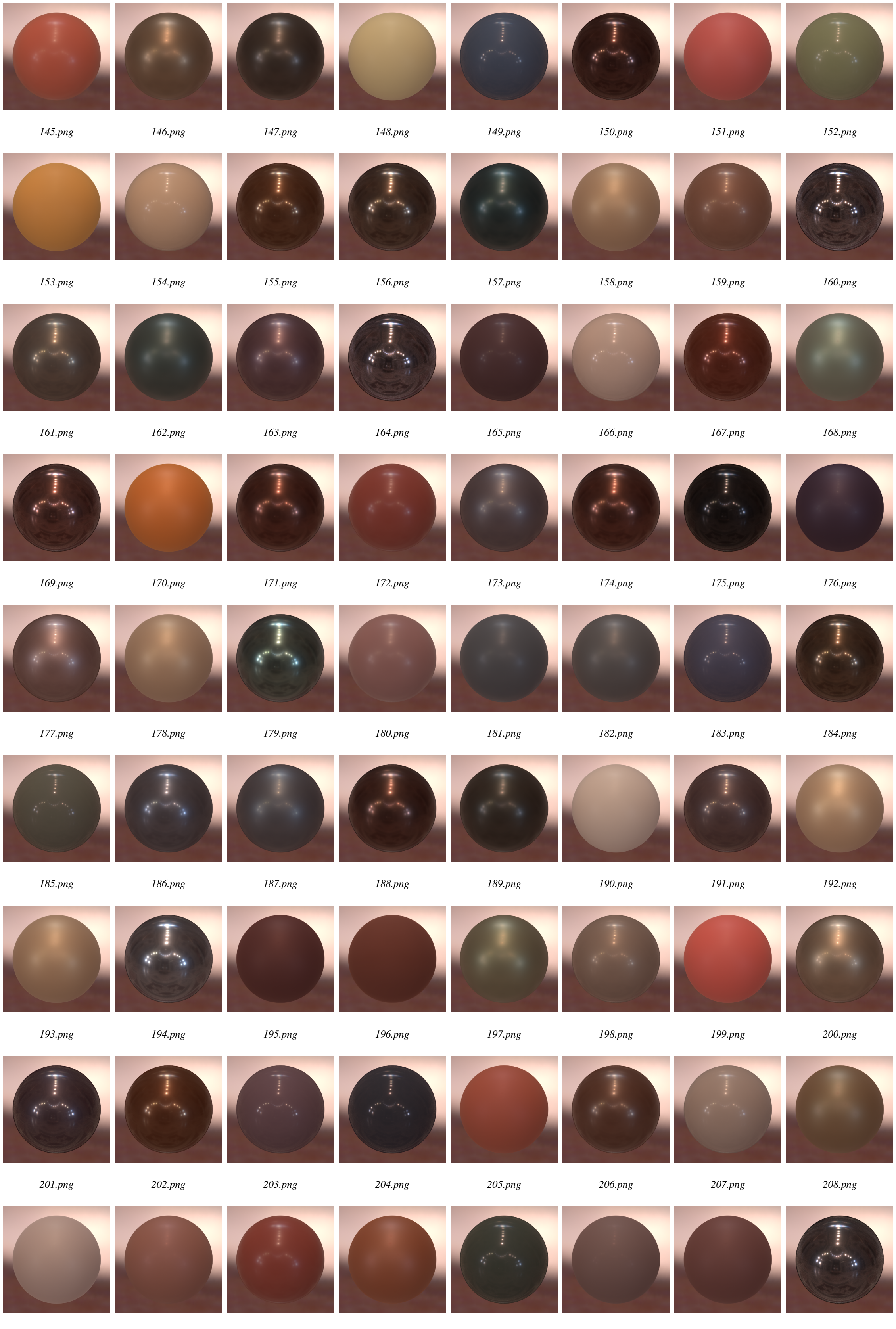


Figure S.20: *Stimuli of our experiments including 94 brdfs from the MERL database and 306 new generated brdfs.*

Figure S.21: *Stimuli of our experiments including 94 brdfs from the MERL database and 306 new generated brdfs.*

Figure S.22: *Stimuli of our experiments including 94 brdfs from the MERL database and 306 new generated brdfs.*

H Attribute lists

We compiled an extensive list of appearance attributes from previous works in industry and academia. Additionally, seven subjects were asked to provide, for each of our 60 stimuli, at least four attributes that described its appearance, using their own words. We then joined the two lists and reduced the number of entries by clustering semantically equivalent attributes; from this we obtained the following initial list of 28 appearance attributes:

- Plastic-like
- Rubber-like
- Mirror-like
- Metallic-like
- Ceramic-like
- Fabric-like
- Acrylic-like
- Pearlescent
- Velvety
- Organic
- Golden
- Silver
- Polished
- Varnished
- Chromed
- Coated
- Opaque
- Soft
- Matte
- Shiny
- Rough
- Strength of reflections
- Sharpness of reflections
- Tint of the Specular
- Sheen
- Tint of the sheen
- Haze
- Specular Gloss

The initial list of attributes was reduced to be manageable. To do this, we devised an experiment in which subjects had to establish, for each stimulus shown, whether each of the attributes applied to the material or not. The outcome of this experiment (Exp. 1 described in Sec. A.2 in this document) was the following list of perceptual attributes:

- Plastic-like
- Rubber-like
- Metallic-like
- Fabric-like
- Ceramic-like
- Soft
- Hard
- Matte
- Glossy
- Bright
- Rough
- Tint of reflections
- Strength of reflections
- Sharpness of reflections

References

ADELSON, E. H. 2001. On seeing stuff: the perception of materials by humans and machines. In *Photonics West 2001-electronic imaging*, International Society for Optics and Photonics, 1–12.

BURLEY, B. 2012. Physically based shading at Disney. In *ACM SIGGRAPH Courses*.

CHOUDHURY, A. K. R. 2014. *Principles of Colour and Appearance Measurement*. Woodhead Publishing.

DEBEVEC, P. 1998. Rendering Synthetic Objects Into Real Scenes: Bridging Traditional and Image-Based Graphics With Global Illumination and High Dynamic Range Photography. 189–198.

EUGENE, C. 2008. Measurement of total visual appearance: a CIE challenge of soft metrology. Tech. rep., 12th IMEKO TC1-TC7 joint Symposium on Man, Science & Measurement.

FISHER, R. A. 1922. On the interpretation of chi square from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society 85*, 87–94.

FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision 3*, 347–368.

FLEMING, R. W., WIEBEL, C., AND GEGENFURTNER, K. 2013. Perceptual qualities and material classes. *Journal of Vision 13*, 8, 9–9.

HUNTER, R. S., AND HAROLD, R. W. 1987. *The Measurement of Appearance (2nd Edition)*. Wiley.

JARABO, A., WU, H., DORSEY, J., RUSHMEIER, H., AND GUTIERREZ, D. 2014. Effects of approximate filtering on the appearance of bidirectional texture functions. *IEEE Transactions on Visualization and Computer Graphics 20*, 6.

KERR, W. B., AND PELLACINI, F. 2010. Toward evaluating material design interface paradigms for novice users. *ACM Trans. Graph. 29*, 4 (July), 35:1–35:10.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. A data-driven reflectance model. *ACM Trans. Graph. 22*, 3 (July), 759–769.

NGAN, A., DURAND, F., AND MATUSIK, W. 2006. Image-driven Navigation of Analytical BRDF Models. In *Proc. of EGSR'06*.

PEARSON, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine 50*, 5, 157–175.

SCHWARTZ, G., AND NISHINO, K. 2013. Visual material traits: Recognizing per-pixel material context. In *ICCV Workshop on Color and Photometry on Computer Vision*, 883–890.

WESTLUND, H. B., AND MEYER, G. W. 2001. Applying appearance standards to light reflection models. In *Proc. of SIGGRAPH '01*, 501–510.

WILLS, J., AGARWAL, S., KRIEGMAN, D., AND BELONGIE, S. 2009. Toward a perceptual space for gloss. *ACM Trans. Graph. 28*, 4 (Sept.), 103:1–103:15.