# Perceptual Attributes Analysis of Real-world Materials

JIŘÍ FILIP and MARTINA KOLAFOVÁ, The Czech Academy of Sciences, Institute of Information Theory and Automation, Czech Republic

Material appearance is often represented by a bidirectional reflectance distribution function (BRDF). Although the concept of the BRDF is widely used in computer graphics and related applications, the number of actual captured BRDFs is limited due to a time and resources demanding measurement process. Several BRDF databases have already been provided publicly, yet subjective properties of underlying captured material samples, apart from single photographs, remain unavailable for users. In this article, we analyzed material samples, used in the creation of the UTIA BRDF database, in a psychophysical study with nine subjects and assessed its 12 visual, tactile, and subjective attributes. Further, we evaluated the relationship between the attributes and six material categories. We consider the presented perceptual analysis as valuable and complementary information to the database; that could aid users to select appropriate materials for their applications.

CCS Concepts: • **Computing methodologies → Perception**; **Reflectance modeling**;

Additional Key Words and Phrases: BRDF, attributes, perception, visual, tactile, user study

**ACM Reference format:**
Jiří Filip and Martina Kolafová. 2019. Perceptual Attributes Analysis of Real-world Materials. *ACM Trans. Appl. Percept.* 16, 1, Article 1 (January 2019), 19 pages.
https://doi.org/10.1145/3301412

## 1 INTRODUCTION

Textureless reflectance is a basic means of material appearance representation used in computer graphics and related disciplines. A formal representation using a bidirectional reflectance function (BRDF) [25] represents appearance as a distribution function of ratios of outgoing radiance to incoming reflectance for any pair of incoming and outgoing directions. An anisotropic BRDF for a fixed color channel is a four-dimensional function $f_r(\theta_i, \varphi_i, \theta_v, \varphi_v)$ of illumination and viewing directions specified by spherical angles, and a spatial extension of a BRDF is a bidirectional texture function (BTF) $f_r(x, y, \theta_i, \varphi_i, \theta_v, \varphi_v)$. As the sampling process of bidirectional pairs requires the independent positioning of light and camera over a hemisphere defined by a normal of a material's surface, the acquisition process can be very slow. Although the measurement process can be sped up using mirrors and taking images of multiple viewing directions simultaneously, or by making assumptions of a known material object's geometry [21], the number of available BRDF datasets remains limited.

This research has been supported by the Czech Science Foundation grant GA17-18407S.
Authors' addresses: J. Filip and M. Kolafová, The Czech Academy of Sciences, Institute of Information Theory and Automation (ÚTIA), Pod Vodárenskou věží 4, 182 08, Praha 8, Czech Republic; emails: {filipj, kolafova}@utia.cas.cz.

1544-3558/2019/01-ART1 $15.00
https://doi.org/10.1145/3301412

The first public BRDF database of isotropic measurements was introduced in Reference [21] and additional BRDFs in References [8, 19]. Regarding anisotropic measurements, four detailed datasets are available [24], three probes of fabric materials [4], and an additional 150 in the UTIA BRDF database [3].

These important resources provide comprehensive information about the angularly dependent visual behavior of measured material samples captured as a BRDF; however, little is known about the subjective visual and tactile impression of human subjects apart from a single snapshot of the measured surface. Additional knowledge of such perceptual attributes would benefit users of such data in a variety of applications. For instance, a user from the gaming industry can select a BRDF that has the highest glossiness or leather of the highest perceived quality, and so on. Also, designers could select, for example, a BRDF of wood representing the highest perceived roughness in the visualization of a particular product. Using a psychophysical study, this article analyzes subjects' impressions of actual materials captured in the UTIA BRDF database[1] [3]. We further studied the correlations between perceived attributes and between the attributes and material classes and their properties.

The article is organized as follows. Section 2 overviews a prior work on the topic. Section 3 describes the user study performed, while Section 4 analyses and discuses subjects' responses and relationship of attributes and material categories. Section 5 discusses the results limitations and applications. Section 6 concludes the article.

## 2 RELATED WORK

Our work relates to the analysis of perceptual features of (1) gray-scale/color textures, (2) complex spatially and angularly dependent properties of material appearance, and (3) real specimens possibly using additional multi-modal features.

**Perception of Materials Properties from Photographs**

Analysis of human visual perception of texture has been studied extensively in the past [16], and consequently this research evolved in two major directions. The first represents statistical approaches that tend to represent a texture's visual properties by a set of statistical measures. The second group of approaches is based on the identification of major visual features, often called textons, and represents texture by combinations from a texton vocabulary.

Tamura et al. [34] suggested a computational form of six basic texture properties and evaluated their performance in a psychophysical experiment on 56 gray-scaled textures of Brodatz's catalogue [1]. The features were as follows. *Coarseness*—the bigger is structure element size and/or less its elements are repeated, the coarser it is felt to be. *Contrast*—is based on four factors: (1) dynamic range of gray-levels, (2) polarization of the distribution of black and white on the gray-level histogram or ratio of black and white areas, (3) sharpness of edges, (4) period of repeating patterns. *Directionality*—intensity of monodirectional/bidirectional features. *Line-likeness*—whether has prevailing lines or blobs. *Regularity*—is related to variations of placement rule, where fine textures are often perceived as regular. *Roughness*—expressing imaginary tactile sense.

Rao and Lohse [29] identified a perceptual texture space also by grouping Brodatz's textures. The resulting data were analyzed using hierarchical cluster analysis, non-parametric multi-dimensional scaling (MDS), classification and regression tree analysis, discriminant analysis, and principal component analysis. The authors concluded that the perceptual texture space can be represented by a three-dimensional space with axes describing repetitiveness, contrast/directionality, and coarseness/complexity. The tested attributes were: *granular textures*, *marble-like*, *lace-like*, *random*, *repetitive*, and *directional*. Another image grouping experiment by Heaps and Handel [10] obtained the following basic texture attributes: *complexity*, *connectedness*, *depth*, *hardness*, *linearity*, *naturalness*, *orientation*, *repetitiveness*, *roughness*, *shape*, *shape*, *size*, and *structure*.

Reference [22] performed an experiment with human subjects to obtain a pattern vocabulary governed by grammar rules accounting for: *overall color*, *directionality and orientation*, *regularity and placement*, *color*

[1]http://btf.utia.cas.cz.

*purity*, *complexity,* and *heaviness.* Although this approach utilized a different perceptual mechanism as it extended the scope to color textures, it generally proved Rao and Lohse's conclusions [29] concerning the perceptual dimensions.

Malik and Perona [18] presented a model of human preattentive texture perception based on low-level human perception. Vanrell and Vitria [39] suggested a texton-based four-dimensional texture space with perceptual textons' attributes along each of the dimensions. In follow-up work, they [40] performed dimensionality reduction of the texton representations to create a low- dimensional behavioral texture space where distances between points represent texture distances.

Long and Leow [17] presented an approach attempting to solve the missing link between the perceptual texture space and the space of computational texture features, by reduction of Gabor features represented by a convolutional neural network a four-dimensional texture space.

Researchers studied also human perception of specific attributes. Padilla et al. [26] developed a model of perceived roughness in fractal surfaces. Pont et al. [27] found that observers use the texture as a cue for relief-depth and that surface roughness can be exploited for increasing realism of the standard 2D texture mapping. Ho et al. [11] found that roughness perception is correlated with texture contrast. Motoyoshi et al. [23] suggest that skewness of the texture's luminance histogram and of sub-band filter output are correlated with the original surface gloss and inversely correlated with the surface albedo. Wills et al. [41] ran an extensive experiment with 75 subjects assessing glossiness of 55 BRDFs from MERL BRDF database [21]. They further used the MDS to construct a low-dimensional embedding allowing perceptual parameterization of an analytical BRDF model.

Fleming et al. [7] presented an extensive analysis of human perception of materials. In the first study nine subject judged nine perceptual qualities: *glossiness*, *transparency*, *colorfulness*, *roughness*, *hardness*, *coldness*, *fragility*, *naturalness*, and *prettiness* of 13 exemplars of 10 material classes. The stimuli images featuring materials on various shapes were taken from MIT-Flickr database. In the second study, 65 subjects assigned 42 adjectives describing material qualities to six classes of materials. Authors revealed that distribution of material classes in visual and semantic domain are similar and conclude that perceptual qualities are systematically related to material class membership.

In a follow-up study, Tanaka et al. [35] analyzed subject rating of the same perceptual qualities as a function of visual information degradation. They assessed the qualities of 34 low-chroma specimens divided into 10 material categories. Ten subjects judged real examples, their images in the same environment, and their gray-scale and down-sampled versions. Authors concluded that general perceptual quality decreased by image-based reproduction, also perceptual qualities of images decreased by using their gray-scale variant, and finally, perceptual qualities of *hardness* and *coldness* increased when image resolution was reduced.

A related research in material recognition identifies local material properties, so-called visual material traits [30], encoding appearance of characteristic material properties by means of convolutional features of train patches. In follow-up work, researchers discovered space of locally recognizable material attributes from perceptual material distances by training classifiers to reproduce this space from image patches [31, 33].

**Perception of Angularly Dependent Material Appearance**

An important aspect of material perception is its variation due to interactions between lighting and object geometry and object geometry [5]. Fleming et al. [6] pointed out the importance of real-world illumination for correct matching of surface roughness and specularity. Limited work has been carried out on the perceptual aspects of real BRDF measurements. te Pas and Pont [36] showed that changes in surface BRDF and illumination are often confounded, but that adding complex illumination or 3D texture improves the visual matching. Dependency of the perception of a light source direction on surface BRDF and its 3D shape was shown in References [15, 37]. Matusik et al. [21] psychophysically evaluated large sets of BRDF samples and showed that there are consistent transitions in perceived properties between different BRDFs. They analyzed whether they possess any of the perceptual attributes: *redness, greenness, blueness, specularness, diffuseness, glossiness,metallic-like, plastic-like,* *roughness, silverness, gold-like, fabric-like, acrylic-like, greasiness, dustiness,* and *rubber-like*. They used the subject's characterizations to build a model in both the linear and non-linear embedding spaces. Such a manifold is then used for interpolation between attributes.

Vangorp et al. [38] found that object shape considerably influences the perception of BRDF samples. Filip et al. [2] ran a psychophysical study and demonstrated that a reduction of BTF data without compromising visual quality is directly related to the complexity of illumination and object geometry.

Ramanarayanan et al. [28] presented a concept of visual equivalence. In a series of psychophysical experiments, the authors characterized conditions under which warping and blurring of the illumination maps and warping of the object's surface yield rendered images that are visually equivalent to the reference solutions. Jarabo et al. [13] ran perceptual experiments to investigate the visual equivalence [28] of rendered images for different levels of BTF filtering, and found that blur in a spatial domain is less tolerable than in its angular counterpart. They analyzed whether BTF datasets possess following properties: *high-contrast*, *granular*, *structured*, *rough*, *feature-dense*, *complex-structure*, *flat*, *relief*, *sharp-relief*, *smooth-relief*, *glossy*, *color*, *light*, *soft*, and *hard*.

Finally, Serrano et al. [32] psychophysically analyzed isotropic BRDFs from the MERL database [21] to identify smooth and intuitive material appearance transition between different visual attributes: *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*.

**Perception of Physical Material Specimens**

Psychophysical analysis of physical materials received relatively low attention. Reference [20] analyzed the visual, aural, and tactile attributes of materials and their mutual contribution to perception of characteristic material parameters. Authors used tactile features: *rough–smooth, hard–oft, warm–cold*; visual features: *shiny–matte, simple–complex, colorful–colorless*; and subjective features: *expensive–cheap, old–new, natural–synthetic, beautiful–ugly*. Tanaka et al. [35] ran a user study assessing nine visual attributes of 34 planar material specimens observed in strictly controlled settings. The study further analyzed impact of degraded image representations of real specimens on judging of attributes intensity.

Inspired by the multi-modal analysis of real material specimens, we provided subjects with real material samples used in the creation of the UTIA BRDF database [3], and evaluated not only their visual but also tactile and subjective properties. To our knowledge, this is one of the first attempts to thoroughly analyze perceptual properties of a large BRDF database, not merely based on the visualization of captured or even fitted BRDF data, but using real material specimens.

## 3 PSYCHOPHYSICAL EXPERIMENT

This section overviews the analyzed materials, tested attributes, and experimental procedure.

### 3.1 Tested Materials and Their Categories

The UTIA BRDF database [3] is composed of 150 anisotropic BRDFs (see Figure 1).

To reduce time demands of the study, out of this database, we selected 93 material specimens by removing those that had the same structure yet differ only in a color. The resulting subset consists of 51 fabric, 15 leather, 4 plastic, 16 wood, and 7 other specimens comprising: retro-reflective material, metallic car paint with acrylic coating, glittering glass paint, plaster, black paper, and two specimens of polyester foam as shown in Figure 2.

### 3.2 Material Attributes

Our motivation was to analyze as many aspects of the tested materials as possible. However, as the number of assessed materials was relatively high, we aimed at comparatively low number of attributes to make the study feasible for human subjects. We intentionally avoided using high-level class predictors, e.g., plastic-like, as used in References [21, 32], and focused on important low-level attributes identified in the prior work.

Fig. 1. The images of 150 original materials used for capturing UTIA BRDF database.

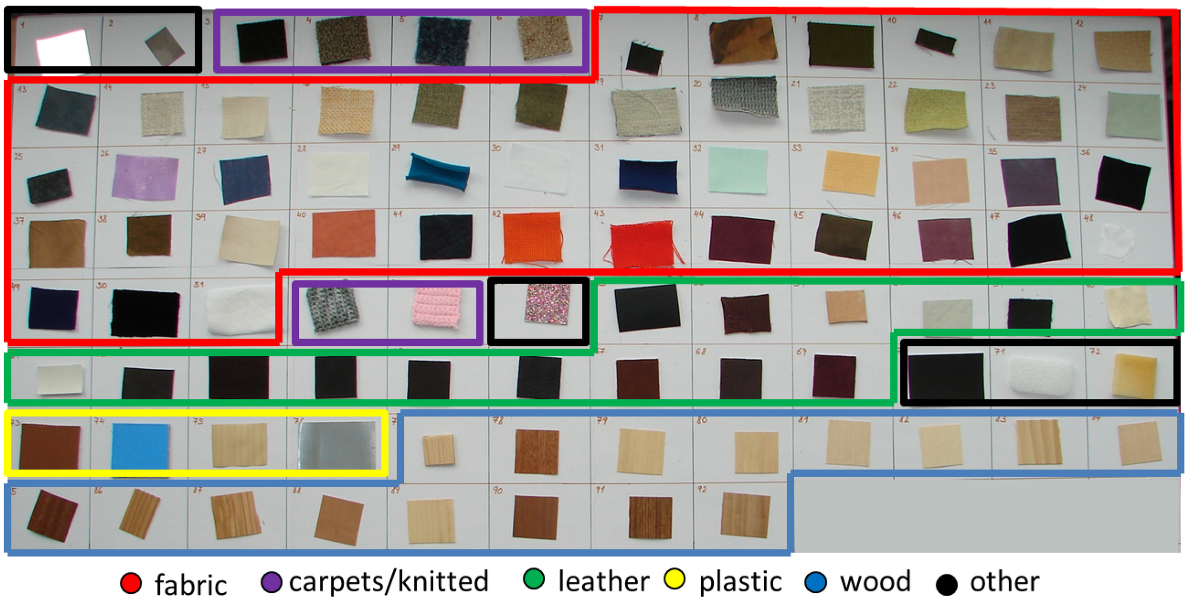



Fig. 2. A tabletop with 93 material specimens used in the experiment. Individual material categories are highlighted by color.

Our goal was a comprehensive multi-modal analysis of the materials. From the visual characteristics, we focused on the general texture attributes [22, 29, 34] 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 specimen, we also extended our attributes to the tactile and subjective domain, similar to Reference [20]. In the tactile domain, we focused on attributes having a 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*.

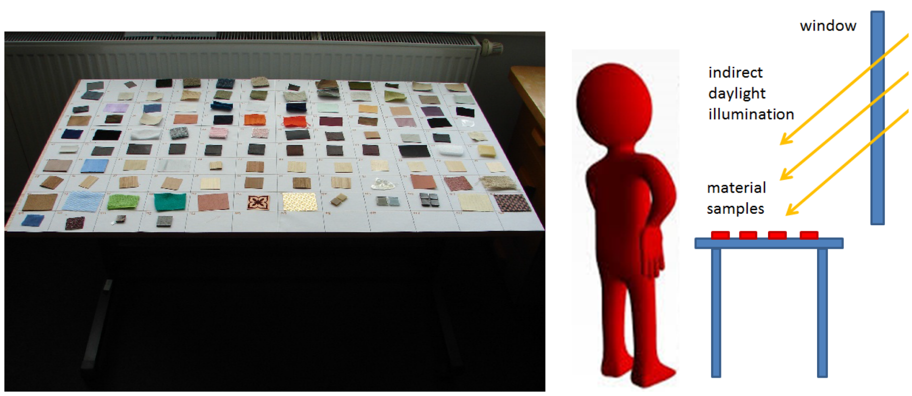
Thus our final list of attributes includes six visual, three tactile, and three subjective attributes as shown in Table 1.

### 3.3 Participants and Experimental Procedure

Nine paid subjects performed the experiment in one session. Their ages ranged from 20 to 59, four were male and five were female. All subjects had normal or corrected-to-normal vision, and all were uninformed with respect to the purpose and design of the experiment.

Table 1. The 12 Material Attributes Evaluated within the User Study

| 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 |




(a) (b)

Fig. 3. A table with the samples (a) and design of the experiment setup (b).

All samples of a typical size 5 × 5 cm were laid down on the table as shown in Figure 2. The table was positioned near a window avoiding direct sunlight as shown in Figure 3(b).

The subject's task was to assess intensity of the attributes shown in Table 1. Prior to the testing session, there was a short training session where individual attributes were explained to the subject and she/he could freely review all samples. At this stage, the subject should have established her/his personal scales for each attribute, as they could see all material at a glance, especially the cases representing extreme values of the attributes. Due to this step, we did not further normalize the subjects' scales in any way. Note that each subject had an opportunity to glance around all the material throughout the experiment. In the testing session, the subject sequentially evaluated individual attributes while observing one material at a time. All materials were presented to subjects at once (see Figure 2), and they could evaluate them in any order. In all cases, subjects evaluated the materials as they were numbered on the table and finished each attribute for all materials, before moving to the next one.

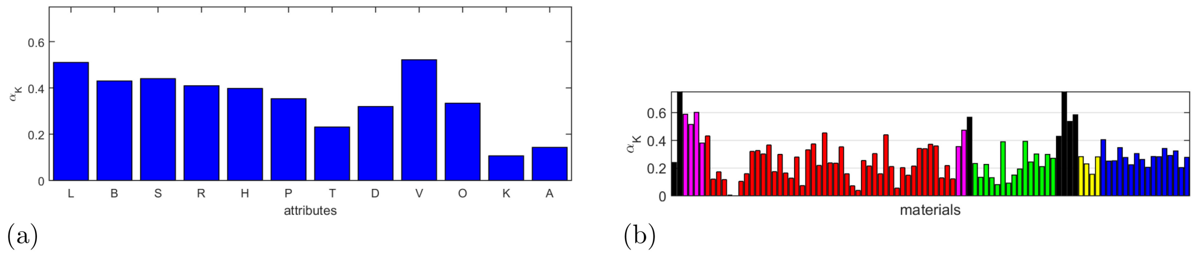



Fig. 4. Subjects agreement assessment using Krippensdorff alpha: (a) for individual attributes, (b) for individual materials.

To provide sufficient options for the subjects, we used a nine-point Likert-like rating scale, where 1 corresponds to the lowest and 9 to the highest intensity of the queried attribute. This range should represent only the span of materials within the study, i.e., the highest rating of, e.g., glossiness should correspond to the glossiest material from the study and not from the real world. We adapted this design as it is dominant in image and video experiments [14].

To avoid confusion between visual and tactile attributes, subjects were not allowed to touch the sample when assessing visual attributes. Also, subjects were not allowed to turn samples over at any time as this could help them to identify the authenticity of materials. The task was clear for all subjects, and as there were no strict time limits, the subjects finished the experiment in between 55 and 160min. During debriefing, the majority of subjects found the experiment interesting yet demanding.

## 4 RESULTS AND DISCUSSION

We divided our analysis to raw data aggregated from the observers, statistical analysis of attributes, in general, and statistical analysis of attributes across material categories.

### 4.1 Raw Data Analysis

Prior to data analysis, we checked on the presence of outliers and assessed agreement across subjects. First, we performed outliers rejection by removal of values differing from mean subject response for more than 5 points. A total of 249 outliers were found representing 2.5% of 10,044 values recorded in the study. By far, most outliers (99) were recoded for the attribute *P–regularity*. Next, we checked subjects' responses agreement using Krippensdorff alpha [9]—a statistical measure of the agreement generalizing several known statistics. The key requirement is agreement observed among independent observers. Output $\alpha_K = 1$ represents unambiguous indicator of reliability, while 0 does not. The $\alpha_K$ values computed for individual attributes are shown in Figure 4(a). The values mean was $\alpha_K = 0.37$, the intuitive attributes (*L–glossiness*, *V–height*) featured the highest agreement (over 0.5), while the lowest agreement values were obtained for subjective features *K–quality*, *A–attractiveness*, and, surprisingly, also for *T–hardness*. The $\alpha_K$ values for all 93 materials are shown in Figure 4(b).

We have also analyzed normality of the data to select appropriate method for statistical hypothesis testing. The Shapiro-Wilk normality test revealed that the data do not comply with normality assumption at significance level 0.05. As the data normality is one of the basic assumptions of ANOVA analysis, we used for any further hypotheses testing non-parametric Kruskal-Wallis tests instead. Hypotheses testing of attributes means using Kruskal-Wallis, repeated measures ANOVA, and Friedman tests confirmed significant differences between attributes means with p-values below 1e-149.

We additionally tested significance of data of each material across different subjects using repeated measures ANOVA and Friedman's tests at significance levels 0.01 and 0.05. We concluded that for repeated measures ANOVA 8 and 7 materials, respectively, and for Friedman test only 2 and 3 materials, respectively,

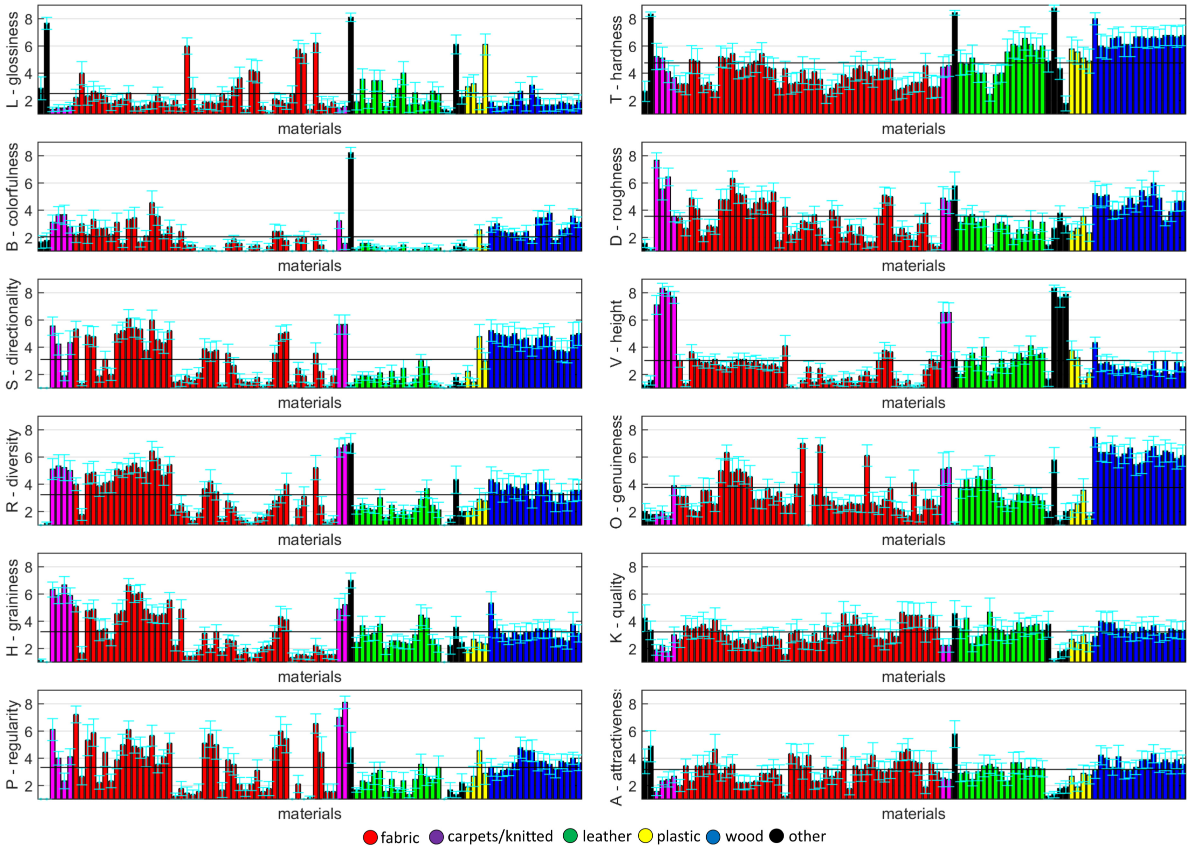



Fig. 5. The mean opinion scores for individual perceptual attributes and all materials. Colors represent individual material categories as shown in the legend (bottom). Errorbars represent standard deviation across subjects and the black horizontal line is a mean value of the attribute across materials.

had p-values above these significance levels. These tests confirmed significant differences between means of attributes values across all observers for majority of materials.

To get insight to typical subjects' responses, we computed the mean opinion score (MOS) obtained as average rating across all subjects. This is standard methodology for subjective quality assessment used especially in the audio and video industries, and recommended by standard international organizations such as ITU [12] or ISO [14]. The mean opinion scores in range 1-9 for individual perceptual attributes and tested materials are shown in Figure 5. The errorbars in the graphs represent the standard error values. Average values of attributes are shown as a black outline in Figure 5. All statistical analyses further on were carried out on all user ratings and not using MOS.

For the sake of our analysis, materials are grouped into six categories: fabric, carpet, leather, plastic, wood, and other as shown in Figure 2, which are color-coded in the same way also in Figure 5. We will discuss each of the attributes briefly while their values are located in the supplementary material.

**Glossiness**—a limited number of glossy materials in the database was reflected by typically low perceptual scores. Exceptions were *car_paint01*, some highly anisotropic fabric materials, *glitter*, and plastics.

**Colorfulness**—subjects generally did not find materials colorful with the exception of *glitter*. Leather materials were viewed as the least colorful, while the most colorful considered were wood, upholstery fabric and carpet

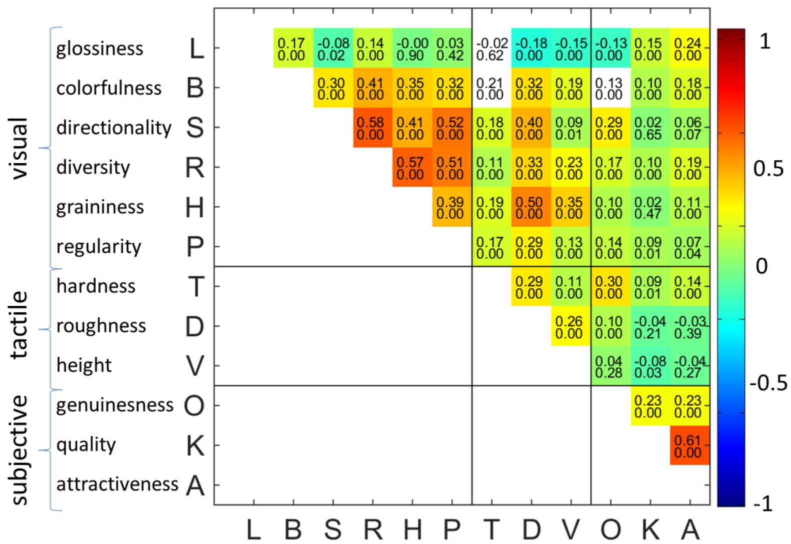



Fig. 6. A Pearson correlation matrix between the tested attributes. Colors toward red represent positive correlation while those toward blue represent negative ones. The white values report significant difference between the attributes.

(the first half of columns in red). Presumably, this was due to texture contrast. This attribute had a low standard deviation across subjects and materials.

**Directionality**—similar to the previous attribute, the category of materials with the least directional structure was leather, while the most were categories: wood, upholstery fabric, and carpet.

**Diversity**—was also related to the richness of visual features and texture contrast. However, contrary to directionality, diversity was less prominent in wood while more apparent in structured fabrics.

**Graininess**—results were, as expected, highly similar to the diversity attribute.

**Regularity**—subjects were quite inconsistent when evaluating this attribute as the standard deviation was the highest, while their responses were low and balanced across groups of materials. An exception was knitted fabric due to its large and regular pattern.

**Hardness**—the first of the tactile properties received relatively high responses, especially in the wood category. The highest response was obtained for the *plaster* material, which was treated as stone.

**Roughness**—this tactile property was highly related to the visual attributes of diversity and graininess with the highest responses being for wood and structured fabric categories.

**Height**—was one of the attributes that can be assessed objectively and thus had the lowest standard deviation across subjects. The samples with the highest height were carpet, crocheted fabric, plaster, and polyester foam.

**Genuineness**—this subjective attribute was assigned especially to wood, plaster, and some fabric materials.

**Quality and Attractiveness**—were subjective perceptual attributes that were highly correlated and for all of the tested materials obtained relatively low scores. An exception was glitter material that was considered attractive. In general, all three subjective attributes had relatively high standard deviations.

## 4.2 Attributes Relationship Analysis

To obtain correlation values between attributes, we computed Pearson and Spearman correlation coefficients between individual pairs of attributes across all subjects ratings of 93 tested materials. The results were very similar, and thus we report only the Pearson correlation values as shown color-coded together with the actual correlation value in Figure 6. Colors toward red represent positive correlation while those toward blue represent

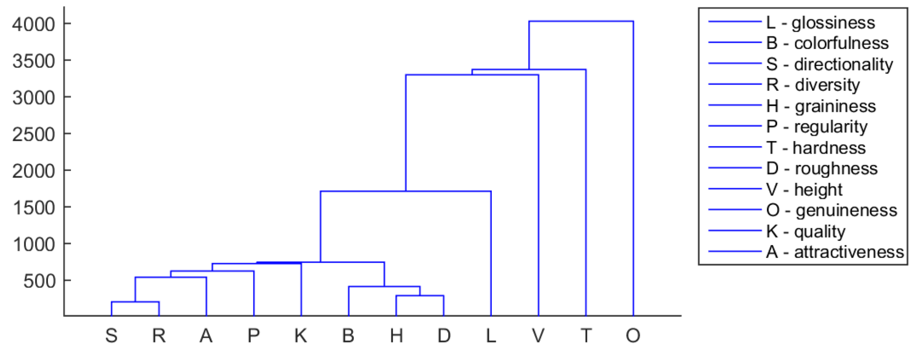



Fig. 7. Multivariate analysis – dendrogram illustrating proximity of attributes.

negative ones. Each pair of attributes in the figure show correlation coefficient $(-1, 1)$ at the top of the cell and corresponding p-values at its bottom. To identify the attributes combinations having high correlations but not being statistically significant, we performed a one-way Kruskal-Wallis test followed by a multiple comparison procedure of attributes' means. This is designed to provide an upper bound on the probability that any comparison will be incorrectly found significant. The differences that are significant at the 0.05 level, i.e., 95% confidence interval, are colored in Figure 6 in white.

One can find higher positive correlations especially between *directionality–diversity* (0.58), *directionality–regularity* (0.52), *diversity–graininess* (0.57), and *diversity–regularity* (0.51). There are also higher positive correlations between tactile *roughness* and *graininess* (0.50). As already mentioned, the attributes of *quality* and *attractiveness* have also similar perceptual responses with a correlation 0.61.

To further investigate attributes relationship, we performed a one-way Multivariate Analysis of Variance (MANOVA) for comparing the multivariate means of the subjects responses grouped by attributes. This analysis tests the null hypothesis that the means of each attribute are the same n-dimensional multivariate vector, and that any difference observed in the data is due to random chance. As the estimated dimensionality is $d = 9$, we can reject the null hypothesis at significance level 0.05, and we expect the data means lie in a nine-dimensional manifold. As a product of the multivariate analysis, a dendrogram depicting attributes proximity is also obtained, as shown in Figure 7. The clusters are computed by applying the single linkage method to the matrix of Mahalanobis distances between group means. Here, we can observe large differences in attributes *O–originality*, *T–hardness*, *V–height*, and *L–glossiness* as opposed to other attributes especially *H–graininess* to *D—roughness*, *S–directionality* to *R—diversity*. These results together with closeness of *K–quality* to *A–attractiveness* support the findings of correlation analysis in Figure 6. The dimensionality $d = 9$ and the dendrogram reveals that one can group attributes (S, R) and (B, H, D).

Finally, multivariate analysis provides us with so-called canonical vectors, which are linear combinations of the original variables chosen to maximize the separation between groups. Figure 8 depicts dependencies of the first three canonical vectors. Here, one can see the clear separation of individual color-coded attributes, as well as the coherence of subjects' responses demonstrated by the narrow distribution of points for each attribute.

## 4.3 Material Categories Relationship Analysis

For the sake of better understanding of perceived differences between individual materials, we focused our further analysis on the six material categories: fabric, carpet, leather, plastic, wood, and other.

Basic statistical analysis of median data statistic across materials in individual groups for averaged subjects' responses is given in Figure 9. The red line represents median value (50th percentile) across all materials in the group, blue bars represent 25th and 75th percentiles, data variance is depicted as a black bar, and outliers are

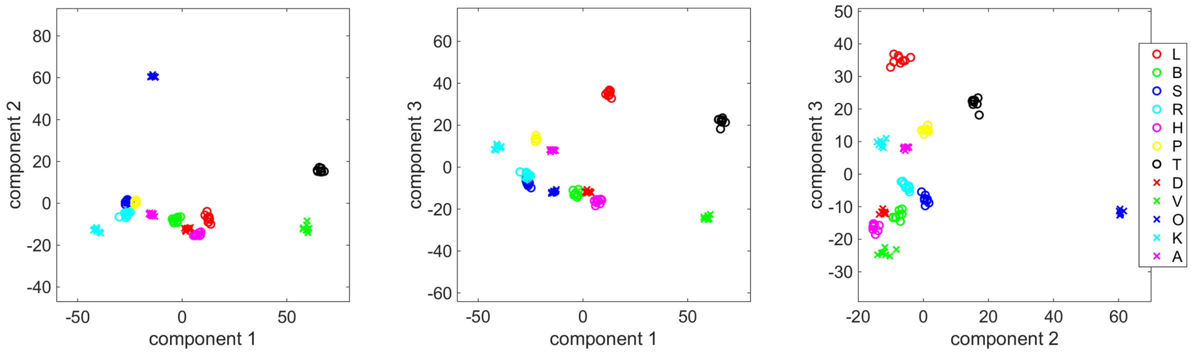



Fig. 8. Multivariate analysis of attributes showing separation of the first three canonical components.

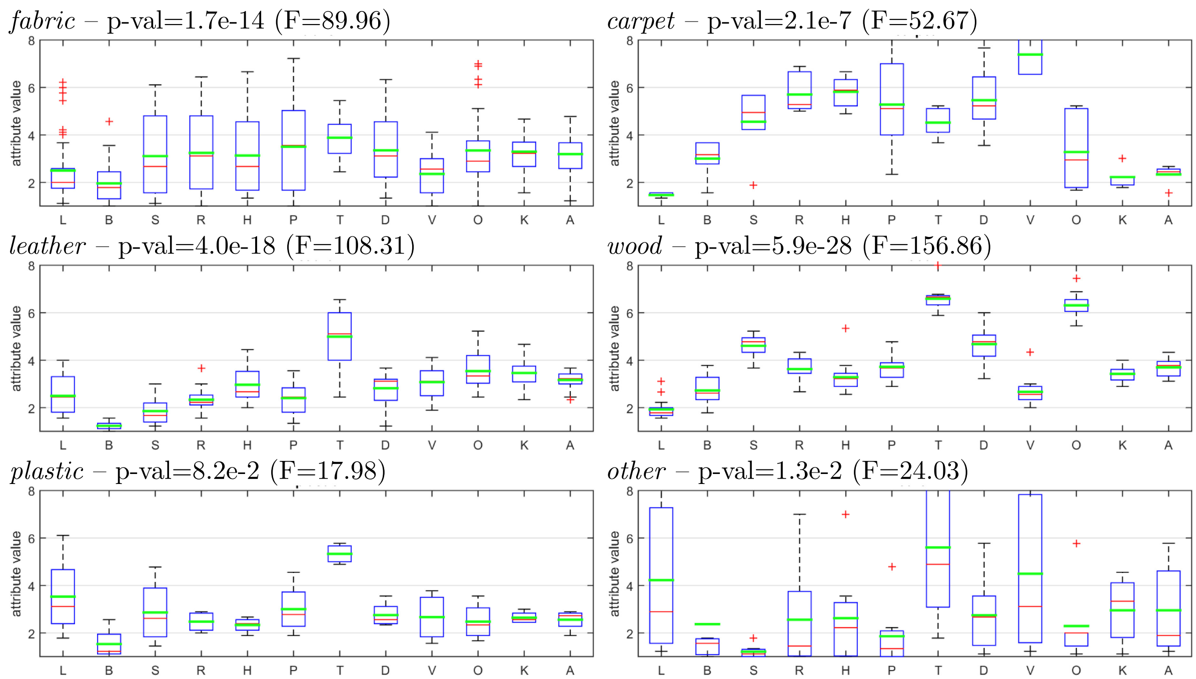



Fig. 9. Attributes median statistics across materials compared to a mean opinion score (green) for individual material categories. The red line represents median value (50th percentile) across all materials in the group and blue bars represent 25th and 75th percentiles.

depicted as red crosses. The figure also comprises corresponding MOS scores as green lines. From the comparison of the median (red) and mean values (green) and from errorbars, we can analyze how well MOS represent the data. Included are also results of Kruscall-Wallis testing of attributes means underlining a high statistical significance with p-values below 0.05 (with exception of group *plastic*).

The attributes values obtained for individual material groups are quite intuitive. Groups *plastic* and *other* have higher glossiness due to the inclusion of car-paint and glitter samples. Also, *carpet* has higher *height* and *diversity, graininess, roughness.* Group *wood* has the highest ranking in *hardness* and *genuineness.* Also, it has together with the group *carpet* the highest directionality. High standard errors for category *fabric* are due to a high variability

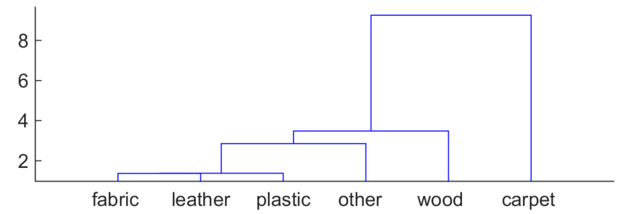



Fig. 10. Multivariate analysis—a dendrogram illustrating the proximity of material categories.

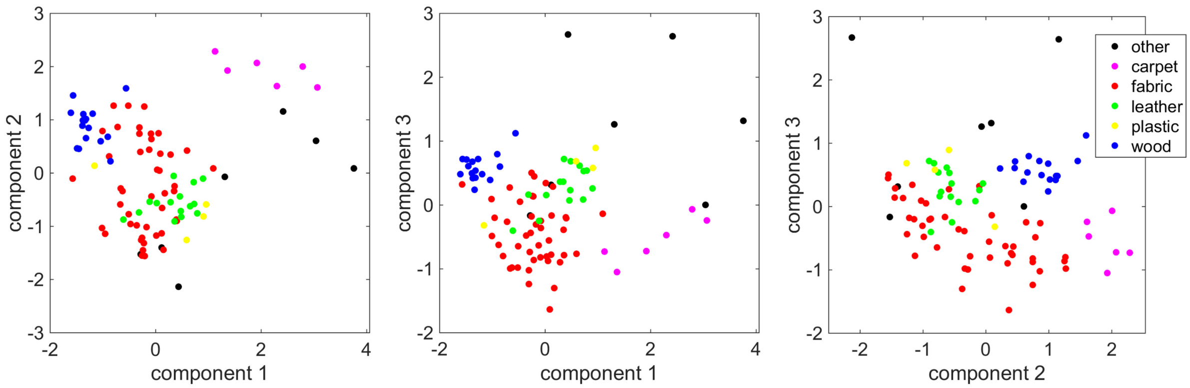



Fig. 11. Multivariate analysis of material groups showing separation of the first three canonical components.

of materials caused by combination of thread materials and weaving patterns (see Figure 2). Also, we noticed considerably higher standard errors for groups *plastic* and *other*, which is due to diversity and the low number of materials in these groups (four and five, respectively). For this reason, we restrict part of our further analysis only to categories *fabric*, *leather*, and *wood*.

Similar to attributes analysis, we performed the Multivariate Analysis of Variance (MANOVA) to compare the multivariate means of the subjects responses grouped by attributes. The estimated dimensionality $d = 4$ suggests that the six material categories can be in our case reduced to a four-dimensional manifold. The dendrogram plot shown in Figure 10 analyses the proximity of individual groups. The analysis sets class *carpet* apart from all the others, while the closest ones are classes exhibiting relatively smooth structure *fabric*, *leather*, and *plastic*.

When we look at the first three canonical vectors resulting from the multivariate analysis (shown in Figure 11), we can see relatively good separation, especially of the material categories *fabric* and *wood* shown in red and blue, respectively. However, *fabric*, *leather*, and *plastic* were the most mutually confounded categories.

Next, similar to Figure 6, we analyzed relationships between attributes using Pearson correlations for three groups of materials: *fabric*, *leather*, and *wood* as shown in Figure 12. Our analysis revealed that *fabric* has a higher positive correlation between visual attributes, and visual–tactile attributes. This suggests that our perception and representation of properties of fabric is highly related to tactile properties, which is different from *leather* and *wood* classes. Also *fabric* class has neutral or slightly negative correlations between material structure related attributes and subjective attributes *K–quality* and *A–attractiveness*. The attributes whose correlations were significantly consistent across all three categories are pairs *diversity–graininess*, and *quality–attractiveness*.

We also evaluated the consistency of attributes across different material categories. To this end, for each attribute, we tested a null hypothesis that its mean values are drawn from the same population regardless the material class. We used an unbalanced Kruskal-Wallis test as the number of materials in individual groups differ. Table 2 shows that all p-values are below 0.002, thus favoring an alternative hypothesis that means for

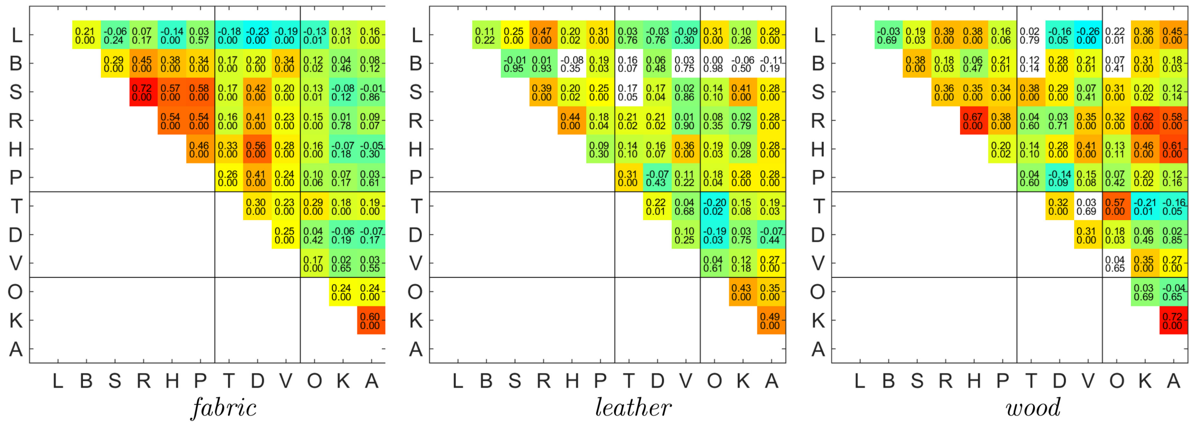



Fig. 12. A Pearson correlation matrix between the tested attributes for material categories *fabric*, *leather*, and *wood*. Colors toward red represent positive correlation, while those toward blue represent negative ones. The white values report significant difference between the attributes.

Table 2. Results from Kruskal-Wallis Hypothesis Testing Using Analysis of Variance for Individual Attributes Across Different Material Groups

| | Visual | | | | | | Tactile | | | Subjective | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Attribute | L | B | S | R | H | P | T | D | V | O | K | A |
| p-value | 2.3e-3 | 4.7e-6 | 7.0e-7 | 4.4e-4 | 2.3e-3 | 1.9e-3 | 1.2e-8 | 1.7e-5 | 3.1e-4 | 6.8e-8 | 2.8e-3 | 1.6e-3 |
| F | 18.12 | 32.52 | 36.67 | 22.42 | 18.57 | 19.00 | 45.43 | 29.72 | 23.16 | 41.70 | 18.09 | 19.41 |

individual materials are draw from statistically different populations at significance level 95%. The lowest significance across material categories was obtained for attributes: *glossiness*, *graininess*, *regularity*, *quality*, and *attractiveness*, while the highest significance was obtained for *hardness*, *genuineness*, and *directionality*. Table 2 comprises also values of F-statistic, which is a ratio of two quantities (variation between sample means/variation within the samples) that are expected to be roughly equal under the null hypothesis, while increases with significance of the alternative hypothesis. Even though the condition of data normality was not fulfilled, we tested our data also using a single-factor repeated measures ANOVA, and in all cases we obtained significantly different values of attributes means (obtained p-values lower than 0.001 for significance level 0.05).

Finally, to relate our work to recent thorough studies of human perception of materials [7, 35], we performed a principal component analysis (PCA) of all ratings across subjects. The importance of individual factors shown in Figure 13 demonstrates high contribution of the first four principal components (PC). In fact, the first three components amount to 56% of data variability and the first four of them to 65%. The factors loadings for the first two components are shown in Table 3. While the PC1 is strongly related to structure attributes (*directionality*, *diversity*, *graininess*, *regularity*), the PC2 is mainly related to subjective attributes *genuineness*, *quality*, and *attractiveness*. PC3 positively represents *hardness* and *genuineness* and negatively represents *glossiness*.

We can compare to the previous studies only with the material categories *fabric*, *leather*, and *wood*, while the rest of categories are either missing or are underrepresented in the UTIA database. Both previous studies have shown that these three categories are perceptually very close each other. Our results show that while class *wood* is well separated from the other two classes the class *leather* is hard to separate from class of *fabric*, where the best discrimination is offered by PC3 dominant for *genuineness* and *height*. Although this behavior depends on intraclass variance of samples, we find a similar pattern in PCA analysis of Tanaka et al. [35]. As to other

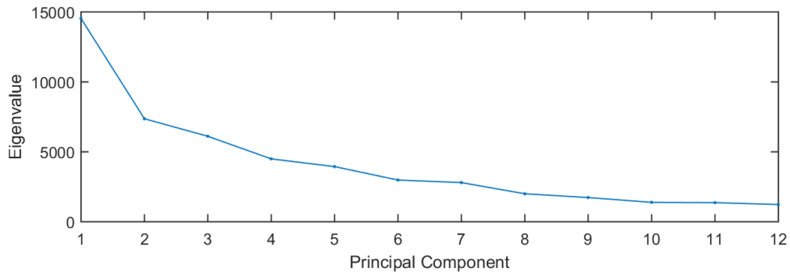



Fig. 13. Eigenvalues of PCA analysis.

Table 3. Loading of the First Two Principal Components for Individual Attributes

| | PC 1 | PC 2 | PC 3 |
|---|---|---|---|
| glossiness | −0.0098 | 0.1759 | **−0.4302** |
| colorfulness | 0.1924 | 0.0019 | −0.0802 |
| directionality | **0.4035** | −0.1089 | −0.0598 |
| diversity | **0.4187** | −0.1056 | −0.3123 |
| graininess | **0.3606** | −0.1845 | −0.0822 |
| regularity | **0.4164** | −0.1294 | −0.2344 |
| hardness | 0.2950 | 0.2923 | **0.5669** |
| roughness | 0.3310 | −0.2219 | 0.2111 |
| height | 0.1721 | −0.2125 | 0.1317 |
| genuineness | 0.2636 | **0.4623** | **0.3522** |
| quality | 0.0962 | **0.5135** | −0.2505 |
| attractiveness | 0.1280 | **0.4880** | −0.2749 |

materials tested, we have found that class of rough fabrics (denoted as *carpet*) is clearly separable in all PCs, while class of *plastic* cannot be well discriminated using the three PCs. This might be due to the fact that this class also contains samples imitating appearance of structured wood and leather.

Distribution of the samples for the first three PCs in each experiment is shown in Figure 14. Circles for each material represent projected positions of averaged rating scores for all nine participants. Both PCA and MANOVA (also used in this section) provide low-dimensional descriptions of multivariate data. However, while PCA searches for components describing data variability and describes the inter-entity variance structure, MANOVA searches for components that optimize group differences, i.e., data separability, and thus describes the inter-group variance structure. Therefore, one can find a more clear separation of individual classes using components resulting from MANOVA experiment in Figure 11, though the conclusions on data dimensionality and classes proximity remain the same as for the PCA.

We also show results of selected attributes on three main groups of materials (*fabric*, *leather*, *wood*) in comparison with recent research [7, 35]. The selected attributes are those that are the same or have very close meaning, i.e., *naturalness*≈*genuineness* and *prettiness*≈*attractiveness*. In all cases, we normalized the range of scores to an interval [0,1]. In Figure 15 one can see that while our results have, in general, slightly lower absolute values, the trends and relative distances across the material classes are well preserved and are consistent with the previous work. Differences in *colorfulness* and *roughness* are possibly due to different types of materials used or to different types of stimuli (real-world images, images of a planar specimen, real specimen).

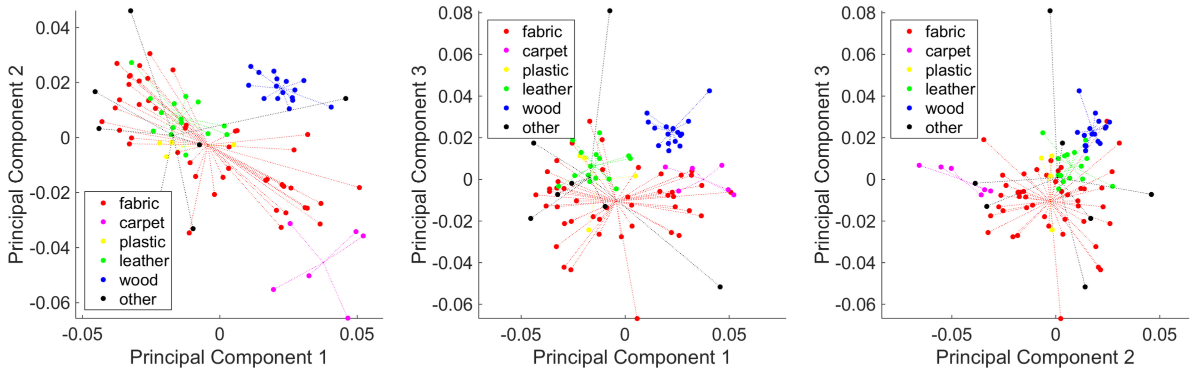



Fig. 14. Results analysis using PCA and distribution of the materials in the first three principal components. Lines join each sample to the projected mean location of each cluster.

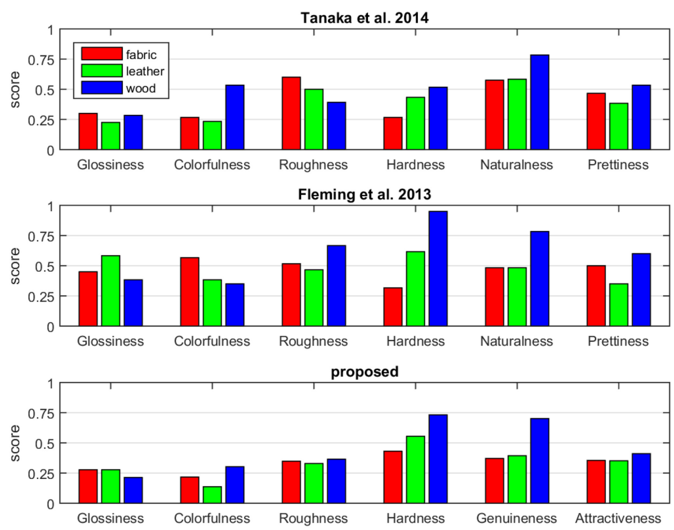



Fig. 15. A comparison of obtained results with state-of-the-art research [7, 35] on a subset of six attributes and three main material classes: *fabric*, *leather*, *wood*.

## 5 DISCUSSION

The proposed psychophysical study has several limitations. The first of them is a relatively low number of participants, which negatively impacts data variance and agreement. Next, the numbers of samples in individual material classes was, due to original database, not balanced. While fabric materials have the highest number of samples, some other categories have only several representatives. This makes a comparison of individual material categories more challenging.

In spite of these limitations, our study of perceptual qualities of selected materials from UTIA BRDF database has shown several interesting aspects. As expected, subjects' agreement was the best across physical attributes of materials (*glossiness* and *height*), while the worst for subjective attributes (*quality*, *attractiveness*). Our analysis also revealed a strong correlation of attributes related to visual and tactile structure of materials, such as *directionality*, *diversity*, *graininess*, and *roughness*, especially for fabric. In contrast, fabrics were considered attractive and of high quality for soft, smooth samples without strong material structure. The leather materials were

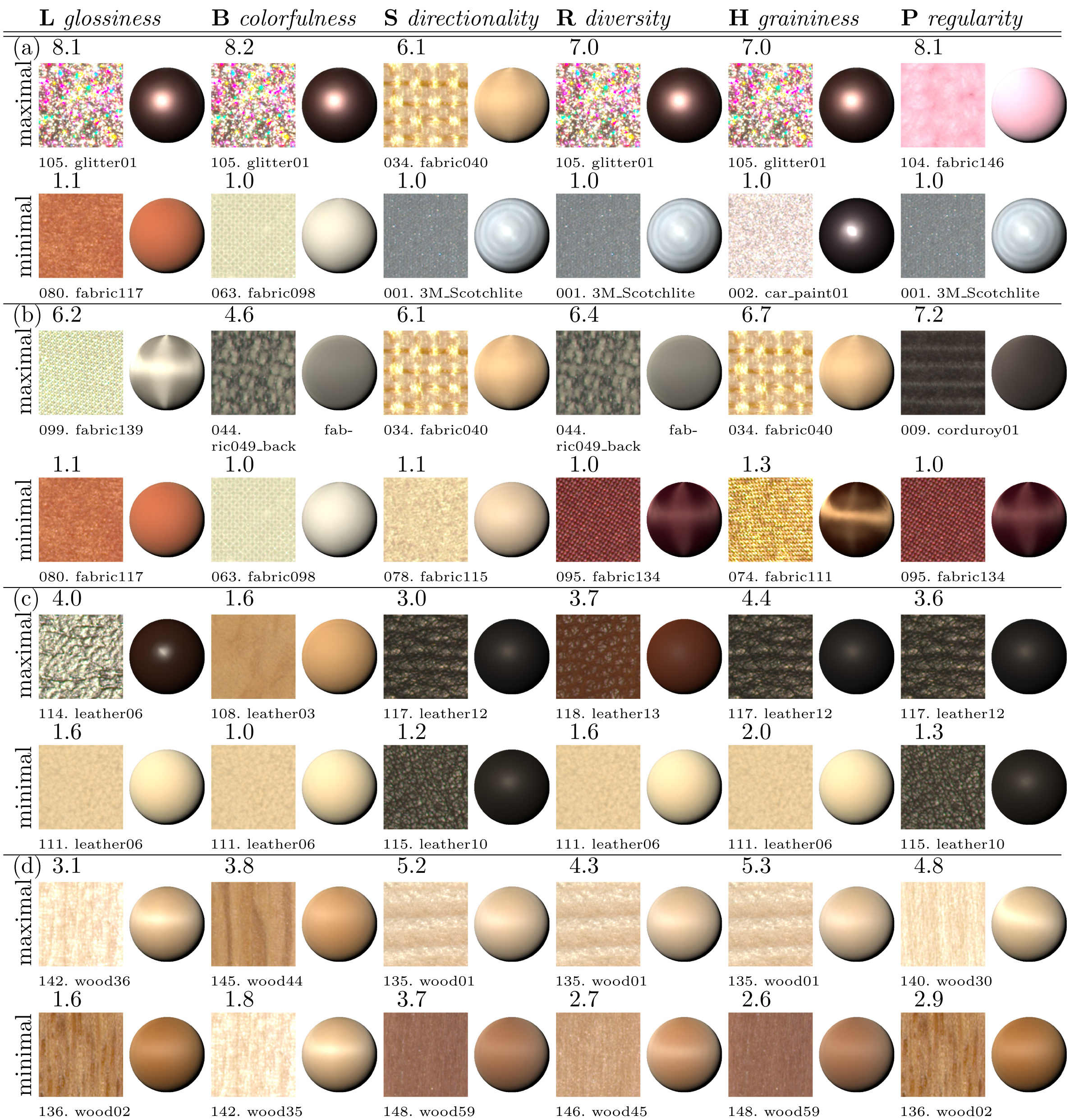



Fig. 16. A list of materials having maximal and minimal values of the visual attributes across (a) all materials, (b) fabric, (c) leather, and (d) wood.

considered the more natural when they were more soft and smooth. Our study also shows that the closest material categories are *fabric* and *leather*, while the most different material category is *carpet*.

Based on our results, follow-up researchers can adjust an appropriate portfolio of materials remaining to be analyzed. Although the span of materials in the tested database is limited, we consider the presented study as an additional convenient source of information to designers and virtual artists using the UTIA BRDF database. The


Fig. 17. A list of materials having maximal and minimal values of the tactile and subjective attributes across (a) all materials, (b) fabric, (c) leather, and (d) wood.

user ratings and their thorough analysis presented here can aid them in selection of appropriate corresponding BRDFs for their application. As an example of such information, we show examples of materials having maximal and minimal values of perceptual attributes for an entire tested dataset and for its subsets: *fabric*, *leather*, and *wood*. Figure 16 shows results for visual attributes while Figure 17 shows results for tactile and subjective ones. The figures include images of material structure, renderings BRDF on a sphere for a point light, numbers and names of the materials in the BRDF database, and actual values of the attributes. Here the results are consistent with an intuitive expectation for most of the attributes. See the supplementary material for raw subject data and MOS for individual attributes and all tested materials. To encourage complementary analysis of these materials, specimens of a majority of them are available and can be distributed upon individual requests.

## 6 CONCLUSIONS

This article analyzed 12 perceptual attributes of a subset of 93 materials from the UTIA BRDF database containing fabric, leather, wood, plastic, and other materials. Our user study with nine subjects revealed interesting relationships between both perceptual attributes and categories of visual, tactile, and subjective attributes. We also provide an insight as to how subjects' perceptual attributes assignment differs for three major material classes: fabric, leather, and wood. Due to the practical context of our work, we expect that practitioners in computer graphics will use our data as an additional reference source when working with UTIA BRDF database. In a future work, we plan to identify links between perceptual attributes and computational features of tested materials.

## ACKNOWLEDGMENTS

We thank the reviewers for their insightful comments and all participants for undertaking the demanding user study. We also thank to our colleague Radomír Vávra for proofreading and advices for improvement.

## REFERENCES

[1] P. Brodatz. 1966. *A Photographic Album for Artists and Designers (Brodatz Texture Database)*. Dover Publications.
[2] J. Filip, M. J. Chantler, P. R. Green, and M. Haindl. 2008. A psychophysically validated metric for bidirectional texture data reduction. *ACM Trans. Graph.* 27, 5 (Dec. 2008), 138.
[3] J. Filip and R. Vavra. 2014. Template-based sampling of anisotropic BRDFs. *Comput. Graph. Forum* 33, 7 (Oct. 2014), 91–99. DOI: https://doi.org/10.1111/cgf.12477
[4] J. Filip, R. Vavra, and M. Havlicek. 2014. Effective acquisition of dense anisotropic BRDF. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR'14)*. 2047–2052.
[5] Roland W. Fleming. 2014. Visual perception of materials and their properties. *Vision Res.* 94, 0 (2014), 62–75.
[6] R. W. Fleming, R. O. Dror, and E. H. Adelson. 2003. Real-world illumination and perception of surface reflectance properties. *Journal of Vision* 3, 5 (2003), 347–368.
[7] Roland W. Fleming, Christiane Wiebel, and Karl Gegenfurtner. 2013. Perceptual qualities and material classes. *J. Vision* 13, 8 (2013), 9–9.
[8] Johannes Günther, Tongbo Chen, Michael Goesele, Ingo Wald, and Hans-Peter Seidel. 2005. Efficient acquisition and realistic rendering of car paint. In *Proceedings of the 10th Workshop on Vision, Modeling, and Visualization (VMV'05)*.
[9] Andrew F. Hayes and Klaus Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. *Commun. Methods Measures* 1, 1 (2007), 77–89.
[10] C. Heaps and S. Handel. 1999. Similarity and features of natural textures. *J. Exp. Psychol.: Human Percept. Perform.* 25, 2 (1999), 299.
[11] Y. X. Ho, M. S. Landy, and L. T. Maloney. 2007. Conjoint measurement of gloss and surface texture. *Psychol. Sci.* 19 (2007), 194–204.
[12] ITU. 2008. *ITU-R.REC.P.910. Subjective Audivisual Quality Assessment Methods for Multimedia Applications.* Technical Report.
[13] A. Jarabo, H. Wu, J. Dorsey, H. Rushmeier, and D. Gutierrez. 2014. Effects of approximate filtering on the appearance of bidirectional texture functions. *IEEE Trans. Visual. Comput. Graph.* 20, 6 (June 2014), 880–892. DOI: https://doi.org/10.1109/TVCG.2014.2312016
[14] B. Keelan. 2003. ISO 20462: A psychophysical image quality measurement standard. In *Proceedings of the SPIE (SPIE'03)*, vol. 5294. 181–189.
[15] B.-G. Khang, J. J. Koenderink, and A. M. L. Kappers. 2006. Perception of illumination direction in images of 3D convex objects: Influence of surface materials and light fields. *Perception* 35, 5 (2006), 625–645.
[16] Michael S. Landy and Norma Graham. 2004. Visual perception of texture. In *The Visual Neurosciences.* MIT Press, 1106–1118.
[17] H. Long and W. K. Leow. 2002. A hybrid model for invariant and perceptual texture mapping. In *Proceedings of the 16th International Conference on Pattern Recognition*, Vol. 1. IEEE, 135–138.
[18] J. Malik and P. Perona. 1990. Preattentive texture discrimination with early vision mechanisms. *J. Optic. Soc. Amer.* 7, 5 (1990), 923–932.
[19] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, and Kenneth E. Torrance. 2000. Image-based bidirectional reflectance distribution function measurement. *Appl. Opt.* 39, 16 (June 2000), 2592–2600.
[20] Rodrigo Martín, Julian Iseringhausen, Michael Weinmann, and Matthias B. Hullin. 2015. Multimodal perception of material properties. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception (SAP'15)*. 33–40. DOI: https://doi.org/10.1145/2804408.2804420

[21] W. Matusik, H. Pfister, M. Brand, and L. McMillan. 2003. A data-driven reflectance model. *ACM Trans. Graph.* 22, 3 (2003), 759–769.

[22] A. Mojsilovic, J. Kovacevic, D. Kall, R. J. Safranek, and S. Kicha Ganapathy. 2000. The vocabulary and grammar of color patterns. *IEEE Trans. Image Process.* 9, 3 (2000), 417–431.

[23] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson. 2007. Image statistics and the perception of surface qualities. *Nature* 447, 10 (2007), 206–209.

[24] A. Ngan, F. Durand, and W. Matusik. 2005. Experimental analysis of BRDF models. *Eurograp. Symp. Render.* 2 (2005), 117–126.

[25] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsburg, and T. Limperis. 1977. Geometrical considerations and nomenclature for reflectance. *NBS Monograph 160* (1977), 1–52.

[26] S. Padilla, O. Drbohlav, P. R. Green, A. Spence, and M. J. Chantler. 2008. Perceived roughness of $1/f^{\beta}$ noise surfaces. *Vision Res.* 48, 17 (2008), 1791–1797.

[27] S. C. Pont, P. Sen, and P. Hanrahan. 2007. $2\frac{1}{2}$D texture mapping: Real-time perceptual surface roughening. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization.* 69–72.

[28] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. 2007. Visual equivalence: Towards a new standard for image fidelity. *ACM Trans. Graph.* 26, 3 (2007), 76:1–76:10.

[29] A. Ravishankar Rao and G. L. Lohse. 1996. Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Res.* 36, 11 (1996), 1649–1669.

[30] G. Schwartz and K. Nishino. 2013. Visual material traits: Recognizing per-pixel material context. In *Proceedings of the IEEE International Conference on Computer Vision Workshops.* 883–890.

[31] G. Schwartz and K. Nishino. 2015. Automatically discovering local visual material attributes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15).* 3565–3573.

[32] Ana Serrano, Diego Gutierrez, Karol Myszkowski, Hans-Peter Seidel, and Belen Masia. 2016. An intuitive control space for material appearance. *ACM Trans. Graph.* 35, 6, Article 186 (Nov. 2016). DOI: https://doi.org/10.1145/2980179.2980242

[33] Lavanya Sharan, Ce Liu, Ruth Rosenholtz, and Edward H. Adelson. 2013. Recognizing materials using perceptually inspired features. *Int. J. Comput. Vision* 103, 3 (July 2013), 348–371.

[34] H. Tamura, S. Mori, and T. Yamawaki. 1978. Textural features corresponding to visual perception. *IEEE Trans. Syst., Man Cybernet.* 8, 6 (1978), 460–473.

[35] Midori Tanaka and Takahiko Horiuchi. 2015. Investigating perceptual qualities of static surface appearance using real materials and displayed images. *Vision Res.* 115 (2015), 246–258.

[36] S. F. te Pas and S. C. Pont. 2005. A comparison of material and illumination discrimination performance for real rough, real smooth and computer generated smooth spheres. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization.* 57–58.

[37] S. F. te Pas and S. C. Pont. 2005. Estimations of light-source direction depend critically on material BRDFs. *Percept., ECVP Abstract Suppl.* 34 (2005), 212.

[38] P. Vangorp, J. Laurijssen, and P. Dutre. 2007. The influence of shape on the perception of material reflectance. *ACM Trans. Graph.* 26, 3 (2007), 77:1–77:10.

[39] M. Vanrell and J. Vitria. 1997. A four-dimensional texture representation space. In *Pattern Recognition and Image Analysis*, Vol. 1. 245–250.

[40] M. Vanrell, J. Vitria, and X. Roca. 1997. A multidimensional scaling approach to explore the behavior of a texture perception algorithm. *Mach. Vision Appl.* 9, 5/6 (1997), 262–271.

[41] Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a perceptual space for gloss. *ACM Trans. Graph.* 28, 4, Article 103 (Sept. 2009), 15 pages.


Received August 2017; revised September 2018; accepted September 2018