CGI2018 manuscript No.
(will be inserted by the editor)

Evaluating Physical and Rendered Material Appearance

Jiří Filip · Martina Kolafová · Michal Havlíček · Radomír Vávra · Michal Haindl · Holly Rushmeier

April 10, 2018

Abstract Many representations and rendering techniques have been proposed for presenting material appearance in computer graphics. One outstanding problem is evaluating their accuracy. In this paper, we propose assessing accuracy by comparing human judgements of material attributes made when viewing a computer graphics rendering to those made when viewing a physical sample of the same material. We demonstrate this approach using 16 diverse physical material samples distributed to researchers at the MAM 2014 workshop. We performed two psychophysical experiments. In the first experiment we examined how consistently subjects rate a set of twelve visual, tactile and subjective attributes of individual physical material specimens. In the second experiment, we asked subjects to assess the same attributes for identical materials rendered as BTFs under point-light and environment illuminations. By analyzing obtained data, we identified which material attributes and material types are judged consistently and to what extent the computer graphics representation conveyed the experience of viewing physical material appearance.

Keywords material appearance · rendering · BTF · perception · psychophysics · MAM2014

1 Introduction

An efficient transfer of realistic appearance of real-world materials to virtual reality has been one of the ultimate challenges of computer graphics. In entertainment and storytelling, a particular material appearance may be selected to evoke a viewer response. For industrial design, material appearance may be rendered to preview a physical design. In either scenario, there is no reliable measure of the visual fidelity of the virtual material's appearance. Moreover, the appearance needs to be faithful with changing view and illumination conditions. In other words, we are looking for visual equivalence but not in a sense of comparison of two virtual representations as introduced in [23], but rather between physical materials and their virtual representations.

Fig. 1 Photos of 16 materials from MAM 2014 dataset.

In this paper, we propose using human judgements of material attributes to assess the fidelity of virtual material appearance. To explore this idea we use phys-

J. Filip, M. Kolafová, M. Havlíček, R. Vávra, M. Haindl
The Czech Academy of Sciences, Institute of Information Theory and Automation, The Czech Republic

H. Rushmeier
Yale University, USA

ical specimens of 16 materials presented at the Workshop on Material Appearance Modelling (MAM) 2014 [25] (sample set No.26) as shown in Fig. 1. The collection includes one natural material *mica* and the rest are man-made utility materials such as sandpaper, fabric, tile, burlap, plastic-weave, etc. This set represents a diverse selection of materials the subjects are familiar with from daily life. They vary in many physical aspects, e.g., natural or man-made origin, opacity or translucency, height, texture, structure, etc.

First, we perform a psychophysical study to identify the main perceptual attributes of these physical material specimens. We selected twelve visual, tactile and subjective attributes to be analyzed. The outcome of this study is an understanding of the consistency of judging these attributes for the range of materials. For a second study we acquired all these materials as bidirectional texture functions (BTFs) [3]. We selected BTFs since they capture samples with height and opacity variations, and include non-local effects such as occlusions, masking, subsurface scattering and inter-reflections. We conducted the second study to find how well renderings using BTF preserve appearance of the physical samples. The results of the studies show the consistency of human judgement of material appearance attributes, and how these observations can be used to assess the effectiveness of computer graphics presentations of material appearance.

Contributions of this paper are:

- An openly accessible database of BTF measurements from the MAM dataset.
- A proposal for evaluating the accuracy of material appearance rendering.
- A psychophysical study of how consistently people judge appearance attributes of material samples.
- A study of how judgements of material attributes from standard BTF sample renderings compare to judgements of real material samples.

The paper contents are as follows: Section 2 summarizes related work. Section 3 overviews the proposed approach and Section 4 introduces selected perceptual attributes. Section 5 describes the perceptual assessment of real samples, while Section 6 evaluates their renderings. Section 7 compares the results of both studies and Section 8 summarizes the paper.

2 Prior Work

Understanding human perception of materials is one of primary challenges of research fields ranging from psychology to computer graphics applications [1,8].

A large body of research has been devoted to identify links between material perceptual attributes and physical properties [9] and to establish a link between texture perceptional and computational spaces [24,15,20].

Although, originally researchers used textural images [2] to represent materials, this changed with advent of illumination- and view-dependent scanning facilities capturing bidirectional reflectance distribution functions BRDF [22], spatially-varying bidirectional reflectance distribution functions SVBRDF or BTF [3] representations. Various perceptual studies related to these representations have been performed. For example, the impact of object shape on the perception of material appearance represented by an isotropic BRDF model was studied in [30]. In other work [5], a psychophysical study demonstrated that a reduction of BTF data without compromising visual quality, is directly related to BTF data variance and to the complexity of illumination and object geometry. BTF data were also used in [6] to investigate the effect of shape and texture on subjects' visual attention. The authors concluded that a flat textured surface receives only half of the fixations in comparison with shaped surface, and that average local variance of a curved surface texture can predict observers' gaze attention to both texture and its underlying geometry, i.e., the more higher frequencies and regularities are present in the material texture, the easier it is to identify possible differences, requiring a lower number of shorter fixations. The perceptual effects of BTF filtering were analyzed in spatial and angular domains in [13]. It was shown that preserving contrast is more important in static than in dynamic images and indicates that greater levels of spatial filtering are possible for animations. Filtering can be performed more aggressively in the angular domain than in the spatial domain. Mylo et al. [21] exploited a link between certain perceived visual properties of a material and specific bands in its spectrum of spatial frequencies. They applied the results for editing of BTFs that produces plausible results.

Researchers have derived a number of attributes in psychophysical studies to evaluate as to what extent they are possessed by a virtual material in a form of texture [19], BRDF [17,27] or by a real specimen [16].

Several different techniques have been proposed to assess the quality of renderings. Meseth et al. [18] used virtual stimuli to psychophysically compare performance of material photographs, BTF renderings, and flat textures modulated by BRDFs under identical illumination conditions. Ramanarayanan et al. [23] developed metrics that predict the visual equivalence of rendered objects under warping and blurring of illumination and warping of object surfaces. Havran et al. [10] developed a surface, optimized to a high coverage of illumination

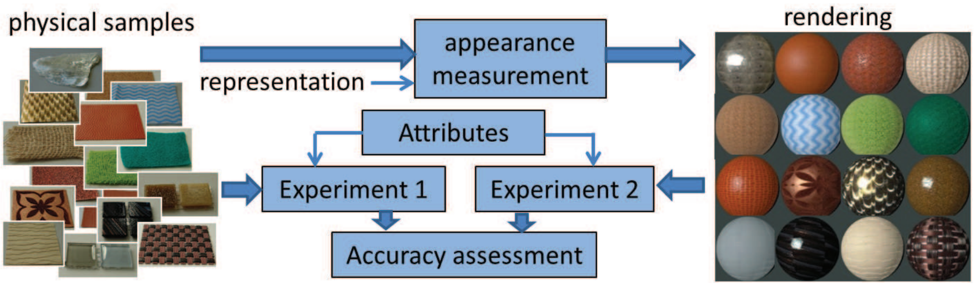

Fig. 2 An overview of the proposed material appearance accuracy assessment approach.

and viewing angles, that can be used for a single image comparison of various material representations.

Although many studies were done analyzing material appearance either of real materials or their digital representations, we are not aware of any study systematically comparing the perception of physical and virtual materials from a set of widely available samples. This is important as it allows researchers to follow up and compare or complement our results with their future findings.

3 Proposed Appearance Evaluation Method

When it comes to the evaluation of material appearance accuracy, standard computational assessment methods are often unreliable, even when they use photographs of the material. Many materials, especially in the consumer market, are carefully designed to deliver a specific user experience that often cannot be reproduced using a single photograph. Therefore, industrial designers have started to use virtual light booths, allowing the direct comparison of a real product and its digital representation rendered under strictly controlled lighting and viewing conditions. In this paper, we continue in this direction and assume that material appearance is accurately rendered if people would draw the same conclusions of material attributes based on renderings that they would make looking at the real, physical materials. A basic scheme of the approach is shown in Fig. 2. Therefore, we suggest two psychophysical visual studies. The first assesses subjects' experience from real material specimens, while the second does the same with virtual representation, in our case BTF, of the same materials. To obtain information of user experience from the material, we asked subjects to rate a preselected group of perceptual attributes described in the following section.

4 Selected Material Attributes

Our goal is to analyze as many aspects of the tested materials as possible; however, as the number of assessed materials is relatively high, we chose a lower number of attributes to make the study more workable for human subjects. We intentionally avoided using high-level class predictors, e.g., plastic-like, as used in [17] and [27] and focused on important low-level attributes identified in prior work.

Our goal was a comprehensive multi-modal analysis of the materials. For the visual characteristics we focused on general textural attributes [29, 24, 19] and selected six of them spanning reflectance, color, and structural properties of materials: *glossiness, colorfulness, directionality, diversity, graininess*, and *regularity*. As our study uses a physical material specimens, we extended our attributes also to the tactile and subjective domain similar to [16]. Tactile attributes are included for other researchers who are interested in evaluating tactile interfaces and they are not used in the second experiment. In the tactile domain we focus on attributes having clear physical interpretation: *hardness, roughness* and *height*. In the subjective domain we found it interesting to assess the authenticity and quality of materials and selected: *genuineness, quality,* and *attractiveness*. Thus our final list of attributes includes six visual, three tactile, and three subjective attributes as shown in Tab. 1.

Table 1 The twelve material attributes evaluated within the experiment.

id	attribute	ranges [1–9]	
Visual			
L	**glossiness**	matte	glossy
B	**colorfulness**	single-color	multiple-colours
S	**directionality**	no-directionality	directional
R	**diversity**	simple	complex
H	**graininess**	smooth	rough
P	**regularity**	random	regular
Tactile			
T	**hardness**	soft	stiff
D	**roughness**	smooth	rough
V	**height**	flat	deep
Subjective			
O	**genuineness**	man-made	natural
K	**quality**	ordinary	luxurious
A	**attractiveness**	usual	attractive

5 A psychophysical assessment of physical specimens

In our study, we asked subjects to rate individual preselected attributes of physical specimens of 16 materials as shown in Fig. 1. For this experiment we paid particular attention on analysis how consistently the subjects rate individual perceptual attributes across the set of materials.

5.1 Participants and Experimental Procedure

Twenty two paid subjects performed the experiment. There were 12 male and 10 female participants aged from 20 to 60. All subjects had normal or corrected to normal vision and all were uninformed with respect to the purpose and design of the experiment.

All samples of a typical size of 6×6 cm were laid down on the table as shown in Fig. 3. The table was

Fig. 3 Analysed materials as seen in the psychophysical experiment.

positioned near a window to avoid direct sunlight. Subjects also had available a LED point-light with a turntable that they could use freely for a more detailed examination of material properties. At the beginning of the session the meaning of individual attributes were explained to the subject and she/he could freely review all samples. Then the subject assessed individual attributes for all materials using a score ranging from 1 to 9 (representing the lowest and the highest intensity of the attribute, see Tab. 1), while the range should represent only the samples within the study.

Subjects always started with assessment of visual attributes. To avoid confusion between visual and tactile attributes, subjects were not allowed to touch the sample when assessing visual attributes. Also subjects were prohibited from turning samples over so as to prevent identifying the authenticity of materials. There were no strict time limits, and subjects finished their evaluation of all materials in between 10 and 30 minutes.

5.2 Results

First of all, we analyzed the reliability of subjects' responses by means of the Krippensdorff's alpha [11] α_K – a statistical measure of the agreement generalizing several known statistics. The key requirement is agreement observed among independent observers. Output $\alpha_K = 1$ represent unambiguous indicator of reliability, while 0 not. Results given in Fig. 4 demonstrate subject agreements for individual attributes (a) and tested materials (b). The highest agreement ($\alpha_K = 0.55 - 0.65$) was achieved for attributes *L–glossiness, T–hardness, D–roughness*, while the other attributes had similar values ($\alpha_K = 0.3{-}0.4$). The lowest agreement ($\alpha_K = 0.25$) was obtained for *H–graininess.*

Materials with the highest agreement (α_K over 0.5) are *04–burlap, 11–silver-gold, 13–glass-tile, 14–blue-black-gold-tile.* Surprisingly low agreement ($\alpha_K = 0.006$) was recorded for material *09–basketball.*

(a)

(b)

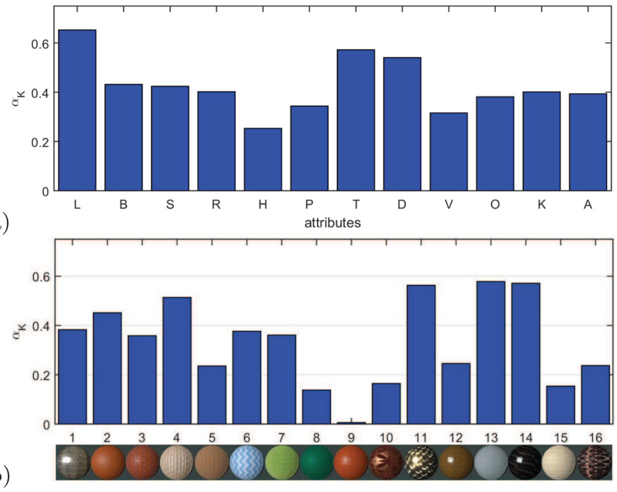

Fig. 4 Statistical analysis of the agreement across subjects in the real-samples experiment for individual attributes using Krippensdorff's alpha across: (a) attributes and (b) materials.

The data were also analyzed using a single factor repeated measures ANOVA demonstrating that all p-values are below confidence level 0.05 and favoring an alternative hypothesis that means for individual materials are draw from statistically different populations at significance level 95%. We also successfully verified that data normality, the basic ANOVA assumption, using the Shapiro-Wilk parametric test.

For aggregation of the subjects' values, we applied the mean opinion score (MOS) obtained as an average rating across all subjects. This is a standard methodology for subjective quality assessment used especially in the audio and video industries, and recommended by standard international organizations such as the ITU

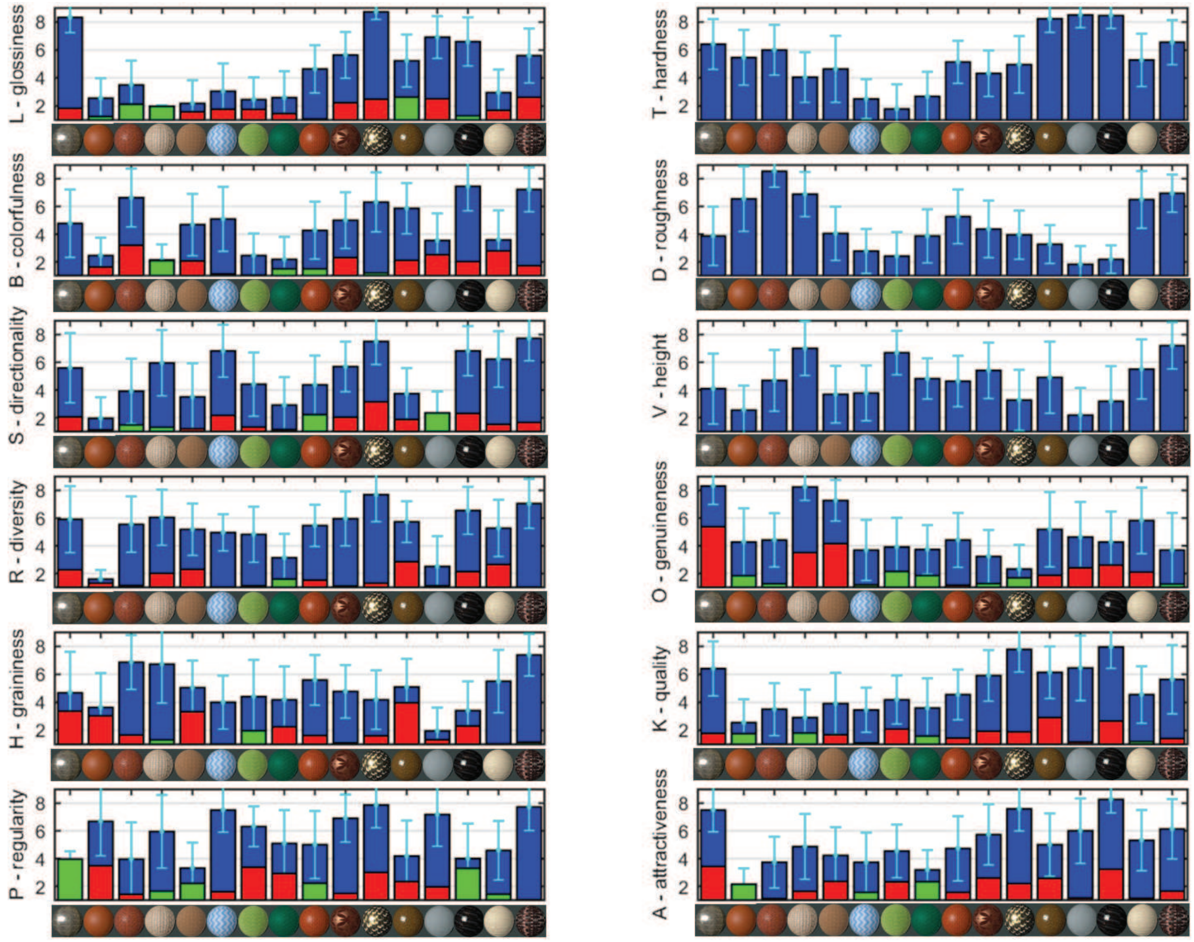

Fig. 5 Perceptual attributes for all materials obtained as mean values across tested subjects. Errorbars represent standard deviation across subjects. The green and red subbars represent positive and negative deviations, respectively, from the original values when the attributes were assessed from materials' rendering using BTFs. Note that tactile attributes (T, D, V) were not assessed this way.

[12] or ISO [14]. We also performed outliers rejection by removing values differing more than 60% of score range from MOS (a total 29 values out of 4224). MOS in range 1-9 for individual perceptual attributes and tested materials are shown as blue bars in Fig.5. Errorbars in the graphs represent the standard deviation across subjects. On the left are graphs representing visual attributes, while on the right are graphs for tactile and subjective attributes.

Next, to analyze attributes relationship, we computed Pearson and Spearman correlation coefficients between mean opinion scores of all attributes across all tested materials. The results were similar (mean and maximum differences 0.014 and 0.237 respectively) and thus we report only the Pearson correlation values ρ in Fig. 6. Each cell contains both color-coded and numerical correlation value and p-value.

Colors toward red represent a positive correlation while those towards blue represent negative ones. Each pair of attributes in the figure show correlation coef-

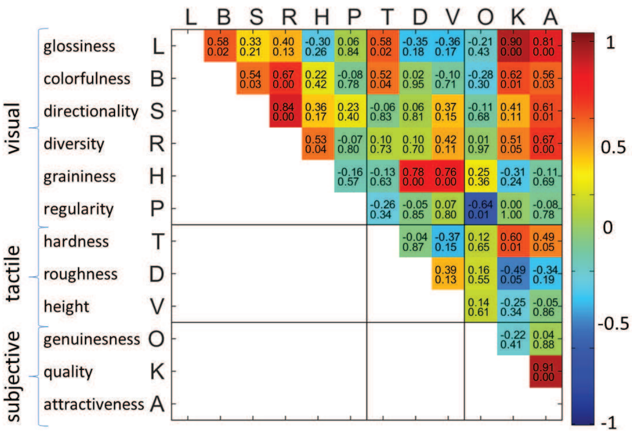

Fig. 6 A Pearson correlation matrix between the tested attributes. Colors toward red represent positive correlation while those towards blue represent negative ones. In each cell the top number specifies correlation value and the bottom one specifies corresponding p-value.

ficient (-1,1) at the top of the cell and corresponding p-values at its bottom. One can find strong sta-

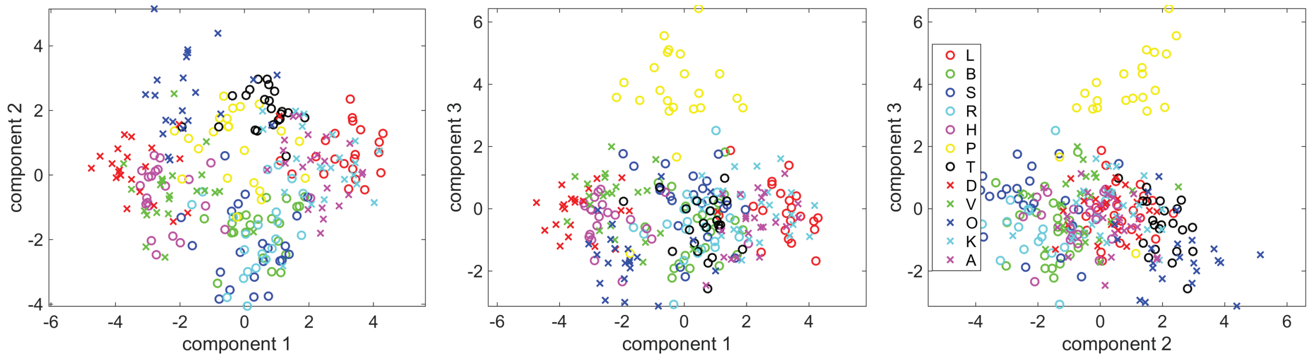

Fig. 7 Multivariate analysis of material attributes showing their separation of the first three canonical components.

tistically significant positive correlations ($\rho > 0.7$, p-value< 0.01) between:

- *glossiness* and *attractiveness* ($\rho = 0.81$),
- *glossiness* and *quality* ($\rho = 0.90$),
- *quality* and *attractiveness* ($\rho = 0.91$),
- *graininess* and *roughness* ($\rho = 0.78$),
- *graininess* and *height* ($\rho = 0.76$),
- *directionality* and *diversity* ($\rho = 0.84$).

The strongest negative correlation was found $\rho = -0.64$ between *regularity* and *genuineness*.

To further investigate relationships of attributes, we performed a one-way Multivariate Analysis of Variance (MANOVA) for comparing the multivariate means of subjects' responses grouped by attributes. This analysis tests the null hypothesis that the means of attributes are the same n-dimensional multivariate vector, and that any difference observed in the data is due to random chance. We estimated dimensionality is $d = 8$ at a significance level 5% and can reject the null hypothesis, and thus expect the data means lie in 8-dimensional manifold. As a product of the multivariate analysis we also obtain a dendrogram depicting attributes proximity as shown in Fig. 8. The clusters are computed by applying the single linkage method to the matrix of Mahalanobis distances between the groups means.

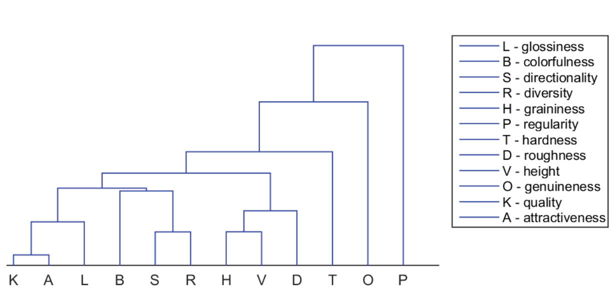

Fig. 8 Multivariate analysis – dendrogram illustrating proximity of attributes.

These results support the findings of the correlation analysis in Fig. 6. The dimensionality $d = 8$ and dendrogram suggest that the most distinct attributes in order of difference are *P–regularity, O–genuineness, T–hardness, B–colorfulness, D–roughness*, while the rest attributes are due to their proximity grouped three groups: (1) *K–quality, A–attractiveness, L–glossiness*, (2) *S–directionality* and *R—diversity*, (3) *H–graininess* and *V—height*.

Further, multivariate analysis provides us with so called canonical vectors, which are linear combinations of original variables, chosen to maximize separation between groups. Fig. 7 depicts dependencies of the first three canonical vectors, providing an insight on variability and overlap of subjects' responses to individual attributes (each attribute marker in the graphs corresponds to one of the subjects).

Finally, in Fig. 9 we show pairs of materials having the highest and the lowest perceived response to individual attributes.

Discussion – The study has revealed that although none of the attributes or materials was judged consistently, we learned which of the attributes are the most reliable as well as how they are mutually related. The most consistent data were obtained for *L–glossiness*, *T–hardness* and *D–roughness*. We assume that this is due to the fact that the meaning of these attributes is easy to understand, and they are even commonly instrumentally measured as physical properties of materials. Another important finding is that attributes *K–quality* and *A–attractiveness* have almost identical perceptual responses. The same can be said about pairs *S–directionality*, *R–diversity* and *H–graininess*, *V–height*. However, here we should note that while for *S,R* the response is possibly due to missing materials with random diverse structure and the perception of diversity comes mainly from regular patterns, for *H,V* it is due to that subjects judged height as span of height variability within structure grains, e.g., relatively thick tiles samples (materials 12,13,14) received lower judgement of this attribute than thin crinkle paper (material 15).

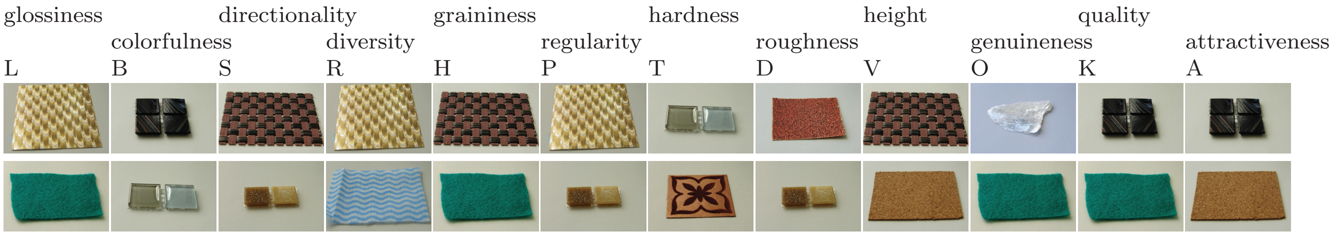

Fig. 9 Each attribute with materials having the highest (the top row) and lowest (the bottom row) average attribute values.

6 A Psychophysical assessment of renderings

In the second experiment, we evaluated the selected material attributes on rendered images which reproduced the appearance of physical material samples by means of a bidirectional texture function (BTF).

6.1 BTF Data Acquisition and Processing

For BTF data acquisition, we used a gonioreflectometer [7] to capture the appearance of tested specimens. This setup consists of the measured sample on a rotating stage, and two independently controlled arms with a camera (one axis) and a light source (two axes). This allows for flexible measurements of nearly arbitrary combinations of illumination and viewing directions. Although camera view occlusion by the arm with the light source may occur, it can be analytically detected, and in most cases, alternative positioning is possible. The angular repeatability of light and camera positioning is 0.03 degrees across all axes. The inner arm holds the LED light source 1.1 m from the sample and produces a narrow and uniform beam of light. The outer arm holds an industrial, full-frame 16Mpix RGB camera AVT Pike 1600C. The sensor's distance from the sample is 2 m. In our experiments, we used a lens achieving maximum resolution of 353 dpi (i.e., $72\mu m$/pixel). The hemispheres of incoming and outgoing directions were uniformly sampled by means of 81 directions [26] giving a total of $81\times81 = 6\,561$ HDR images, where each image was taken for a unique bidirectional pair of camera and light positions.

Once the BTF data are captured, we seek a minimal spatial sample whose repetition conveys an original material's appearance. To achieve this, the image stitching algorithm [28] is applied, resulting in a seamless spatially repeatable tile. Note that this technique cannot in principle produce seamless repetition for materials with a non-regular structure or low frequency features, e.g., *mica* or *blue-black-gold-tile*.

Each captured BTF generates thousands of HDR images. As the handling of such a large number of files is inconvenient due to time-demanding IO operations, we introduce a new straightforward format encapsulating all images into one binary file together with meta-information (e.g., image size, color-space, spatial resolution, etc.). The format is called BIG - Big Image Group and the recommended file extension is .big. The data are stored in a uncompressed form so as they can be quickly accessed using several application-dependent query functions. We implemented the IO functions in C++ as publicly available software (http://btf.utia.cas.cz). The format allows significantly faster and more convenient manipulation of the measured data. We use this format for storing and sharing captured data with the greater research community. More details on data format are given in the Appendix of this paper.

Table 2 Sizes of image tiles and data files.

material	tile size [pix.]	data size [GB]
01-mica	160x160	1.88
02-sand-fine	112x106	0.89
03-sand-coarse	130x116	1.11
04-burlap	154x126	1.42
05-cork	112x110	0.93
06-towel	197x 87	1.26
07-green-cloth	110x155	1.25
08-green-felt	110x108	0.89
09-basketball	103x 98	0.76
10-flocked-paper	395x389	11.27
11-silver-gold	236x253	4.38
12-brown-tile	169x160	1.98
13-glass-tile	100x107	0.80
14-blue-black-gold-tile	224x225	3.70
15-crinkle-paper	224x190	3.12
16-basket-weave	190x224	3.12

Due to the variable size of structural patterns of materials, the size of a stored BTF tile varies between 100×100 and 400×400 pixels. Thus the captured and tiled BTF data stored in the BIG format range in size between 0.8 and 11.3 giga-bytes for each material from the MAM2014 dataset as shown in Tab. 2.

Materials visualized from captured BTFs are shown for point-light illumination in Fig. 10. Note that rendering is obtained from the raw data and no data compression is applied.

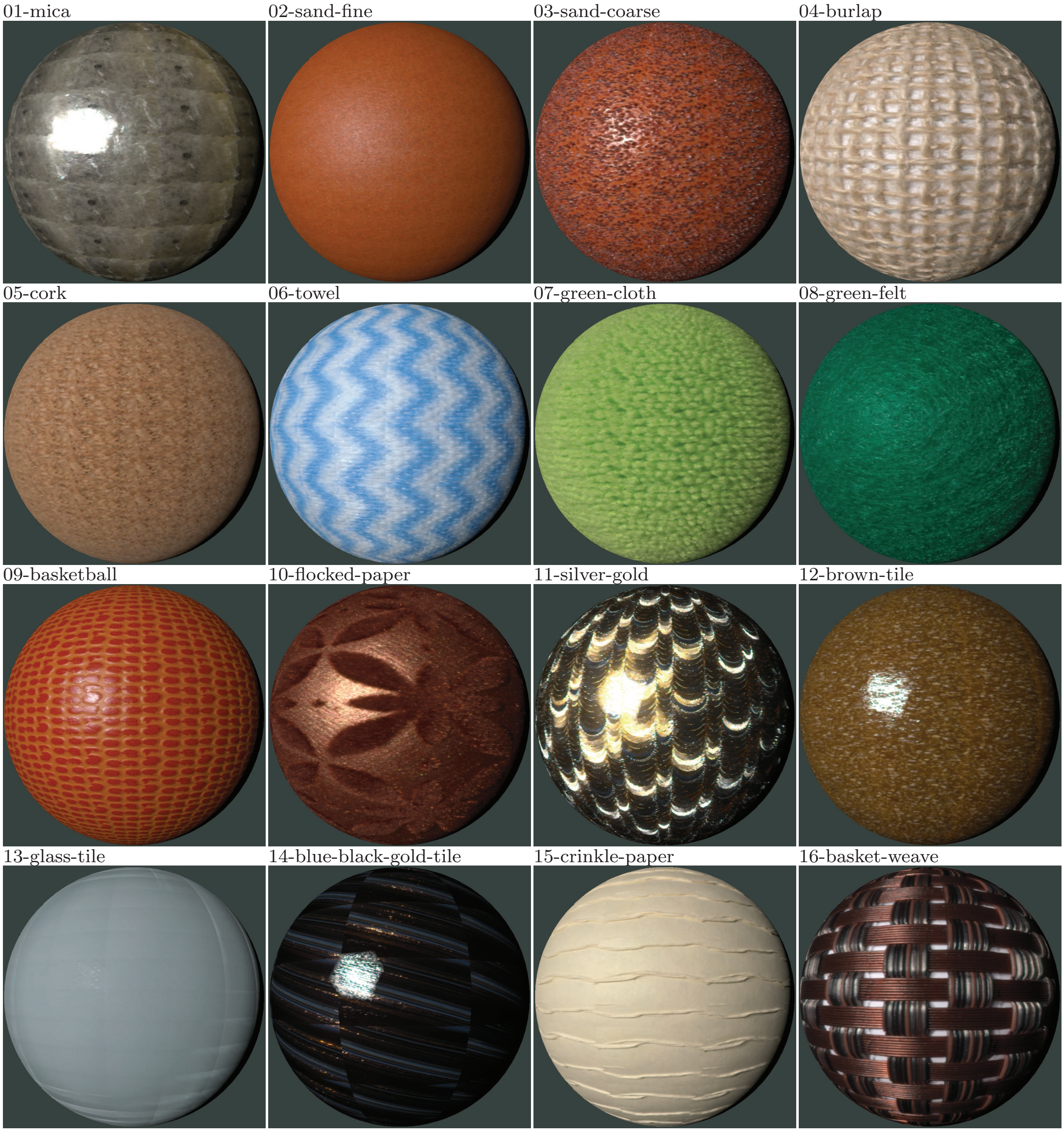

Fig. 10 Materials from MAM 2014 dataset (set 26) captured as BTFs rendered for a point-light.

6.2 Participants and Experimental Procedure

Once the BTFs of materials were captured, we moved to the assessment of their perceptual attributes. To obtain a sufficient number of anonymous volunteers, we performed a web-based study, i.e., run in an uncontrolled environment and at various screens, while subjects were advised to run it in a dim room environment. The materials are not rendered in the same way as the materials were presented physically, since a flat surface would provide low angular variations and becomes unattractive for observers [6]. Instead, we mapped materials on a spherical shape as has become standard for presenting BTF data. We render appearance on spheres directly from the raw BTF data and used a single point light source, best exposing high frequency features [4], as well as environment illumination (fixed orientation of *grace* environment). We combine this information in stimuli images as shown in example in Fig. 11. To allow subjects anchoring perceptual scales, each stimuli image features on its top an overview of thumbnails of all assessed materials for point-light illumination. As subjects could assess the materials only visually, we removed the three tactile attributes from the study. Individual attributes were evaluated by between 30 and 43 participants (the number of finished sessions varies across attributes). Subjects were advised to perform the study in full-screen mode and in a dark environment.

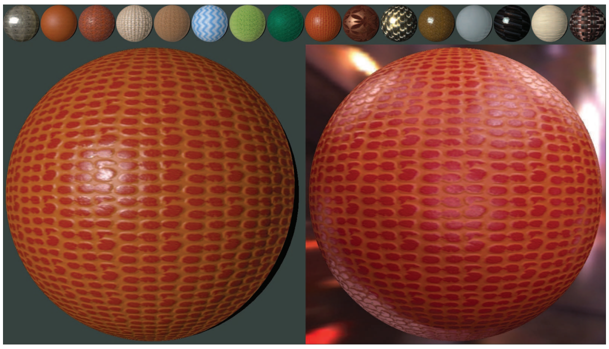

Fig. 11 An example of material rendering stimulus image showing material rendering for point-light and environment illumination side-by-side with thumbnails of all materials under point-light illumination.

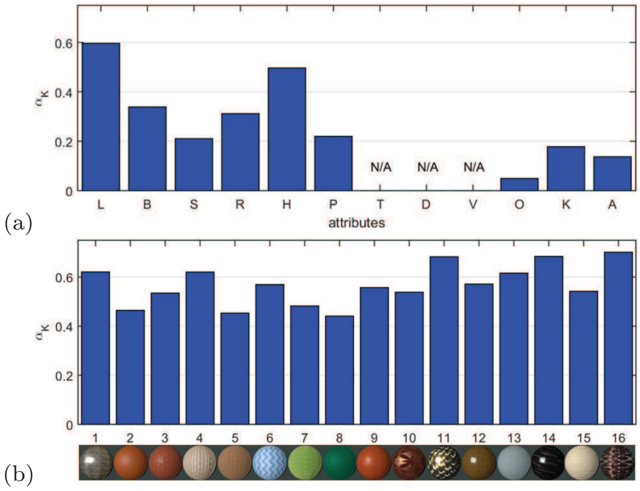

Fig. 12 Statistical analysis of the agreement across subjects in web-based experiment for individual materials using Krippensdorff's alpha across: (a) attributes and (b) materials.

6.3 Results

First, we again assessed data reliability using Krippensdorff's alpha as shown in Fig. 12. In contrast to the values in the first experiment, better agreement was achieved for individual materials rather than for individual attributes. Attributes having $\alpha_K > 0.5$ were only *L–glossiness* and *H–graininess*. Low agreement, i.e. $\alpha_K < 0.2$, was recorded for *S–directionality*, *P–regularity* and all subjective attributes. In contrast, the agreement for individual materials was relatively high for all of them, i.e. α_K between 0.45 and 0.75.

The obtained mean opinion scores for individual attributes and materials are compared to results from the first experiment in Fig. 5. Differences from the first experiment are shown as green and red subbars indicating higher and lower values respectively. The standard deviations across subjects are not shown, but the values resembled those in the first experiment.

Discussion – The second experiment has shown that assessment of any representation is influenced by several aspects. Users in our experiment could not move with the material and relied only on a span of viewing and illumination angles as defined by the illumination environment and surface geometry. This might be insufficient as the appearance of some materials changes substantially with regards to their orientation as shown in Fig. 13. Also the sphere used as test geometry could look unnatural for some materials that are typically available as flat, e.g., mica or tiles. Moreover, agreement between subjects across different materials is more balanced for rendered representations (see Fig. 12-b); whereas, it is more scattered for the experiment with real samples (see Fig. 4-b). This might suggest that rendered images in some sense "filtered" some unique material properties.

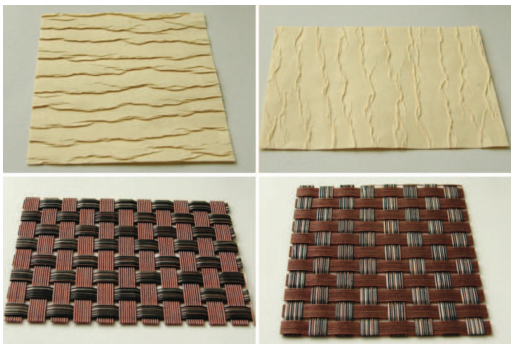

Fig. 13 Two examples of material appearance change for different orientations of a material.

7 Comparison of Physical and Rendered Material Appearance

This section combines data obtained from subjects observing real samples and their digital representations using a BTF. We computed a correlation between mean opinion scores from both experiments as shown in Fig. 14. While the bar's height corresponds to the correlation between the perception of physical samples and corresponding rendered images, their darkness is proportional to data agreement across subjects as obtained by the multiplication of α_K from both experiments. There are separate results for individual attributes and materials, and individual bars include p-values.

(a)

(b)

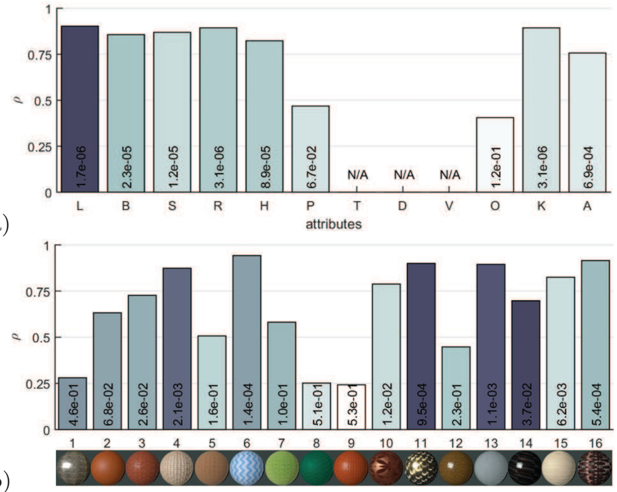

Fig. 14 Peason's correlation coefficients computed between values obtained from assessment of physical samples and their renderings for individual: (a) attributes, (b) materials. Each bar includes correlation p-value and its darkness is proportional to data agreement across subjects.

For attributes (a) we observe comparably high values ($\rho > 0.75$) for most of the attributes with *L–glossiness* having by far best subjects responses agreement. The only exceptions were *P–regularity* and *O–genuineness* where ρ dropped below 0.5. This indicates that a real experience of these attributes was not delivered by means of BTF.

For materials (b) the results are more challenging to interpret. We observe high correlations ($\rho > 0.75$) for man-made materials having an obvious regular structure. In contrast, natural material or materials with random structure have generally lower correlations. As mentioned earlier, the material with a familiar structure *09–basketball* (see Fig. 11) was one of the three materials having $\rho < 0.3$. We hypothesize that subjects were able to identify properties of materials with a clearly regular structure due to their familiarity from everyday-life; therefore, they evaluated their perceptual attributes more consistently than for materials with random structures.

Although we consider these results as one of first steps towards analyzing the perceptual quality of material appearance representations, there are still open questions. For instance, how the perception improves when HDR or stereo displays are used? Or how a user's experience changes when animation is used instead of static images? An important aspect is also an interaction of material appearance with shape and illumination. It is clear that all these aspects should be covered in future research to obtain unequivocal conclusions on the quality of rendered material appearance.

8 Conclusions

This paper assessed the perceptual accuracy of material appearance reproduction by directly comparing it with subject judgement of real physical samples. We analyzed sixteen materials of MAM 2014 dataset and captured their appearance using a bidirectional texture function (BTF) and assessed differences in perception of real material samples and their rendered counterparts. The results suggest that BTF representations conveyed the majority of observed visual attributes; however, there are differences in perception of different material types. The perception of man-made materials with regular structure was better delivered in virtual reproduction than that of natural materials with random structure. To stimulate further work on the measurement and understanding of material appearance, we publicly provide captured and uncompressed BTFs of the MAM 2014 dataset for research purposes.

References

1. Adelson, E.H.: On seeing stuff: the perception of materials by humans and machines (2001)
2. Brodatz, P.: A Photographic Album for Artists and Designers (Brodatz Texture Database). Dover Publications (1966)
3. Dana, K., van Ginneken, B., Nayar, S., Koenderink, J.: Reflectance and texture of real-world surfaces. ACM Transactions on Graphics **18**(1), 1–34 (1999)
4. Filip, J.: Analyzing and predicting anisotropic effects of BRDFs. In: ACM SAP, pp. 25–32 (2015)
5. Filip, J., Chantler, M., Green, P., Haindl, M.: A psychophysically validated metric for bidirectional texture data reduction. ACM Transactions on Graphics **27**(5), 138 (2008)
6. Filip, J., Vacha, P., Haindl, M.: Analysis of human gaze interactions with texture and shape. In: S+SSSPR Workshop, Springer LNCS 7252, pp. 160–172 (2011)
7. Filip, J., Vávra, R., Haindl, M., Zid, P., Krupicka, M., Havran, V.: BRDF slices: Accurate adaptive anisotropic appearance acquisition. In: In proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013, pp. 4321–4326 (2013)
8. Fleming, R.W.: Visual perception of materials and their properties. Vision Research **94**(0), 62 – 75 (2014)
9. Goodman, T.: Measurement of naturalness: Physics and perception. In: Proceedings of the 3rd International Conference on Appearance, pp. 21–24. MIT Press
10. Havran, V., Filip, J., Myszkowski, K.: Perceptually Motivated BRDF Comparison using Single Image. Computer Graphics Forum pp. 1467–8659 (2016)
11. Hayes, A.F., Krippendorff, K.: Answering the call for a standard reliability measure for coding data. Communication methods and measures **1**(1), 77–89 (2007)

12. ITU: ITU-R.REC.P.910. subjective audivisual quality assessment methods for multimedia applications. Tech. rep. (2008)
13. Jarabo, A., Wu, H., Dorsey, J., Rushmeier, H., Gutierrez, D.: Effects of approximate filtering on the appearance of bidirectional texture functions. IEEE Transactions on Visualization and Computer Graphics **20**(6), 880–892 (2014)
14. Keelan, B.: ISO 20462: A psychophysical image quality measurement standard. In: Proceedings of the SPIE, vol. 5294, SPIE 2003, p. 181189 (2003)
15. Long, H., Leow, W.: A hybrid model for invariant and perceptual texture mapping. In: Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 1, pp. 135–138. IEEE (2002)
16. Martín, R., Iseringhausen, J., Weinmann, M., Hullin, M.B.: Multimodal perception of material properties. In: Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, SAP '15, pp. 33–40 (2015)
17. Matusik, W., Pfister, H., Brand, M., McMillan, L.: A data-driven reflectance model. ACM Transactions on Graphics **22**(3), 759–769 (2003)
18. Meseth, J., Müller, G., Klein, R., Röder, F., Arnold, M.: Verification of rendering quality from measured BTFs. In: Third Symposium on Applied Perception in Graphics and Visualization, pp. 127–134 (2006)
19. Mojsilovic, A., Kovacevic, J., Kall, D., Safranek, R., Kicha Ganapathy, S.: The vocabulary and grammar of color patterns. Image Processing, IEEE Transactions on **9**(3), 417–431 (2000)
20. Motoyoshi, I., Nishida, S., Sharan, L., Adelson, E.: Image statistics and the perception of surface qualities. Nature **447**(10), 206–209 (2007)
21. Mylo, M., Giesel, M., Zaidi, Q., Hullin, M.B., Klein, R.: Appearance bending: A perceptual editing paradigm for data-driven material models. In: Vision, Modeling & Visualization. The Eurographics Association (2017)
22. Nicodemus, F., Richmond, J., Hsia, J., Ginsburg, I., Limperis, T.: Geometrical considerations and nomenclature for reflectance. NBS Monograph 160 pp. 1–52 (1977)
23. Ramanarayanan, G., Ferwerda, J., Walter, B., Bala, K.: Visual equivalence: towards a new standard for image fidelity. ACM Transactions on Graphics **26**(3), 76:1–76:10 (2007)
24. Ravishankar Rao, A., Lohse, G.: Towards a texture naming system: Identifying relevant dimensions of texture. Vision Research **36**(11), 1649–1669 (1996)
25. Rushmeier, H.: The MAM2014 sample set. In: Proceedings of the Eurographics 2014 Workshop on Material Appearance Modeling: Issues and Acquisition, MAM '14, pp. 25–26 (2014)
26. Sattler, M., Sarlette, R., Klein, R.: Efficient and realistic visualization of cloth. In: Eurographics Symposium on Rendering, pp. 167–178 (2003)
27. Serrano, A., Gutierrez, D., Myszkowski, K., Seidel, H.P., Masia, B.: An intuitive control space for material appearance. ACM Trans. Graph. **35**(6), 186:1–186:12 (2016)
28. Somol, P., Haindl, M.: Novel path search algorithm for image stitching and advanced texture tiling. In: WSCG 2005, pp. 155–218 (2005)
29. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. Systems, Man and Cybernetics, IEEE Transactions on **8**(6), 460–473 (1978)
30. Vangorp, P., Laurijssen, J., Dutre, P.: The influence of shape on the perception of material reflectance. ACM Transactions on Graphics **26**(3), 77:1–77:10 (2007)

Appendix – BIG Data Format Description

The format stores image data by providing a list of image files to be included (so far PNG and EXR (half/float) image files are supported) together with optional meta data such as list of corresponding incoming and outgoing directions, color-space, spatial resolution, measured material name and descriptions. The stored binary data can be either loaded to the RAM or alternatively, for large datasets one can open a datafile and seek the requested data from a hard drive. The latter option is considerably slower but still acceptable for many offline rendering scenarios. Once the file is loaded/opened one can use a standard "get-pixel" query function returning RGB triplet for specific spatial UV coordinate and image index. A transformation between image index and incoming/outgoing angles is up to the user and depends on an initial ordering of files during the saving process. Also we do not attempt to provide any compression of data as this could potentially impact visual quality and rendering speed. The compression can be easily added by extension of the format.

Since the proposed format is universal (it can include any LDR/HDR data), it allows an unified representation of any image-based information, e.g., movies, dynamic textures. The format also enables management of numerous scattered files that are difficult to handle without any metadata. The source codes for saving/loading of data to/from the format are made publicly available (http://btf.utia.cas.cz) to promote its wide usage and allowing easy adoption by various users for visualization and data analysis software packages. The format is composed of data chunks consisting of chunk ID, its size and data, as shown in a list of current data chunks and their brief description in Fig. 15.

ID	***parameter***	***datatype***	***description***
1	Image Width	longlong	>0
2	Image Height	longlong	>0
3	Number of Channels	longlong	>0
4	Number of Images	longlong	>0
5	Data Type	longlong	-1=mix, 2=half, 4=float
6	Data		
7	Index Map	char	if Data Type=-1, specifies Data Type of each image
8	Colour Space ID	longlong	-1=unknown, 1=sRGB, 2=RGB, 3=XYZ
9	Pixels per mm	float	
10	DPI	float	
14	List of Directions - instrument coordinates		(4 x float x Numer Of Images)
15	List of Directions - spherical coordinates		(4 x float x Numer Of Images)
16	Llist of Directions - idealized spherical directions		(4 x float x Numer Of Images)
17	Text (File)	char*	
18	Material Name	char*	
19	XML File	char*	
20	Number of tiles	longlong	>0

legend
required in input
detected automatically
optional in input
recommended in input

Fig. 15 A list of data chunks available in the BIG format.

Compliance with Ethical Standards

Funding: This research has been supported by the Czech Science Foundation grant 17-02652S and the US National Science Foundation grant IIS-1218515.

Conflict of Interest: The authors declare that they have no conflict of interest.