The role of objective and subjective measures in material similarity learning

Johanna Delanoy
Universidad de Zaragoza, I3A

Manuel Lagunas
Universidad de Zaragoza, I3A

Ignacio Galve
Universidad de Zaragoza, I3A

Diego Gutierrez
Universidad de Zaragoza, I3A

Ana Serrano
Universidad de Zaragoza, I3A

Roland Fleming
Justus-Liebig-UniversitÃďt Giessen

Belen Masia
Universidad de Zaragoza, I3A

ABSTRACT

Establishing a robust measure for material similarity that correlates well with human perception is a long-standing problem. A recent work presented a deep learning model trained to produce a feature space that aligns with human perception by gathering *human subjective measures*. The resulting metric outperforms objective existing ones. In this work, we aim to understand whether this increased performance is a result of using human perceptual data or is due to the nature of feature learnt by deep learning models. We train similar networks with *objective measures* (BRDF similarity or classification task) and show that these networks can predict human judgements as well, suggesting that the non-linear features learnt by convolutional network might be a key to model material perception.

CCS CONCEPTS

• **Computing methodologies → Appearance and texture representations**.

KEYWORDS

neural networks, material perception

ACM Reference Format:

Johanna Delanoy, Manuel Lagunas, Ignacio Galve, Diego Gutierrez, Ana Serrano, Roland Fleming, and Belen Masia. 2020. The role of objective and subjective measures in material similarity learning. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The ability to determine the degree of similarity between two material samples is a key aspect of material appearance modeling, with applications in acquisition, editing, retrieval, rendering, compression, or sampling, to name a few. Establishing a robust measure for material appearance similarity that correlates well with human perception remains a long-standing problem. Traditional similarity metrics computed in BRDF space have been shown to correlate poorly with human perception [Serrano et al. 2016]. Alternatives include metrics computed in image space [Ngan et al. 2006; Pereira and Rusinkiewicz 2012], and works that have sought low-dimensional perceptual embeddings from which similarity metrics can be derived [Serrano et al. 2016; Soler et al. 2018; Wills et al. 2009]. Among the latter, a recent work presented a learning-based model that, taking an image with a rendition of a material as input, yields a representation of the material in a feature space that correlates with *perceived* material appearance similarity [Lagunas et al. 2019]. This model is based on a deep neural network (DNN), and trained with *human subjective measures* of similarity gathered through crowdsourcing.

Their learned model is shown to outperform existing objective metrics in reproducing human assessment of similarity, presumably thanks to (i) the subjective measures used during the training, as well as (ii) the ability of the model to learn deep features. Interestingly, it has been shown that deep learning features can lead to a representation that correlates with perceptual judgements [Zhang et al. 2018]. Consequently, we consider here to what extent each of the two characteristics above contribute to the success of the network, and whether subjective measures are actually needed.

In this work, we propose and analyze a learning-based model analogous to that of Lagunas et al. in architecture, loss function, and training scheme, but making use of an objective BRDF metric for training instead of the human subjective judgements used in the previous work. Then, we compare the two models (subjective and objective), analyzing which one correlates better with human judgements. We find that the ability of this proposed network to predict human judgements is almost on par with the network that actually used human subjective measures during training, suggesting that the non-linear deep features learned by the model may be the key to understanding and modeling material perception. We also compare these similarity networks with a standard network trained to classify materials as a reference task. Although this network is able to distinguish each material from the database, we show that it does not cluster perceptually close materials.

2 METHODOLOGY

We train three different networks, namely a *classification network* (trained to classify each image given the represented material), a *human similarity network* (original network from [Lagunas et al.

2019]) and a *BRDF similarity network* (analogous to the latter but trained to mimic the *Cube root cosine weighted* BRDF metric [Fores et al. 2012]).

We compute Representation Dissimilarity Matrices (RDMs) between a 10% random subset of the training images. The dissimilarity is measured as the $l2$ norm between normalized feature vectors. By organizing these matrices according to various properties over the images, we can understand how the space is organized and which are the properties that are best captured by the network. We used image properties (shape and illumination) and material properties (color of the material and reflectance properties).

3 ANALYSIS

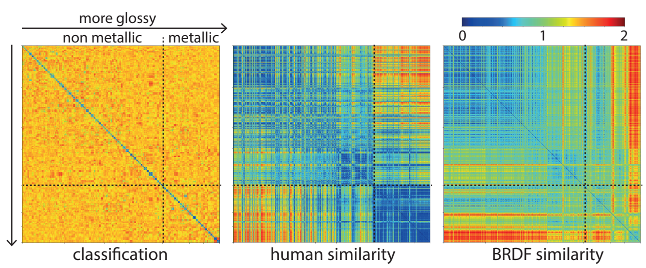

Figure 1: The RDMs are organized by reflectance for the three networks. The x- and y-axis are organized by surface reflectance properties of increasing glossiness : starting with diffuse materials to very glossy plastics, and following with metals. The separation between metallic and non metallic materials is depicted by the dotted black line.

Analysis of RDMs. In Figure 1, we show the RDMs organized by the reflectance properties of the materials. The classification network creates strong correlations only between images showing the same material but does not create relationships between ones with similar reflectance. Although classification networks have been used as a reference for perception metrics in recent works [Zhang et al. 2018], we show here that the resulting feature space does not align with material perception.

On the other side, the human similarity network strongly correlates materials that have a similar reflectance. We see a clear distinction between metallic and non metallic materials and big blocks for the completely diffuse materials (top-left block) and highly specular plastics (bottom-right blocks in the non-metallic materials). This shows that the network learned to distinguish between different reflectance properties in a way that correlates with human perception.

The RDM of the BRDF similarity network exhibits a less clear structure and reveals an unevenly distributed space. Diffuse materials appear to be in a small cluster in the space but all other materials are spread out with a majority of large distances. Notably, the network does not make any clear separation between metallic and non metallic materials, but seems to cluster materials with a similar glossiness (very glossy plastics and very glossy metals).

Agreement with human perception. We measure the agreement of the distance in feature space with human judgement in the same way as Lagunas et al. [2019], both for color and for gray-scale images. Results are summarized in Table 1.

Table 1: Agreement with human judgement for the three different networks, on color and gray-scale images. Agreement for the classification networks increases significantly when removing color. The two similarity networks performs almost equally.

	classification	human sim.	BRDF sim.
color	0.59	0.82	0.80
gray-scale	0.68	0.81	0.80

Although color seems to play a minor role for the two networks trained on material similarity, the accuracy increases significantly when using gray-scale images for the classification network (9% more agreement). This network thus seems to use colors in a way that human did not use when making their judgement. Removing the color bias allows the network to judge only on the reflectance of the material and get closer to human perception.

Notably, the BRDF similarity network yields very similar results to the human similarity network (80% agreement). In [Lagunas et al. 2019], it was shown that using the BRDF metric by itself aligns only 67% of the time with human judgements. However, the network trained with this metric creates a space that reflects human judgement as well as the one trained to directly mimic human perception. This suggests that the structure of the network leads to a representation that aligns with human perception, as soon as the network is fed with a metric that sufficiently represents human judgement.

ACKNOWLEDGMENTS

This research has been partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (CHAMELEON project, grant agreement No 682080), the Spanish Ministry of Economy and Competitiveness (project TIN2016-79710-P) and the German Research Foundation (project 222641018-SFB/TRR 135 TP C1)

REFERENCES

Adria Fores, James Ferwerda, and Jinwei Gu. 2012. Toward a perceptually based metric for BRDF modeling. In *Color and imaging conference*, Vol. 2012. Society for Imaging Science and Technology, 142–148.

Manuel Lagunas, Sandra Malpica, Ana Serrano, Elena Garces, Diego Gutierrez, and Belen Masia. 2019. A similarity measure for material appearance. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 135.

Addy Ngan, Frédo Durand, and Wojciech Matusik. 2006. Image-driven Navigation of Analytical BRDF Models.. In *Rendering Techniques*. 399–407.

Thiago Pereira and Szymon Rusinkiewicz. 2012. Gamut mapping spatially varying reflectance with an improved BRDF similarity metric. In *Computer graphics forum*, Vol. 31. Wiley Online Library, 1557–1566.

Ana Serrano, Diego Gutierrez, Karol Myszkowski, Hans-Peter Seidel, and Belen Masia. 2016. An intuitive control space for material appearance. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.

Cyril Soler, Kartic Subr, and Derek Nowrouzezahrai. 2018. A versatile parameterization for measured material manifolds. In *Computer graphics forum*, Vol. 37. Wiley Online Library, 135–144.

Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a perceptual space for gloss. *ACM Transactions on graphics (TOG)* 28, 4 (2009), 1–15.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 586–595.