https://doi.org/10.2352/lim.2023.4.1.27

Towards High-Level, Intuitive Descriptors of Material Appearance

Belen Masia; Universidad de Zaragoza - I3A; Zaragoza, Spain

Abstract

Material appearance perception depends both on the physical interaction between light and material, and on how our visual system processes the information reaching our eyes. Currently, there is a disconnect between the physical properties underlying material appearance models used in simulations, and perceptual properties that humans rely on when interpreting visual depictions of material appearance. Our goal is to bridge this gap, creating high-level, intuitive descriptors of visual appearance that are linked to the underlying physical properties of the material. This can in turn benefit final applications such as material appearance editing, acquisition, gamut mapping, or compression. Here, we review two approaches proposing representations of appearance that are better aligned with human perception: one based on the use of intuitive attributes, and another exploring the use of natural language descriptions of material appearance. All our data and models are publicly available.

Introduction

Our perception of material appearance depends on the physical interaction between light and material, but also on the final image that reaches our retina and the processing that the human visual system does of it. The former can be accurately simulated for a wide variety of materials by means of light transport algorithms, which work with physical magnitudes. However, when humans need to interact with the resulting material depictions –for instance, to edit or to specify a certain appearance–, these physical parameters and simulations can fall short, because humans do not interpret visual information in terms of physical magnitudes. An example of this is the depiction of materials in art by the Dutch and Flemish masters of the 17th century: They were experts at faithfully representing complex material appearances such as satin or velvet, yet they were not doing so in a physically-accurate manner (see Fig. 1).

Two ideas emerge from this that motivate our work: First, there is a disconnect between the physical properties underlying material appearance models used in simulations, and perceptual properties that humans rely on when interpreting visual depictions of material appearance. Second, there are alternative representations of appearance that can be better correlated with the way humans perceive it. Our goal is to bridge the gap, creating high-level, intuitive descriptors of visual appearance that are linked to the underlying physical properties of the material. This can in turn benefit final applications such as material appearance editing, acquisition, gamut mapping, or compression.

A critical aspect in material appearance modeling is that the perceived appearance does not only depend on the material's properties. We define material appearance as "the visual impression we have of a material" [1], and it is also influenced by extrinsic factors, including the geometry of the object, the illumination conditions, or the viewpoint, as well as human perception [2, 3]. Multiple studies have been conducted to analyze the role of these factors on appearance perception [4, 5].

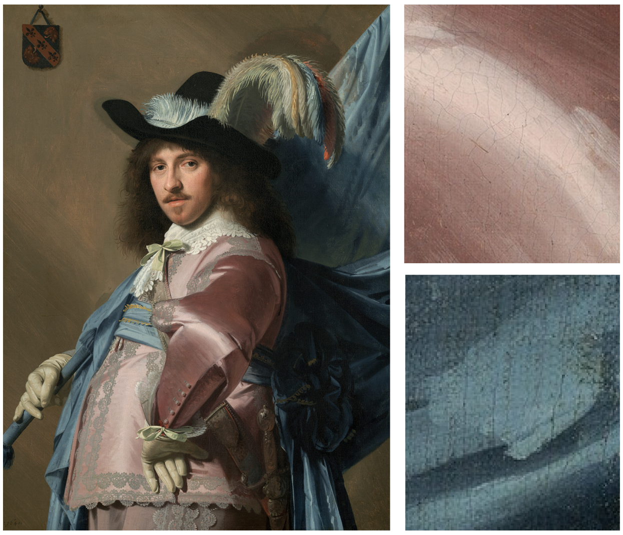

Figure 1: *Andries Stilte as a Standard Bearer*, by Johannes Cornelisz Verspronck (1640). The complexity of material appearance perception is illustrated by this painting, where the highly realistic appearance of the materials depicted in it disappears when looking at close-up regions. Image courtesy of Diego Gutierrez.

Motivated by the influence of extrinsic factors, and by the ability of our visual system to achieve perceptual constancy, we do not restrict ourselves to working in material space (e.g., BRDF space), but rather favor working in image space, in which representations of the material (images) are closer to the proximal stimuli reaching our visual system. Besides, and as an effective means to extract relevant features and model the complex, non-linear interactions between them that ultimately lead to our perception of appearance, we often resort to learning approaches that are trained on human judgements of appearance. An example of this is our work on building a similarity measure of material appearance [6]. In it, we first build a large dataset of images depicting materials, each one seen from different geometries and illuminations, and gather similarity judgements for them (in the form of triplets, with a 2AFC paradigm). We then train a model based on a deep learning architecture, which learns a feature space for materials that correlates with such perceived appearance similarity[1]. The key idea here is precisely that we combine the information about the physical properties of the material contained in the images, since we have the same material under different geometries and lighting conditions, with the subjective data on appearance similarity. A traditional image similarity metric would not be able to generalize across shape or illumination, while a BRDF-based metric would be unable to predict human similarity judgements.

The rest of this paper reviews two efforts towards building high-level, intuitive representations of material appearance. Both

[1] Full dataset and model available at: `http://webdiis.unizar.es/~mlagunas/publication/material-similarity/`

are data-driven approaches in which perceptually-meaningful latent spaces are learnt from a combination of carefully-crafted image datasets with corresponding human subjective data. The first work explores the use of high-level attributes to specify and edit material appearance in a predictable, intuitive manner. The second focuses on natural language descriptions of (fabric) materials, and leverages a specialized dataset of images and descriptions to fine-tune large vision-language models, creating a meaningful latent space for fabric appearance that enables applications such as fine-grained material retrieval and automatic captioning.

Editing Appearance via High-level Attributes

Material appearance models have become very successful at conveying extremely photorealistic appearance at manageable rendering costs. Whether these models are analytical, procedurally-generated, or data-driven, they all share a common limitation: the difficulty of *editing* a given material appearance to generate desired variations of it. Among these types of representations, data-driven ones such as BTFs or measured BRDFs, usually resulting from sophisticated capture setups, are notoriously difficult to manipulate and control. This is mainly due to the high-dimensional and non-linear nature of their parameter spaces, unaligned with perceptual dimensions of appearance. We thus focus here on this type of materials, and specifically on measured BRDFs.

Numerous works have tried to facilitate material appearance editing, from the seminal work of Pellacini et al. [7] building a perceptually-meaningful space for gloss, to appearance models designed for optimal balance between simplicity, robustness and artistic control [8]. For the particular case of data-driven models, the work of Matusik et al. [9] was particularly influential, both to our and other works: they applied dimensionality reduction techniques, defined perceptual traits over the low-dimensional space, classified materials according to these traits (in a binary manner, whether a material possessed a trait or not), and used this to define navigation directions along the resulting space. When attempting to build frameworks for intuitive editing of appearance, two main questions arise: First, which dimensions (or parameters, or attributes) defining appearance should be exposed to the user. Second, how are these dimensions aligned (or how can we align them) with the underlying physically-based model, amenable to rendering pipelines. Our work aims to answer both.

Finding a set of attributes or parameters that provide an intuitive representation of material appearance is a long-standing problem, for which no definite solution or methodology exist [10], and naming can further depend on the field [2]. The set of attributes must be reduced enough to be manageable, but comprehensive enough to allow for rich appearance edits, even for inexperienced users. In our case, we compiled an extensive list of attributes used in the literature from both industry and academia, and reduced it to fourteen attributes by means of a user study involving 60 stimuli and 26 participants (see original paper for details [11]). These attributes, covering both high- and mid-level features, are: 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.

Having established a set of attributes, the next step is to associate it to the underlying model of material appearance. Significant reduction of the high dimensionality of measured BRDFs is attained through PCA decomposition after log-relative linear mapping of the reflectance data; we retain the first five principal components, which are loosely related to characteristics of material appearance [12]. We then seek to *learn* a mapping between the attributes that a given material exhibits, and the underlying coefficients in the aforementioned PCA basis; this requires both training data, and a model for such mapping. Training data is collected in the form of Likert-scale ratings per attribute per material sample (400 materials), in a large-scale crowdsourced experiment with 400 participants and a total of 56,000 ratings. As a model we use radial basis function networks with one hidden layer: for each attribute we train one network, mapping the five dimensions of the PCA representation to the corresponding value of the attribute. Editing the value of an attribute to modify a given BRDF, which is our ultimate goal, therefore requires the inverse of this mapping. The solution to this inverse problem is not unique, and we therefore formulate it as a minimization, which we solve via gradient descent. This enables editing the appearance of measured BRDFs by means of high-level, intuitive attributes, as shown in Fig. 2 for several examples. More details can be found in the original paper [11], and the code and training dataset are publicly available[2].

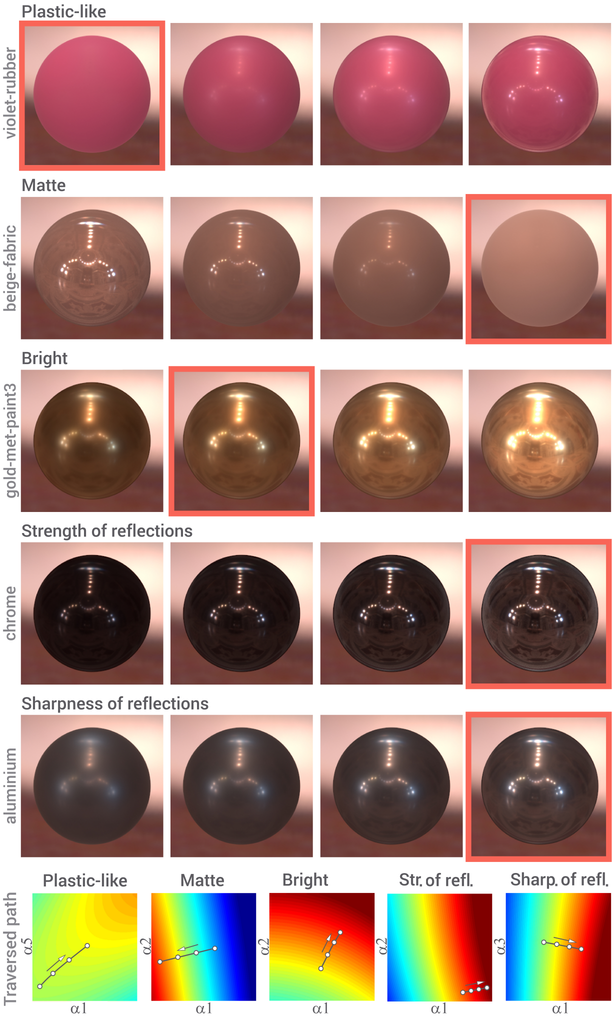

Figure 2: Editing measured BRDFs by varying the values of our high-level attributes. Each row shows a measured BRDF from the MERL database [9] (framed in light red), and the results obtained when linearly increasing or decreasing the value of the specified attribute in our space. The last row depicts the path followed in the 5D PCA space (we show the most representative 2D slice) when computing each of the other rows. Image from [11].

[2] Please refer to the project webpage: `http://webdiis.unizar.es/~aserrano/projects/Material-Appearance.html`

Our work enables not only editing, but also other applications such as attribute-specific similarity computation, or BRDF gamut mapping [13]. Follow up works have resorted to deep learning-based models to learn appearance representations, ranging from generative models that enable a certain degree of editing and morphing [14, 15], to models that focus on reliably predicting gloss and other attributes across varying geometries and illuminations [16]. Further, the construction of low-dimensional latent spaces of material appearance has been demonstrated both in supervised [17] and unsupervised learning setups [18, 19]. While these approaches are extremely promising, their degree of control, robustness and ability to generalize deserves further exploration. Finally, a recent work [20] has focused on creating a perceptual embedding for editing using NMDS and crowdsourced subjective data; different to our work, they determine the perceptual traits a posteriori, based on the collected similarity data. Their approach is demonstrated on metal-related materials, and extension to other categories remains as promising future work.

Describing Appearance using Natural Language

Natural language is a high-level, intuitive, accessible and common means of communicating information. Being able to reliably and precisely use natural language as a descriptor for applications such as generation, editing or retrieval of material appearance would immensely facilitate these applications. Not only would it help reduce their initial learning curve and make them more accessible to users from different backgrounds, but could also potentially help build a unified, universal space for representing appearance.

In opposition to its intuitiveness and "ease of use", using natural language as a descriptor of appearance poses numerous questions. First, we need to know whether there is a common lexicon and structure, shared by most people, when describing material appearance using natural language. If variability were too high, using natural language as a robust, universal means to reliably describe appearance could be unfeasible. Second, it is unclear whether natural language alone would be effective for precisely communicating material appearance: Can natural language descriptions discriminate between two material samples of similar appearance, and to what extent? These are open questions for the community, and we have taken initial steps towards answering them.

At the same time, and provided we had positive answers to the previous questions, we need models capable of taking natural language as one of their inputs. We have recently witnessed an impressive leap forward in natural language processing models, and in particular in what large vision-language models can achieve by linking visual content to text. Models such as CLIP [21] or BLIP [22] are capable of creating latent spaces where the representations of images and their natural language descriptions are close together (BLIP further includes a generative stage for image captioning), effectively linking them. They are, however, supervised models trained on hundreds of millions of image-text pairs. Therefore, questions arise as to whether we can directly use them for material appearance concepts; and, if not, whether we can adapt them, e.g., through fine-tuning, for their use in material appearance-related areas.

Our work in this area seeks to answer both sets of questions: those related to the existence of a shared understanding of language as it relates to material (fabric) appearance, and those related to the use of large vision-language models for tasks requiring fine-grained representations of appearance. Since the space of material appearance is extremely vast, to keep the task tractable, we focus on *fabric* materials. This allows us to validate our methodology in a reasonably constrained yet sufficiently expressive subset: fabrics exhibit a large variability in reflectance, colors, structure, or patterns, while being an ubiquitous and widely familiar material class.

We make three main contributions towards this goal. First, we collect a dataset[3] linking 3,000 photorealistic images of fabric materials to 15,000+ free-text natural language descriptions of such fabrics, provided by participants in a user study who were native English speakers and were familiar with fashion or design (Fig. 4 includes some sample descriptions provided by humans). Second, we analyze the descriptions provided and find that: (i) there is a common lexicon (ca. 500 words are enough to cover 95% of the 15,000+ descriptions gathered), (ii) there are common attributes (we identify eleven traits or attributes that emerge from the descriptions), (iii) there is a common structure followed by users when describing appearance, and (iv) there is high similarity between descriptions of the same fabric given by different users. These insights provide a foundation for our third contribution, in which we explore the use of large vision-language models for applications such as: fine-grained text-based retrieval, image-based search, and description generation of fabric images (see Figs. 3 and 4). We show how fine-tuning these models with a low amount of high-quality, specialized data provides a significant improvement over the native versions of these models, trained only on their original datasets. We refer the reader to the original publication [23] for more details on each contribution. Our work suggests that natural language descriptions may indeed be sufficient to convey material appearance in a fine-grained manner, and that learning-based models can link such descriptions to robust latent representations of appearance, which can in turn be used for a variety of applications.

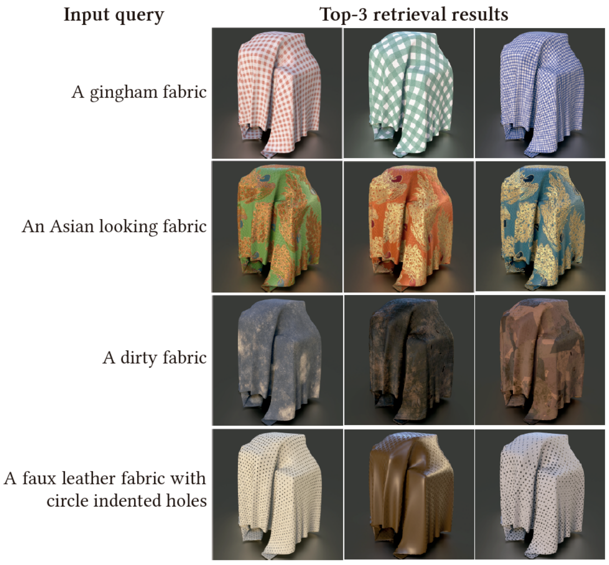

Figure 3: Examples of text-based retrieval results obtained with our model (the CLIP model, fine-tuned with our specialized dataset). We show input text queries, including domain-specific concepts, and the top-3 retrieval results. Image from [23].

Future Directions

Numerous challenges and interesting avenues of work lie ahead; to name a few: the search for a unified, general treat-

[3] The dataset is publicly available at: `https://valentin.deschaintre.fr/text2fabric`

Input	Human description	Ours	Native BLIP
	This fabric resembles canvas and is woven with black and blue yarn giving it a dull gray appearance from afar. It has a slightly rough texture and no sheen. There are golden sunbursts that appear to made with fabric paint.	This fabric has a blue background with small yellow suns all over it the sun is designed in repeating lines of five - point stars the linen thread could be used to make curtains or upholster	This is a print of gold flowers in an ocean blue background, and placed over the black
	This is blue cotton fabric. It is made up of yellow colored sunlight patterns. it appears to be soft and is mostly used for cushions and table linen.	This fabric is a muted gold / pale yellow on a slate blue it has an appearance of being woven, like a sun with five - point star patterns similar to the stars themselves	The cover is made out of blue and yellow flowers
	This is a medium weight fabric. The base color is a coffee brown. The overall pattern is continual broken zigzag lines of copper orange throughout.	This fabric is black in color, with a pattern of metallic gold zigzags these are arranged diagonally throughout the material it looks tightly knit and soft to the touch	The navy blue chey print on this tie is made of silk and has a thin pattern
	This fabric is dark gray with zigzag designs throughout the fabric in orange. Its texture is soft and smooth. The material seems most likely to be polyester. The orange zigzag features a little sparkle.	This fabric is dark gray with a pattern consisting of many jagged triangles throughout these triangular shapes are gold the material appears shiny and smooth, even satin it would make an elegant dress	A blue scarf with orange pattern on it

Figure 4: Examples of description generation results obtained with our model (the BLIP model, fine-tuned with our specialized dataset) for two fabric samples (unseen during training). *From left to right:* Descriptions provided by humans, by our model, and by native BLIP (trained on the original BLIP dataset, with no fine-tuning). Fine-tuning on a relatively small amount of specialized data clearly improves the performance over the original model trained on a large amount of generic data (see text for details). Image from [23].

ment of different material types or categories, which is a longstanding goal amidst a fragmented space of solutions; the use of neural networks to represent BRDFs, which despite its recentness is quickly gaining traction in the field; or the reliance on large quantities of gathered data, that is a challenge of current learning-based solutions, but for which recent results of unsupervised learning approaches are promising. Last, but not least, we need to continue in our quest to fully understand the behavior and operation of the human visual system.

Acknowledgments

We would like to thank all co-authors of the works described here, as well as the reviewers of such works, and the participants in the user studies. This work has received funding from the European Research Council (ERC) within the EU's Horizon 2020 research and innovation programme (project CHAMELEON, Grant No. 682080) and under the Marie Skłodowska-Curie grant agreement No. 956585 (project PRIME).

References

[1] J. Dorsey, H. Rushmeier, and F. Sillion. *Digital modeling of material appearance*. Elsevier, 2010.

[2] E. H. Adelson. On seeing stuff. *Human Vision and Electronic Imaging VI*, 4299:1–13, 2001.

[3] R. Fleming. Visual perception of materials and their properties. *Vision Research*, 94:62–75, 2014.

[4] M. Olkkonen and D. H. Brainard. Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception*, 2(9):1014–1034, 2011.

[5] M. Lagunas, A. Serrano, D. Gutierrez, and B. Masia. The joint role of geometry and illumination on material recognition. *Journal of Vision*, 21(2), 2021.

[6] M. Lagunas, S. Malpica, A. Serrano, E. Garces, D. Gutierrez, and B. Masia. A similarity measure for material appearance. *ACM Trans. Graph.*, 38(4), 2019.

[7] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. *Proceedings of ACM SIGGRAPH*, pages 55–64, 2000.

[8] B. Burley. Physically Based Shading at Disney. SIGGRAPH Courses, 2012.

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

[10] A. K. R. Choudhury. *Principles of Colour and Appearance Measurement*. Woodhead Publishing, 2014.

[11] A. Serrano, D. Gutierrez, K. Myszkowski, H.-P. Seidel, and B. Masia. An intuitive control space for material appearance. *ACM Trans. Graph.*, 35(6), 2016.

[12] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi. On optimal, minimal brdf sampling for reflectance acquisition. *ACM Trans. Graph.*, 34(6), 2015.

[13] T. Sun, A. Serrano, D. Gutierrez, and B. Masia. Attribute-preserving gamut mapping of measured BRDFs. *Computer Graphics Forum*, 36(4):47–54, 2017.

[14] Y. Guo, C. Smith, M. Hašan, K. Sunkavalli, and S. Zhao. MaterialGAN: Reflectance Capture Using a Generative SVBRDF Model. *ACM Trans. Graph.*, 39(6), 2020.

[15] J. Delanoy, M. Lagunas, J. Condor, D. Gutierrez, and B. Masia. A generative framework for image-based editing of material appearance using perceptual attributes. *Computer Graphics Forum*, 41(1):453–464, 2022.

[16] A. Serrano, B. Chen, C. Wang, et al. The effect of shape and illumination on material perception: model and applications. *ACM Trans. Graph.*, 40(4):1–16, 2021.

[17] B. Hu, J. Guo, Y. Chen, M. Li, and Y. Guo. DeepBRDF: A Deep Representation for Manipulating Measured BRDF. *Computer Graphics Forum*, 2020.

[18] K. R. Storrs, B. L. Anderson, and R. W. Fleming. Unsupervised learning predicts human perception and misperception of gloss. *Nature Human Behaviour*, 5(10):1402–1417, 2021.

[19] C. Liao, M. Sawayama, and B. Xiao. Unsupervised learning reveals interpretable latent representations for translucency perception. *PLOS Computational Biology*, 19(2):e1010878, 2023.

[20] W. Shi, Z. Wang, C. Soler, and H. Rushmeier. A low-dimensional perceptual space for intuitive brdf editing. In *EGSR*, 2021.

[21] A. Radford, J. W. Kim, C. Hallacy, et al. Learning transferable visual models from natural language supervision. In *Intl. Conf. on Machine Learning*, pages 8748–8763, 2021.

[22] J. Li, D. Li, C. Xiong, and S. Hoi. BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. *International Conference on Machine Learning*, 2022.

[23] V. Deschaintre*, J. Guerrero-Viu*, D. Gutierrez, T. Boubekeur, and B. Masia. The visual language of fabrics. *ACM Trans. Graph.*, 2023.

Author Biography

Belen Masia (PhD 2013) is an Associate Professor in the Department of Computer Science at Universidad de Zaragoza. Her research focuses on the areas of material appearance, applied perception, and virtual reality. Masia is the recipient of a Eurographics Young Researcher Award (2017) and a Eurographics PhD Award (2015), among others. She has served as an Associate Editor for ACM Transactions on Graphics, Computers and Graphics, and ACM Transactions on Applied Perception.