https://doi.org/10.2352/ISSN.2169-2629.2018.26.294

Behavioral Investigation of Visual Appearance Assessment

Davit Gigilashvili, Jean-Baptiste Thomas, Jon Yngve Hardeberg, Marius Pedersen;
Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway

Abstract

The way people judge, assess and express appearance they perceive can dramatically vary from person to person. The objective of this study is to identify the research hypotheses and outline directions for the future work based on the tasks observers perform. The eventual goal is to understand how people perceive, judge, and assess appearance, and what are the factors impacting their assessments. A series of interviews were conducted in uncontrolled conditions where observers were asked to describe the appearance of the physical objects and to complete simple visual tasks, like ranking objects by their gloss or translucency. The interviews were filmed with the consent of the participants and the videos were subsequently analyzed. The analysis of the data has shown that while there are cross-individual differences and similarities, surface coarseness, shape, and dye mixture have significant effect on translucency and gloss perception.

Introduction and Motivation

Vision is one of the primary senses humans use to perceive and interpret the surrounding. "Visual perception is the ability to interpret the surrounding environment by processing information that is contained in the visible light" [1]. On the other hand, "appearance is the visual sensation through which an object is perceived to have attributes as size, shape, colour, texture, gloss, transparency, opacity etc." [2] Appearance is a complex psychophysical phenomenon that depends not only on the stimuli, but on a broad spectrum of various factors, e.g. memory of the observer [3]. For an easier understanding of appearance, it has been split into several distinct attributes that compose the appearance. CIE defines four major appearance attributes: color, gloss, translucency and texture [2, 4] that interact and influence each other [5, 6, 7].

Advances in computer graphics and simplicity of controlling the parameters have lead to widespread usage of synthetic images for appearance research (e.g. [8, 9]). On the other hand, RGB images of the real objects are frequently used for material appearance analysis, especially in computer vision (e.g. [10, 11, 12]). Despite the clear advantage of using synthetic or real images, the appearance and perception still differ from that of real-life situations. The interaction can be considered less natural due to the presence of the intermediate media and lack of the imperfections in synthetic images [13]. Lack of possibility to touch the objects, limited or no possibility to move them, and lack of the effect of the head movement can be named as further disadvantages of using images for studying appearance.

There has been examples of using real objects for studying appearance [14, 15]. However, experiments were held in controlled laboratory conditions, the observation geometry was fixed and observers were not allowed to touch the objects. This makes the setup artificial and is rarely to be encountered in real life.

Therefore, we decided to use real objects for our study; allowing observers to freely interact with them. The geometry of the measurement can impact the appearance. Bidirectional Reflectance Distribution Functions (BRDF) [16, 17], gloss [18, 19], or color [20, 21] are all measured for predefined geometries. However, observation geometries in real life vary a lot. This is the main reason why we allow the observers to freely interact with the objects. This is primarily a qualitative study to identify traits of appearance assessment by human observers. Analysis of the consistency of human behaviour might potentially outline the directions for further studies, and eventually leading to a better understanding of appearance perception.

The scope of this paper covers the results obtained from the experiments. Particular procedures and processes that lead observers to the results discussed below will be analyzed in the future work. Below we introduce the experimental setup, quantitative results of the experiments followed by the research hypotheses generated from them.

Experimental Setup

In our experiments, we used resin objects of the *Plastique* artwork collection described by Thomas *et al.* [13]. The objects are referred by their codes in task descriptions, as labelled in [13]. The collection of objects is composed of spheres, parallelepipeds, and female bust figures of three levels of surface coarseness and four hues (blue, yellow, white, and achromatic/transparent).

The interviews were conducted in uncontrolled conditions, under a mixture of daylight and artificial fluorescent illumination. The experimenter measured light intensity (in lux) and color temperature of the illumination (in Kelvin) with a light meter before and after the interview. The video and audio of the interview was recorded from two perspectives, front and side. A screenshot from a sample video can be seen in Figure 1. Nine boxes with different sets of the physical objects were used for eleven tasks of the interview (Figure 2). A checkerboard, a pen with text on it, and a white paper were placed on the table without explicit explanations. However, the participants were informed that they could freely interact with the objects. We expected that the white paper, as a homogeneous background, and a checkerboard, as a heterogeneous background, could be used by the observers for judging translucency. Besides, a pen with a text on it could be used to check whether reading through the object was possible. The observers were asked to wear gloves, in order to protect the objects.

17 observers, 11 men and 6 women have been interviewed in total with average age of 35.7 years. 4 out of them were the authors of this paper. 14 observers were experts in the field, while three of them were naïve to visual appearance studies. 2 observers were color deficient.

The interviewees were encouraged to explain their decisions and comment their actions while completing the tasks. The boxes

Figure 1: A screenshot from a sample video.

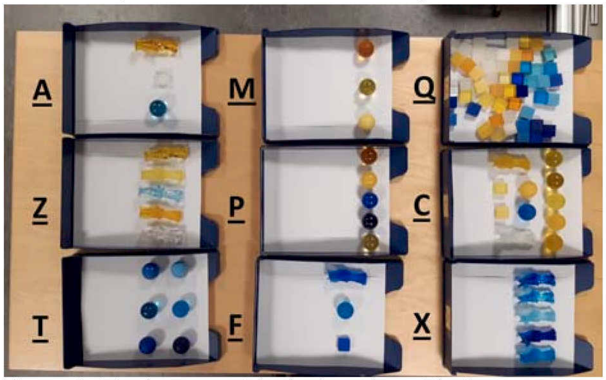


Figure 2: The boxes used in the interview. The letters are randomly assigned to the boxes and are not related to the appearance of the objects.

with respective oral instructions were introduced to the interviewees in the following order:

Task 1 (box Q):

- *Objects*: There are 48 rectangular parallelepipeds of different color, coarseness and translucency in the box.
- *Tasks:* 1) The first task is to cluster the objects into any number of groups the participant considers natural. 2) Experimenter asks the participant to discuss and explain the reasoning of clustering this way. 3) Experimenter asks the participant whether there could be any other way of creating groups that look natural. 4) Experimenter selects one of the groups of the cluster and asks the participant to sub-cluster this group even further.

Task 2 (box C):

- *Objects:* There are 5 yellow spheres of different coarseness and translucency in the box. Besides, there are 6 more objects: two female busts, two spheres and two rectangular parallelepipeds.
- *Tasks:* 1) The first task is to order the 5 spheres in any way the participant considers natural. They are encouraged to use any dimensions they think fit. 2) The participant is given 6 additional objects and is asked to locate the object in relation to the order he/she created with the first five spheres. The observer is expected to fail to order all objects within the created order, and thereby, to generate some questions how to locate the new object. The outcome is to identify potential cues to create an appearance ordering system.

Task 3 (box X):

- *Objects:* There are 5 blue female bust objects from the *Plastique* collection in the box. Object codes: 140, 154, 157, 158, 161.[13]
- *Tasks:* 1) The first task is to describe the appearance of the objects. Besides, the observers are asked, which objects look softer or harder, lighter or heavier, without touching them. 2) The participant can now touch the objects. The participant is asked to rank the object by their gloss/shine.

Task 4 (box M):

- *Objects:* There are 3 yellow spheres of different surface coarseness and translucency in the box. Object codes: 86, 95, 109.
- *Tasks:* 1) The first task is to describe the appearance of the objects with participants' own words. 2) The second task is to rank the object by their gloss/shine.

Task 5 (box P):

- *Objects:* There are 5 spheres of different colors, coarseness and translucency in the box. Object codes: 79, 82, 88, 94, 112.
- *Tasks:* 1) The first task is to describe the appearance of the objects with participants' own words. 2) The second task is to rank the object by their gloss/shine. The goal is to observe, whether difference in color and translucency impacts the result.

Task 6 (box F):

- *Objects:* There are 3 blue objects in the box: one sphere, one rectangular parallelepiped, and one female bust. Object codes: 42, 101, 155.
- *Tasks:* 1) The first task is to describe the appearance of the objects with participants' own words. 2) The second task is to rank the object by their translucency. However, word "translucency" is not be mentioned explicitly throughout the experiment, as it could be ambiguous for some of the interviewees; "how light is going through" is used instead.

Task 7 (box X):

- *Objects:* There are 5 blue female bust objects in the box. Although the box has already been used in the experiment, the experimenter has re-introduced the box in the pile discretely.
- *Tasks:* 1) The first task is to describe the appearance of the objects. 2) The second task is to rank the object by their translucency.

Task 8 (box A):

- *Objects:* There are 3 objects of different shape and color in the box: yellow female bust, achromatic rectangular parallelepiped, and blue sphere. Object codes: 2, 103, 151.
- *Tasks:* 1) The first task is to describe the appearance of the objects. 2) The second task is to rank the object by their translucency. The goal is to observe, whether color and shape impact translucency perception.

Task 9 (box Z):

- *Objects:* There are 5 female bust objects of different colors in the box. Object codes: 115, 152, 160, 163, 167.
- *Tasks:* 1) The first task is to describe the appearance of the objects. Besides, the observers are specially asked, which objects looks softer or harder, heavier or lighter, without touching them. 2) The participant can now touch the objects. The participant is asked to rank the object by their translucency.

Task 10 (box A):

- *Objects:* There are 3 objects of different shape, color, and surface coarseness in the box: yellow female bust, achromatic rectangular parallelepiped, and blue sphere. Although the box has already been used in the experiment, this is not revealed to the participant.

- *Tasks:* 1) The first task is to describe the appearance of the objects. 2) The second task is to rank the object by gloss/shine.

Task 11 (box T):

- *Objects:* There are 6 blue spheres of different surface coarseness and dye mixture in the box. Object codes: 75, 76, 80, 83, 100, 102.

- *Tasks:* 1) The first task is to describe the appearance of the objects. 2) The second task is to cluster them into "opaque" and "non-opaque" categories. We are interested, whether level of light transport is critical for opacity or transparency identification.

The objects used for tasks 3, 4, 5, 7 and 9 are labelled and illustrated on Figure 4.

Analysis and Results

We provide quantitative analysis on 9 boxes, while the first two ones will only be considered in a qualitative way due limited space. Nevertheless, the behavioral patterns and detailed analysis will be considered in a future communication. Behavioral patterns for boxes Q and C are very complex and therefore, left beyond the scope of this paper.

The ranking experiment results are quantified as follows: ranked objects are given points from 5 to 1, where 5 points correspond to the most glossy/translucent one. In case of ties, the average point of the tied objects is assigned to each of them. For instance, if first three objects are tied, each of them gets 4 points, while if only first two are tied, each gets 4.5 points.

The results are visualized as boxplots, given on Figure 3. In order to check statistical significance of the differences, ranked objects were considered as pairs. Afterwards, sign tests have been conducted and Bonferroni correction [22] was applied to avoid the bias due to the multiple testing. Alpha threshold was set to 0.05.

It is worth mentioning that the experimental protocol was not identical for all observers. Some observers were clearly instructed that they could have ties, while in other cases, this was not clearly mentioned by the interviewer. Therefore, the observers might have assumed that they were forced to choose and cross-individual differences might be accounted for this factor.

Task 1 (box Q)

Color or hue was a dominant attribute used by the observers to group the objects. 13 out of 17 participants used this single criterion for clustering, while the criteria used by 4 other observers were the combination of color and translucency, transparency, "surface properties", and "material properties". However, the number of groups created based on color varied, leading to a color naming problem. The second level criteria were mostly gloss and translucency, either separately, or in combination.

Task 2 (box C)

12 observers had 2-dimensional arrangement for defining the space, while 5 observers had 1-dimensional order. 14 observers used translucency as one of the criteria. The dimensions increased in 8 cases after getting access to additional objects. However, 13 observers mentioned that either they would not have changed their space in case they had access to all objects at once, or they were uncertain what they would have done. As suggested by Thomas et al. [13], people usually tend to stick to the standards they create and feel comfortable with.

Task 3 (box X)

The task was reasonably fast taking about 5 minutes on average. Seven observers had binary ranking - grouping the objects into two: "glossy" and "matte" categories. While others had more than two steps with some ties possible. There is very clear separation between the objects, as A, B and C are always considered less glossy than D and E. On the other hand, there is no consensus among observers about ranking within "glossy" and "matte" groups, especially, between D and E. All differences are statistically significant except for that between A and B, and D and E. 5 people considered D more glossy, 5 people ranked E as more glossy, while 7 people tied them. The analysis of their argumentation revealed two different approaches: people opting for D mostly argued that as the object is lighter and more translucent, more light is coming from it and therefore, it appears more glossy. On the other hand, people opting for E argued that it has larger tonal range, as the contrast between brightest and darkest points is larger, and therefore, the object appears more glossy. In the latter case, we can think that people use the contrast gloss (as defined by Hunter [23]) as an additional cue.

Task 4 (box M)

All observers ranked object C as the most glossy one, while the difference between A and B is not statistically significant. Objects A and B have the same level of surface coarseness, while they substantially differ in transparency. In this particular case, we achieved the same gloss perception with the same surface coarseness, even when other material properties are different. We can hypothesize that similar gloss appearance can be achieved with similar surface coarseness. This is in agreement with microfacet BRDF model [24, 25, 26]. However, the limits of this hypothesis need to be understood. As we have demonstrated for Task 3, transparency and lightness can impact gloss perception among some individuals, even when surface properties are the same.

Task 5 (box P)

All five objects have the same surface coarseness. According to the hypothesis drawn from the Task 4, their perceived glossiness is expected to be the same. It is interesting that there is no clear trend in ranking and no statstically significant difference among perceived gloss of the object. The only statistically significant difference was observed between D and E. Five observers decided that all objects have same amount of glossiness. In spite of this, other observers forced themselves to use various cues for ranking. While some used the same argumentation, as in case of the objects from box X (ligher and more translucent ones being more glossy, i.e. objects A and B), others used the clarity of their own image reflected on the surface, listing C, D, and E

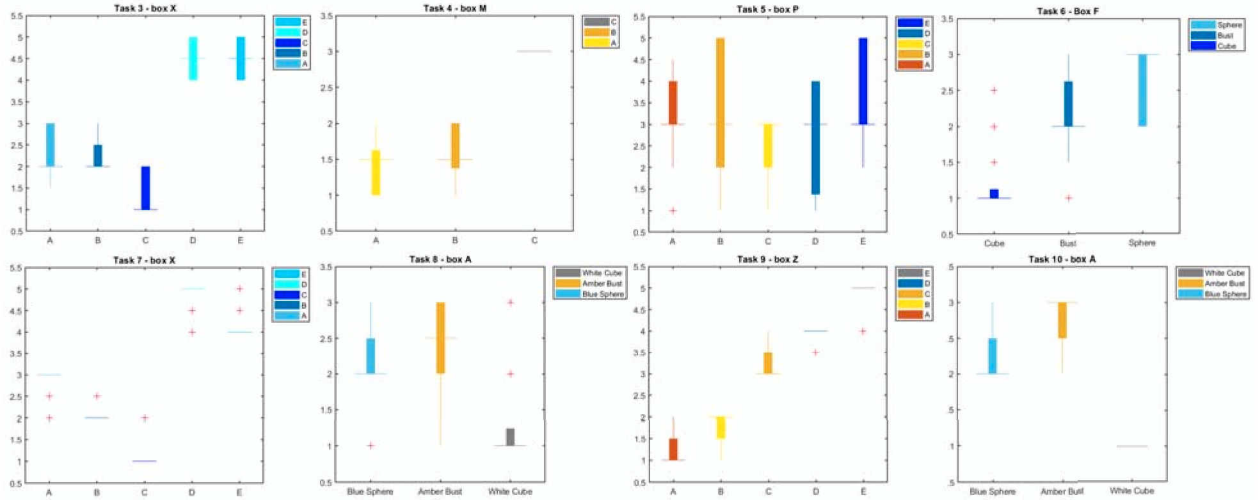
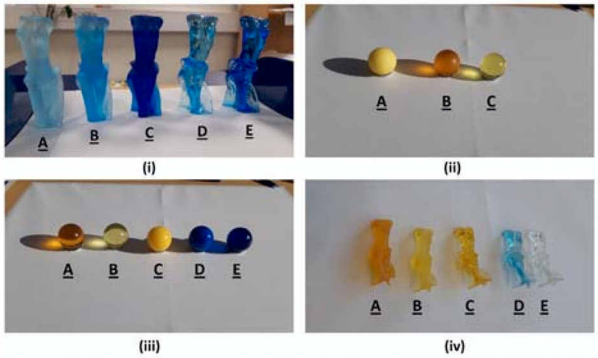


Figure 3: Boxplots for observer scores. Central mark -median; bottom and top edges - 25^{th} and 75^{th} percentiles, respectively; Whiskers extend to the extreme data points excluding outliers; red '+' symbol - outliers. We can observe clear separation for Tasks 3 and 7, clear order can be seen for Task 9, while no difference is significant for Task 5.



(i) (ii)

(iii) (iv)

Figure 4: The objects used: **(i)** Tasks 3 and 7 (box X). **(ii)** Task 4 (box M). **(iii)** Task 5 (box P). **(iv)** Task 9 (box Z).

as more glossy ones. This implied that they come a bit closer to the object and then, the intrinsic properties of the material permitted them to infer differences. Hunter [23] defined six types of perceptual gloss. Apparently, specular gloss that is "the most commonly measured parameter in experiments as an approximation for the physical measurement of perceptual gloss" [27] is widely used by the observers. On the other hand, we might argue that distinctness-of-reflected-image gloss is a secondary cue for judgement used by some observers. However, we think that the observers use different reflections from different light sources rather than different types of gloss as a cue. When the reflection of a very intense point light source is equivalent (the sun in our experiments), the observers might have tried to estimate ambient structured light in the room, which was too low to generate a very bright specularity. Therefore, the observers tried to evaluate distinctness of the reflected image. However, considering the data we have at hand, no statistical correlation have been found between average intensity of illumination (mean of the illumination in Lux at the beginning and the end of the experiment) and usage of distinctness-of-reflected-image as a cue. However, illumination has changed rapidly for some experiments due to meteorological conditions and thus, we need more controlled conditions to examine the hypothesis.

Task 6 (box F)

The decisions were very consistent about the rectangular object, 13 observers considering it least translucent and thereby, making the difference statistically significant. However, the difference between the bust and the sphere is statistically negligible. The rectangular object has more coarse surface than other objects, while surface coarseness is the same for the bust and the sphere. On the other hand, the sphere has one-level-less amount of blue dyes than the rectangular object and the female bust. Ranking the cube as least translucent can be accounted for the combination of its compact shape in comparison with the bust, higher amount of dyes in comparison with the sphere, and higher surface coarseness in comparison with both objects. Despite the fact that the bust has higher proportion of the dyes, we still have insignificant difference in perceived translucency with the sphere. This can be explained with the presence of thin areas in the bust, while the sphere is a compact and thick object. Objects of the same shape with varying material properties are often used in appearance studies (e.g. [8, 14, 28]). However, our data has some indications that shape might compensate for the difference in intrinsic material properties and generate the similar translucency perception of the overall object even if the material is less translucent. To test this hypothesis, further experiments are needed using different levels of dye mixture, and same level of surface coarseness.

Task 7 (box X)

In contrast with the first occurrence of this box, the results are very consistent among observers. All differences are statistically significant. 14 people ranked them in the following order from least translucent to the most translucent one: C (least translucent), B, A, E, D (most translucent). In this case, dye mixture and surface coarseness factors do not contradict and compensate each other that makes ranking simple for the observers.

Task 8 (box A)

13 observers considered the cube as least translucent one. Although there have been three observers who ranked this object first. The reason can be the experimental protocol, as the phrase "how light is going through" used by the experimenter was interpreted differently. While some participants understood this phrase as the complexity or simplicity of the light interaction with the objects, others judged simply the amount of light transmitted through them. The ambiguity of the instruction makes sphere and cube the only pair that are significantly different.

Task 9 (box Z)

The observers demonstrated very high consistency when ranking the objects by translucency. While shapes are the same in contrast with Task F, surface coarseness and dye mixing should be impacting perceived translucency. All differences are statistically significant except for that between A and B. Object A and B have a more rough surface than objects C, D, and E. Their surface scatters the light, and blurs the content behind. Besides, Object A has higher portion of yellow dyes, and therefore, considered mostly less translucent than B. However, four observers discarded the "color difference" and ranked them as equally translucent. While object E has smooth surface and no colorants inside, it is intuitive that the object is considered most translucent. There is more neutral transparent material in bluish object D than that in yellowish object C. However, as absorption and scattering properties of the two colorants are different, the effect of their concentrations are not directly comparable. The fact that bluish object is considered more translucent can be accounted for more complex cognitive factors too. Most observers described bluish object as precious and glassy, i.e. something associated with transparent material. On the other hand, yellow one was compared with jelly, less precious plastic, or amber - something to be less prone to transparent. The most interesting case is ranking object C over B, despite having higher concentration of the colorants. We can hypothesize that translucency perception is impacted by surface roughness and lightness of the object. What are the limits of the impact by each factor needs further investigation of the objects with varying surface roughness and dye concentration.

Task 10 (box A)

All observers considered the cube least glossy. However, the difference between the sphere and the bust is not statistically significant. This is an interesting case where objects with similar surface coarseness, but with different shapes and color intensity evoke similar gloss perception.

Task 11 (box T)

There has been interesting inconsistency in what observers consider the limit of being opaque or translucent, as particular objects were sometimes classified as opaque, and sometimes as non-opaque. Even when people observed a certain translucency for some of the opaque spheres, they still classified them opaque. We suggest that opacity does not imply the absence of translucency. However, this topic requires further investigation.

Discussion

After analyzing the data, we can say that expert observers are more scrupulous with taking decisions, judging objects from many different observation geometries, moving objects, trying to look through them and moving head to detect specularities, while non-expert observers decide faster. The interesting trends have been identified in the vocabulary usage, as experts tend more to use common appearance attributes "color", "gloss", "translucency" and "texture". Parallels with familiar objects using words like "icicle","gelatine","amber","milky","honey" etc. have been widely used. This phenomenon has been also observed in the paper by Thomas *et al.* [13] Nonetheless, the full analysis of behavioral patterns and vocabulary statistics will be conducted in further work. On average, each experiment took 1 hour and 7 minutes. Non-expert observers were 16 minutes faster spending 54 minutes on average, while the experiment took 70 minutes for the experts. However, small number of non-expert observers makes difficult to generalize the finding.

The quantitative data has shown that in some cases people are very consistent in what they consider glossy or translucent. Decision making is very easy and the objects are clearly separated. Although in other cases opinions vary a lot and the observers made diametrically different decisions. While poor experimental protocol could impact the result in some cases, there is clear indication that for this dataset cues used by different people vary and that the surface coarseness, dye concentration, and shape of the object play significant role. Furthermore, complex cognitive factors could also contribute to the final outcome.

The major questions can be drawn from above mentioned analyses: whether the trends observed for this dataset can be generalized to other objects and materials, and what are the extent surface coarseness, shape, and dye composition can impact and alter gloss and translucency perception? Considering the dataset, the interview, and the conditions, it is not possible to derive a general model of perception from these data. However, we still could identify some interesting trends to define research hypotheses for our future experiments.

Conclusion and Further Work

We have conducted a set of experiments investigating appearance assessment using real objects in uncontrolled conditions. Quantitative results show interesting cross-individual differences and similarities. We suggest that surface coarseness, material composition, and shape impact gloss and translucency perception.

It is worth mentioning that different tasks generated contradictory research hypotheses. For instance, considering tasks 4, and 10, we demonstrated that similar gloss perception is achieved, when the surface coarseness is nearly identical. On the other hand, task 3 has shown that transparency and lightness also impact gloss perception. Another hypothesis is that shape is significant factor for translucency perception and in some cases, can even outweight the impact from intrinsic material properties. Considering the results of the task 9, we suggest that when the shapes are identical, surface coarseness and dye mixture have most significant impact on translucency perception. The results of task 11 lead us to the hypothesis that opacity does not imply absence of translucency. We plan follow-up experiments to investigate those topics.

Finally, we also plan to conduct a comprehensive study of behavioral patterns and vocabulary better to understand the processes that lead us to given quantitative results. As we are limited to resin objects in this experiment, other materials and computer graphics could be used to generalize the findings.

References

[1] Yasir Nawab, Syed Talha Ali Hamdani, and Khubab Shaker, *Structural Textile Design: Interlacing and Interlooping*, CRC Press, 2017.

[2] Christian Eugène, "Measurement of "total visual appearance": a CIE challenge of soft metrology," in *12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement*, 2008, pp. 61–65.

[3] Thorsten Hansen, Maria Olkkonen, Sebastian Walter, and Karl R Gegenfurtner, "Memory modulates color appearance," *Nature Neuroscience*, vol. 9, no. 11, pp. 1367, 2006.

[4] M Pointer, "A framework for the measurement of visual appearance," *CIE Publication*, pp. 175–2006, 2006.

[5] Yun-Xian Ho, Michael S Landy, and Laurence T Maloney, "Conjoint measurement of gloss and surface texture," *Psychological Science*, vol. 19, no. 2, pp. 196–204, 2008.

[6] Shin'ya Nishida, Isamu Motoyoshi, Lisa Nakano, Yuanzhen Li, Lavanya Sharan, and Edward Adelson, "Do colored highlights look like highlights?," *Journal of Vision*, vol. 8, no. 6, pp. 339, 2008.

[7] Edul N Dalal and Kristen M Natale-Hoffman, "The effect of gloss on color," *Color Research & Application*, vol. 24, no. 5, pp. 369–376, 1999.

[8] Roland W Fleming and Heinrich H Bülthoff, "Low-level image cues in the perception of translucent materials," *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 3, pp. 346–382, 2005.

[9] Bei Xiao, Bruce Walter, Ioannis Gkioulekas, Todd Zickler, Edward Adelson, and Kavita Bala, "Looking against the light: How perception of translucency depends on lighting direction," *Journal of Vision*, vol. 14, no. 3, pp. 17–17, 2014.

[10] Lavanya Sharan, Ruth Rosenholtz, and Edward H Adelson, "Accuracy and speed of material categorization in real-world images," *Journal of Vision*, vol. 14, no. 9, pp. 12–12, 2014.

[11] Roland W Fleming, Christiane Wiebel, and Karl Gegenfurtner, "Perceptual qualities and material classes," *Journal of Vision*, vol. 13, no. 8, pp. 9–9, 2013.

[12] Lavanya Sharan, Ce Liu, Ruth Rosenholtz, and Edward H Adelson, "Recognizing materials using perceptually inspired features," *International Journal of Computer Vision*, vol. 103, no. 3, pp. 348–371, 2013.

[13] Jean-Baptiste Thomas, Aurore Deniel, and Jon Y Hardeberg, "The plastique collection: A set of resin objects for material appearance research," *XIV Conferenza del Colore, Florence, Italy*, p. 12 pages, 2018.

[14] Bui Minh Vu, Philipp Urban, Tejas Madan Tanksale, and Shigeki Nakauchi, "Visual perception of 3d printed translucent objects," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2016, vol. 2016, pp. 94–99.

[15] Kevin Smet, Wouter R Ryckaert, Michael R Pointer, Geert Deconinck, and Peter Hanselaer, "Colour appearance rating of familiar real objects," *Color Research & Application*, vol. 36, no. 3, pp. 192–200, 2011.

[16] Yannick Boucher, Helene Cosnefroy, Alain Denis Petit, Gerard Serrot, and Xavier Briottet, "Comparison of measured and modeled BRDF of natural targets," in *Targets and Backgrounds: Characterization and Representation V*. International Society for Optics and Photonics, 1999, vol. 3699, pp. 16–27.

[17] Wojciech Matusik, Hanspeter Pfister, Matthew Brand, and Leonard McMillan, "Efficient isotropic BRDF measurement," 2003.

[18] Frank P Nanna and John Jereb, "Gloss measurement system," Sept. 3 1996, US Patent 5,552,890.

[19] Yoshitaka Kuwada, "Gloss measurement apparatus and gloss measurement method," Mar. 16 2010, US Patent 7,679,747.

[20] Raymond G McGuire, "Reporting of objective color measurements," *HortScience*, vol. 27, no. 12, pp. 1254–1255, 1992.

[21] David J Gozalo-Diaz, Delwin T Lindsey, William M Johnston, and Alvin G Wee, "Measurement of color for craniofacial structures using a 45/0-degree optical configuration," *Journal of Prosthetic Dentistry*, vol. 97, no. 1, pp. 45–53, 2007.

[22] C Bonferroni, "Teoria statistica delle classi e calcolo delle probabilita," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, vol. 8, pp. 3–62, 1936.

[23] Richard S Hunter, "Methods of determining gloss," *NBS Research paper RP*, vol. 958, 1937.

[24] Robert L Cook and Kenneth E. Torrance, "A reflectance model for computer graphics," *ACM Transactions on Graphics (TOG)*, vol. 1, no. 1, pp. 7–24, 1982.

[25] Addy Ngan, Frédo Durand, and Wojciech Matusik, "Experimental analysis of brdf models.," *Rendering Techniques*, vol. 2005, no. 16th, pp. 2, 2005.

[26] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance, "Microfacet models for refraction through rough surfaces," in *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, 2007, pp. 195–206.

[27] AC Chadwick and RW Kentridge, "The perception of gloss: a review," *Vision research*, vol. 109, pp. 221–235, 2015.

[28] Philipp Urban, Tejas Madan Tanksale, Alan Brunton, Bui Minh Vu, and Shigeki Nakauchi, "Redefining a in rgba: Towards a standard for graphical 3d printing," *arXiv preprint arXiv:1710.00546*, 2017.