Relative contributions of low- and high-luminance components to material perception

Takehiro Nagai
Department of Informatics, Yamagata University, Yonezawa, Yamagata, Japan

Yuta Hosaka
Department of Informatics, Yamagata University, Yonezawa, Yamagata, Japan

Tomoharu Sato
Department of Intelligent Systems Engineering, Ichinoseki National College of Technology, Ichinoseki, Iwate, Japan

Ichiro Kuriki
Research Institute of Electrical Communication, Tohoku University, Sendai, Miyagi, Japan

Besides specular highlights, image pixels that represent clues to recognizing the object material, such as shading between threads of fabrics, often yield relatively lower luminance in the image. Here, we psychophysically examined how lower and higher luminance components contribute to material perception. We created two types of luminance-modulated images—low- and high-luminance-preserved (LLP and HLP) images—and instructed observers to choose which modified image resulted in a material impression closer to the original. LLP images were created by compressing the luminance contrast of the higher half of the histogram in each original photograph and vice versa. The stimuli were photographs of various samples of stone, wood, leather, and fabric. Although the LLP and HLP images were equally chosen, the choice ratios of the HLP images largely differed across the samples and categories and moderately correlated with the luminance statistics of higher-spatial-frequency sub-bands. These results suggest that either the lower- or higher-luminance components play an important role in material perception, depending on the material category. However, the correlation with sub-band image statistics for stone/wood samples was much weaker than for leather/fabric samples, suggesting that more intricate image characteristics may be involved in evaluating the material impressions of the stone/wood samples.

Introduction

Recently, the influence of image features on the perception of various aspects of material properties has been intensively studied. In psychophysical studies, relatively simple image statistics have been shown to contribute to material perception. The majority of these studies involved the investigation of the relationship between image features and, in particular, glossiness perception. A frequently cited study by Motoyoshi, Nishida, Sharan, and Adelson (2007) suggests that the statistics of the pixel-based and sub-band luminance histograms are strongly related to the perception of glossiness. The researchers claim that object images with positive skewness in their luminance histograms yielded strong perceptions of glossiness. The relatively smaller area, in total, of a specular highlight in an image emerges as minorities in the higher range of a luminance histogram. This turns out to be the long tail part of the histogram and yields a positive skewness. However, this claim is apparently still debatable because the effectiveness of skewness adaptation on perceived glossiness proposed by Motoyoshi et al. has been called into question by subsequent experiments with similar adaptation paradigms (Kim & Anderson, 2010; Kim, Tan, & Chowdhury, 2016). Rather, it has been suggested that other image features may explain perceived glossiness. For instance, the spatial alignment of specular and diffuse reflectance components is reported to be crucial to glossiness perception (Anderson & Kim, 2009; Kim, Marlow, & Anderson, 2011). It has also been reported that some image characteristics of specular highlights, such as their coverage and contrast, better explain the strength of perceptual glossiness, at least for a limited stimulus set (Marlow, Kim, & Anderson, 2012). More recently, information...
about the three-dimensional shapes of objects extracted from object contour and binocular disparity strongly affects glossiness and transparency through a combination with surface luminance variations (Marlow & Anderson, 2015, 2016; Marlow, Kim, & Anderson, 2017).

The glossy reflection of an object surface is attributed to the higher luminance range of the object image. However, it is also true that the darker region of an image sometimes provides the most important clues to the material perception of its body. Generally speaking, not all object surfaces produce specular reflections, e.g., unfinished woods, textile, rock, and leather. Even if the surfaces of these objects are finished or polished with gloss, we correctly perceive the material property of these objects from spatial patterns that are unique to each material class underneath the glossy top layer. The perceptual properties across different surface features, such as apparent roughness/smoothness, heaviness/lightness, hardness/softness, and warmness/coldness, have been compared in a few studies (e.g., Fleming, Wiebel, & Gegenfurtner, 2013; Nagai et al., 2015; Sharan, Rosenholtz, & Adelson, 2014). However, these investigations did not focus on the details of the image features that contribute to material perception.

Sometimes bumpiness or roughness on a very small scale of the surfaces of an object can be detected by local shading from their fine-scale surface structures. High- and low-luminance regions are formed on an object’s surface primarily due to the relationship between the normal directions of small facets and illuminations to yield shadings and specular highlights. For a close-up image of a woven textile surface, such as Figure 1, brighter regions in the image emerge on the front thread, and darker regions emerge mainly on the back thread and the spaces between these threads. This is due to the diffusion and absorption of light via mutual and multiple reflections before an incident ray reemerges from the surface of the fabric. Such darker shades of woven textile on the back thread or darker spaces in between are doubtlessly important in the process of fabric-material impression (Kuriki, 2015). In this case, small-scale changes in the shaded and lighted areas provide the most important clues to the “material” perception of the object body. For such fabric surfaces, the lack of shades between the thread that reside in the darker part of the luminance histogram must impair the fabric perception more severely than the highlights. A similar level of importance of the darker regions should exist in lumber or wood pieces because dark lines of wood grains, i.e., dissections of year rings, are very important clues that aid in the recognition that an object is made of wood. Even for other types of surface textures, such as stone pigmentation, there could be some asymmetries in the roles of darker and brighter sides of the texture luminance histogram on material impression, considering the separation of on and off signals with respect to the background luminance level in an achromatic channel (Beer & MacLeod, 2000) even though the importance of the darker side for such textures is unclear.

Despite the possible importance of lower-luminance components of object images, to our knowledge, there are no published results that directly address a comparison of the impact of lower- and higher-luminance components on material perception aside from glossiness perception (Kim, Marlow, & Anderson, 2012; Wiebel, Toscani, & Gegenfurtner, 2015). As a strategy to achieve quality material perception for the correct visual interpretation of object properties, human observers can add weights to either the darker or brighter regions. For fabric textures, the influence of removing the lower- or higher-luminance components on material perception may be asymmetric according to the weights in the visual system’s strategy. In addition, it can be considered that the human visual system uses such a fine-scale structure of shading patterns to correctly perceive material properties because the initial stage of cortical visual processing begins with the decomposition of the retinal image into small line segments (e.g., Hubel & Wiesel, 1959, 1968). If this is the case, the importance of a higher- or lower-luminance component can differ for different spatial frequency components; e.g., the effects of removing low- or high-luminance components in the high-spatial-frequency sub-bands on material impression of woven textile and wood surfaces may be more profound than those in low-spatial-frequency sub-bands.

On the other hand, it is also possible that the human visual system does not use such fine structure information and achieves material perception from macro-scale statistics. The technology of computer graphics has advanced dramatically in the past several decades, but it is still difficult to render a complete physical
structure of a textile with all individual threads. For example, techniques that create the impression of a silk-like appearance at a glance by skipping the precise simulations of every optical effect around the threads have been intensively studied (e.g., Zhao, Hasan, Ramamoorthi, & Bala, 2013). This approach works effectively in some cases. Most of the available techniques control the macrostructure of the reflectance (luminance) pattern from small areas of the fabric surface to render different kinds of textile images. Therefore, it is still possible that luminance statistics of an image on a relatively macro scale (i.e., a scale such that the rendition of luminance differences between the threads can be neglected) can provide sufficient information about nonglossy objects. Therefore, whether a particular luminance portion or a relatively macro-scale image statistic is significant for material perception is still unresolved.

In the present study, we aim to clarify if the potential influence of global luminance statistics of an image or differences in the effectiveness of the high- and low-luminance components is important for the material perception of an object body. To achieve this objective, we conducted a psychophysical experiment in which we presented object photographs with selectively lowered values for the luminance contrasts of lower- or higher-luminance components.Observers were asked to choose one of the modified images that appeared most similar to the original image in terms of the material impression. If a higher- or lower-luminance component is more important, the removal of these components is expected to significantly impair material impression. In addition, in order to investigate the presence of asymmetry in the dependence on macro- or micro-scale image statistics, we applied a spatial frequency sub-band analysis to examine which component of the image statistics account for the observers’ choices of higher- or lower-luminance preserved images.

## Methods

### Observers

Eleven males and one female participated in the experiment as observers. They had normal or corrected-to-normal visual acuity. None of the observers were aware of the purpose of the experiment. The experimental protocols were approved by the ethical committee of the faculty of engineering, Yamagata University, and followed the code of ethics of the World Medical Association (Declaration of Helsinki). All observers provided written informed consent.

### Apparatus

The stimulus was presented on an organic light emitting display (OLED) PVM-1741 (Sony, Tokyo, Japan) with a 60-Hz refresh rate and a 1,920 × 1,080 resolution. The gamma curves and wavelengths of the red, green, and blue primaries were measured and carefully calibrated with an SR-3AR spectral photometer (Topcon, Tokyo, Japan). A Vostro 3900 personal computer (Dell, TX) was connected to the monitor. GNU Octave and Psychtoolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) on Ubuntu 14.04 LTS were used to control the procedure of the experiment and recorded the observers’ responses. The observers binocularly viewed the monitor at a distance of 50 cm away in an otherwise dark room with their heads positioned on a chin rest. We employed the OLED monitor, which can present stimuli of less than 0.1 cd/m² to precisely represent luminance information of dark regions on the surfaces of an object.

### Stimulus

#### Object image photography

We used photographs of object samples in the Shitsukan (literally, material) sample set (Takei Scientific Instruments, Niigata, Japan), which consists of 302 real object samples made of various materials that form shapes of the same size (approximately 10 × 10 cm) and curvature; the shape was either flat or a wavy macro profile. This sample set was developed to investigate the mechanisms of the human sensory system (including vision) in regard to the perception of object material and surface properties. The samples were imaged under a simulated solar lighting system (SERIC, Tokyo, Japan) with a D700 commercial digital camera (Nikon, Tokyo, Japan). The details of the photography process are described in Nagai et al. (2015). Although the size of the acquired images was 4,256 × 2,853 pixels, the 600 × 600 pixel regions in the center of each object were trimmed so that the observers would use only the texture information on the surface and not the object edge information for material perception. The trimmed image covered an area of 12.8° × 12.8° on the screen.

#### Images

We used 71 samples as stimuli from the photographed images. Some examples are shown in Figure 2a. All selected samples are shown in Figures A1 and A2 in Appendix A. We employed the following two criteria for the selection of the images:

1. We chose only flat samples to prevent judgments based on glossiness and/or shading from macro-
scale curvatures (Ho, Landy, & Maloney, 2008) because our objective is to study perceptions of material features of the object’s body. In particular, glossiness occupies higher-luminance components in the histogram, such as specular highlights (Marlow et al., 2012), and it could compress the contrast of the other (i.e., lower) luminance range.

2. The utilization of samples from a variety of material categories as stimuli is clearly ideal. However, only stone, wood, leather, and fabric samples were employed because the flat samples of the other categories (plastic, metal, and glass) exhibited very poor impressions of material qualities other than glossiness due to the lack of surface bumpiness or texture as a clue to identify the material category.

A total of 11 flat stone samples were used together with 20 samples each for the other categories. The mean luminance of these selected samples was 17.0 cd/m², and the chromaticity of all images was set as that of a D₆₅ illuminant: \((x, y) = (0.313, 0.329)\).

Next, the luminance distributions of the images were manipulated to control the relative contrast of the higher- and lower-luminance components. We created a reduced contrast image for each sample in which the luminance contrast was reduced by a factor of 0.2 from the original image while keeping the mean luminance unchanged. In the next step, about half of the pixels with luminance values that were either lower or higher than the mean were selected and combined with the other half of the original pixels to generate a lower-luminance-preserved (LLP) and a higher-luminance-preserved (HLP) image. For instance, the LLP image of a sample was derived by combining pixels with lower luminance values than the mean from the original image and pixels with higher luminance values than the mean from the reduced-contrast image. This yielded an image with half of the histogram compressed by a factor of 0.2 toward the mean of the original image. Therefore, the LLP images contained the original lower-luminance components, and the higher-luminance components were excluded due to the compression. The HLP image was also created in a similar way: the higher- and lower-luminance components of the original and reduced-contrast images, respectively, were combined.

Figure 2b shows an example set of our stimuli. From the left are the HLP, original, and LLP images. Each stimulus set consisted of three images of an identical sample, simultaneously presented on a D₆₅ gray background with 40 cd/m² of luminance. The original image was always presented at the center, and the LLP and HLP images were on the left or right. The positions of the LLP and HLP images were randomly determined by trial by trial. The images were separated by a gap.

Figure 2. (a) Samples in four material categories used in the experiment. (b) Four examples of stimulus images. In these examples, the left and right images are HLP and LLP images, respectively.
width of 60 pixels (1.3°) and colored with the gray background.

Procedure

A stimulus set was presented on the monitor during each trial (Figure 2b). During the stimulus presentation, the observers identified which of the two (left or right image) represented a “closer material impression” to the original image at the center by clicking a mouse button in the two-alternative, forced-choice manner. The stimulus was presented on the screen until the observer made a response, followed by a uniform gray screen to avoid dark adaptation. The response time of the observers was approximately 2.5 s, on average, for each image. Trials for the 71 stimuli were repeated twice in each session. The first 60 s of each session was a practice interval for the observers to adapt to the stimulus luminance levels and to gain familiarity with the task. During this interval, the observers conducted the same task as in the formal trials for randomly chosen stimuli, but these responses were not subsequently used in the analysis. The observers were not provided with any explicit instructions regarding any material features they should consider. This was intended to avoid possible differences in criteria for evaluating the quality of material perception across images and material categories. For example, if roughness were explicitly identified as an evaluation metric, images of polished stone surfaces may not be appropriately evaluated, or this could lead to the observers making inappropriate guesses, especially for the reduced-contrast images. Thus, the observers were left to rely on their own subjective criteria and strategies for making material judgments and not on any imposed criteria designated by the experimenter.

Results and discussion

Choice ratio

Figure 3 shows the choice ratios calculated for all the responses of the observers. Each symbol represents the ratio for each sample image, and the bars show the average across all samples. In this figure, the higher ordinate values indicate that the HLP images resulted in holding closer material impressions with the original images than the LLP images. Two clear tendencies are observed in this figure. First, the choice ratios for different samples varied from high to low values in each category, suggesting that the relative importance of lower- and higher-luminance components differed across the samples. Second, as an overall result, the choice ratios were biased toward neither lower nor higher values; they uniformly ranged from zero to one except for the wood and leather samples, and the overall average choice ratio was 0.495. These results indicate that neither lower- nor higher-luminance components can be implicated as having a stronger impact on material perception in a general sense. Both the lower- and higher-luminance components seemed to contribute to the perceptions of different material features apart from glossiness, but this depended more on the property of each sample image. This point is discussed in the next section.

The choice ratios were significantly biased in some material categories. Specifically, the choice ratios for the wood category samples appear to be biased toward the LLP images, and the leather category samples exhibited a bias toward the HLP images. The average choice ratios were different with a statistical significance in most category pairs except for the stone–leather pair when assessed using a permutation test with Holm’s multiple comparisons ($p < 0.001$ for stone–wood, stone–leather, wood–leather, and wood–fabric pairs and $p < 0.05$ for a leather–fabric pair). In addition, as mentioned previously, the choice ratios differed significantly between samples within each material category. These results suggest the possibility that the specific region of the luminance histogram of each object image that strongly contributes to material appearance may depend more on certain inherent properties of each sample and not just on the material category.
Correlation to simple image statistics

As described in the Introduction, the image luminance statistics could have affected the choices that were made by the observers during the experiment. In this section, we try to elucidate which image statistics can account for the observers’ choices by comparing the HLP choice ratio and lower-order image statistics (up to the third order: mean, variance, and skewness) of either overall or sub-band images. For instance, the luminance histograms of the original images may strongly affect changes in the histogram shape when generating the LLP and HLP images. If the histogram of an original image had a positive skew, the contrast compression of the high-luminance components would affect the shape of the histogram more strongly than those for low-luminance components in terms of, e.g., luminance contrast. Thus, the image statistics of the original images must strongly affect the choice ratios if the observers’ judgments were based on such image statistics.

To examine this assertion, we examined the relationship between the observers’ choice and simple luminance statistics of the original image. We first calculated three kinds of pixel-based simple luminance statistics: mean, root-mean-square (RMS) contrast, and skewness derived from all pixels in each of the original images. The RMS contrast was the standard deviation of the luminance normalized to the mean luminance of each sample image. Furthermore, we calculated the RMS contrast and skewness of the sub-band images. The sub-band images were obtained by applying Gaussian band-pass filters to the original images; the Gaussian filters had central frequencies of two, four, eight, 16, and 32 cycles per image (cpi), and the bandwidth was one octave. The RMS contrasts in the sub-band images were also normalized to the mean luminance of each sub-band image.

It must be taken into account that the distributions of the second- and third-order statistics across all image samples do not always exhibit a normal (Gaussian) distribution. The histograms of the RMS contrast and skewness across the 71 image samples exhibit a strongly positive skew in both the pixel and sub-band statistics. These skewed histograms of statistical values are likely to cause bias in the analysis using Pierson’s correlation coefficients between the image statistics and observers’ choice ratios. To compensate for such deviations from the normal distribution, we applied a nonlinear transformation to the statistical values before applying correlation analyses. Specifically, we calculated a power to a certain constant for each of these statistics so that the skewness in the distribution was zero. For the image skewness, the minimum value among the samples was subtracted from the original skewness before applying this nonlinear transformation to avoid errors due to the use of negative values. The average exponents among the sub-bands were 0.026 and 0.19 for RMS contrast and skewness, respectively. After applying these transformations, we obtained three-pixel statistics (mean, contrast, and skewness) and 10 sub-band statistics for each image. Finally, we calculated the correlation coefficients between the statistical values after applying the aforementioned transform and deriving z scores of the choice ratios using all samples across material categories. Thereafter, all correlation coefficients were calculated using the z scores of the observers’ choice ratios and not with the crude ratio values. This analysis is different from the kernel analysis because the nonlinear transformation was performed prior to the calculation of the correlation coefficients, and the exponents were not determined for the optimization of the correlation coefficients.

Figure 4 shows the correlation coefficients between these statistics and the choice ratios for all sample images. The rightmost isolated symbols show the coefficients with the pixel statistics derived from the original images. The mean luminance shows a negative correlation with the z-scored choice ratio. This correlation is statistically significant (the correlation coefficient is larger than zero; \( p < 0.001 \) by a nonparametric bootstrap test after a repetition of 10,000 times). Similar but positive correlations with statistical significance are observed for the other pixel statistics: the RMS luminance contrast and the luminance skewness (both \( ps < 0.001 \) by a parametric bootstrap test).

In addition, the correlations of the sub-band statistics with the observers’ choices suggest the effects...
of the spatial frequency on the task. The correlation coefficients increase with the sub-band spatial frequency although the RMS contrast in all frequency bands shows a positive correlation with the observers’ choice. Similar spatial frequency effects are observed in the sub-band skewness; the correlations are significantly different from zero only for skewness of 16- and 32-cpi sub-band images.

After a precise inspection of the results, the observers’ choice ratios were considered to rely on the luminance range of the long-tail part of the skewed histograms. The significant correlations of the pixel luminance statistics in the original images with the choice ratios (isolated symbols labeled “pixel” on the abscissa in Figure 4) raise the possibility that the relative importance of the lower- and higher-luminance components (Figure 3) was a result of the dependence of material judgments on the pixel statistics. For example, if the mean luminance of an original image was relatively low, higher-luminance components could be more salient than lower-luminance ones. Similarly, for images with a positive luminance skewness, higher-luminance pixels may contribute more to the RMS contrast and vice versa. Therefore, these “salient” pixels may have had a stronger impact on material perception and led to a higher correlation between the choice ratios and the statistics in Figure 4. In addition, both the contrast and skewness correlations are stronger in the higher-frequency sub-band images, suggesting that fine-scale luminance statistics may play an important role in material perceptions.

In summary, the correlations of these image statistics with the observers’ choices suggest that the luminance statistics can be a candidate factor that determines the observers’ choices. Namely, the relative impact of lower- and higher-luminance components was likely affected by the simple image statistics to some extent. In particular, it is suggested that the image statistics in spatially fine patterns (i.e., high spatial frequency) are more important for material impressions than those in coarse patterns. It, therefore, seems natural that these fine patterns provide crucial information regarding material impressions because these patterns should reflect different physical sources related to material differences: thread patterns of fabrics/textiles, small asperities of leather surfaces, mosaic patterns of stone samples, and so on.

Differences between material categories

It is also important to consider the reason for the systematic differences in the averaged choice ratios between the material categories as shown in Figure 3. As mentioned in the previous subsection, simple image statistics were significantly correlated with the choice ratios overall. Because images of different material categories tend to have different image statistics, these statistics could account for the differences in the behavior of the observers between categories. The relations between the observers’ choice ratio and the image statistics are summarized in Figure 5. The larger symbols represent the average values for each material category. The image statistics and the mean choice ratios seem to be correlated across the categories to some extent. Thus, it is plausible that the luminance statistics at least partly contributed to the material category differences in the choice ratios. Perhaps equalizing the intensity ranges across all samples would reduce the material category differences to some extent. However, we have not tested this as yet for possible side effects in other aspects: e.g., increasing the average luminance of a black velvet image would deteriorate its velvet-like appearance.

Nevertheless, it is also likely that the image statistics contribute to the evaluation of material impressions for each material category in a different way. For example, it is obvious that the wood and fabric samples yield completely different surface qualities. In addition, to avoid the observers’ judgment in favor of a particular material category, we did not explicitly instruct the observers on which surface qualities should be evaluated. This means that the surface properties used for the judgments could have been different across the categories. Accordingly, the image statistics used to evaluate the material similarity may have been different among the material categories.

To examine this issue, we calculated the correlations between the lower order (first- to third-order) image statistics and the choice ratios for each material category; the results are shown in Figure 6. The correlations of the pixel statistics (isolated symbols on the right) are stronger for the fabric and leather categories than the stone and wood categories in all (first- to third-order) statistics. Because the fabric and leather samples in the pixel statistics (Figure 5) span widely across the range of the horizontal axis, the relationships between the values of this axis and the choice ratios became clear: the higher the mean luminance difference, the lower the choice ratio of the HLP and vice versa. In addition, the correlations of the RMS contrast become higher as the spatial frequencies of sub-band images increase except for the stone samples. This trend is similar to that in Figure 4. These trends again seem to suggest that the higher-spatial-frequency components are more informative for the material evaluation at least in our stimuli. More importantly, these results imply that the dependence of material judgments on image statistics is different for the material categories. Generally, the evaluation of material property for the fabric and leather samples seems to depend strongly on simpler image statistics.
The negative correlations of sub-band skewness, especially for the wood samples (Figure 7c) at higher spatial frequencies, is counterintuitive and its interpretation is complex. This point is discussed in detail in the General discussion.

Which factors induced the apparent differences in response trends between the material categories? To address this question, we conducted a supplementary experiment (details are described in Appendix C). One factor could be the bias of simple statistics in the images as shown in Figure 5. If this is the case, replacing the samples in every category may alter the response bias between the material categories. However, higher-order image features should also contribute to material impression differently across the material categories, and this could have influenced the differences in the relative impacts of lower-/higher-luminance components. To confirm this point, we performed a supplementary experiment using phase-scrambled images as the stimuli. Specifically, the stimuli were derived by randomizing the phases of spatial frequency components of the original images while keeping the spatial frequency amplitude and luminance histogram almost unchanged. This phase-scramble procedure should destroy the higher-order image features and weaken material impressions. The observers’ task was the same as in the main experiment: the judgment of the “material” similarity. The results are shown in Figure C2 (Appendix C). The observers’ responses for fabric samples were almost the same as for the original experiment. This suggests a strong material impression from low-level image statistics that remained after the phase-scramble manipulation. In contrast, the diversity of the choice ratios became smaller than the original experiment for stone and wood samples, indicating the destruction of higher-order image statistics induced by the phase-scramble procedure seemed to make the judgment more difficult for the samples of these categories. In fact, the observers later reported that they mostly judged “image” (not material) similarity for those samples. These differences in the effects of phase randomization across the categories imply that differences in the complexity of image features, not only the bias in the simple image statistics within every category, may have induced the differences in the response trends across the categories.

Figure 5. Relationship between choice ratios and simple pixel statistics (mean, RMS contrast, and skewness) of original images in each category: (a) mean, (b) RMS contrast, and (c) skewness. Symbol colors indicate material categories. Large symbol indicates the average value among samples in each material category.
To summarize, the key factors that contributed most to the observers’ criteria (whether to use image statistics in luminance or their similarity) could have been different for each of the material categories. Both the pixel and sub-band luminance statistics exhibited a positive correlation with the observers’ choices that was stronger for the leather and fabric samples than for the stone and wood samples. These results suggest that the

Figure 6. Correlation coefficients between choice ratios and image statistics in (a) stone, (b) wood, (c) leather, and (d) fabric categories. Rightmost plots in each panel show the results for pixel statistics. The error bars show the standard errors estimated from a nonparametric bootstrap procedure.

Figure 7. (a) Original, (b) 8-cpi, and (c) 32-cpi images of part of wood sample. Contrast is emphasized for better visibility.
observers may have used simple image statistics for fabric and leather samples, and relatively complex image features were more significant for the stone and wood samples although they were not available in the first- to third-order statistics. Similarly, these differences between the material categories could have arisen from the dependence of material impression on higher-order image statistics.

General discussion

Relative importance of the higher- and lower-luminance components

The relative impacts of the lower- and higher-luminance components on material perception differed across the sample images. In addition, there was some systematic bias among the material categories in terms of HLP/LLP preference (Figure 3). According to an analysis of the relationship between the observers’ choice ratios and the lower-order image statistics (mean, contrast, and skewness), the results showed that some of the luminance statistics were moderately correlated with the observers’ choice ratios (Figure 5a). In particular, the skewness of the luminance histogram, either as pixel statistics or sub-band statistics, showed a higher correlation with the observers’ choice ratios. This implies that part of the image with a relatively smaller population and a higher contrast with respect to the background played a significant role. These regions included the shaded area between strings in the textile or local highlights that gave clues to small-scale bumpiness. This study was initiated based on a prediction that the lower-luminance (dark) spots or lines among the relatively brighter flat areas could appear to be more salient to the observers in images with a relatively high mean luminance so that this factor had a more pronounced influence on the observers’ judgments than the higher-luminance components (Figures 2 and 6). As a result, the choice ratios for the wood samples show a bias toward the LLP images, which supports our original prediction (Figure 3). The bias toward HLP images in the leather samples appears to conflict with our prediction. However, the choice was influenced mainly by a small highlight along hump edges of the leather surface structure among otherwise relatively dark and flat areas. Therefore, in terms of the dependence on the salient area, the result was consistent across both categories. In summary, one of the main findings of the present study is that the relatively small areas of high contrast region with respect to the background harbors important clues on material quality judgments. Ultimately, whether these clues are included in the higher-/lower-luminance component or not depends on the material category.

In the present study, the higher-/lower-luminance components were defined with respect to the mean luminance within the luminance range of each individual image; i.e., the image statistics were not defined based on the global luminance range across all images or the visibility range. On the other hand, several studies have dealt with the effectiveness of lower- and higher-luminance components with respect to the background that spanned the entire achromatic range of the visual field from black to white. It has been reported that the on and off signals with respect to the adapting field (gray) in the achromatic channel were mediated by separate mechanisms on unipolar contrast adaptation (Beer & MacLeod, 2000; Sato, Motoyoshi, & Sato, 2016) and spatial interactions in perceived contrast (Pamir & Boyaci, 2016; Sato, Motoyoshi, & Sato, 2012). In the present study, the average stimulus observation time was approximately 2.5 s, and thus, the light adaptation to the luminance levels of each image should be weak even if it occurred. Hence, the adaptation level of the achromatic channel was possibly maintained at approximately equal to the background luminance throughout the experiment. However, the extent of the luminance range was not controlled among samples, and the relative image luminance levels in the achromatic channel gamut could have also affected the results. Therefore, one of our future plans is to explore the role of the low- and high-luminance components within the visual system’s gamut by explicitly controlling the luminance levels of the images relative to the sensitivity range of the achromatic channel.

From a practical point of view, an understanding of the relative effectiveness of low- and high-luminance components on material impression may be efficacious, for instance, in the manipulation of tone reproduction for emphasizing or reducing material impressions in images. Specifically, the enhancement of contrast in lower- or higher-luminance components may enhance the material recognition in the case of wood or leather, respectively. However, in our experiment, material category recognition was not severely impaired for most samples in the HLP and LLP images. Therefore, these effects may not be strong enough to alter recognition in terms of material categories.

Interpretation of negative correlations in sub-band analysis

The interpretation of the correlations for sub-band skewness is complicated. The sub-band skewness for fabric samples (Figure 6d) exhibited consistently positive correlations similar to the overall trend in
Figure 4. As previously mentioned, a positive skewness in the original image yields a proximity of the luminance contrast between the original and HLP images. If the judgments were entirely based on the similarity of contrast between HLP/LLP and the original images, positive correlations would be predicted. This hypothesis holds, at least for our fabric samples. An additional analysis in which the correlation between the luminance contrast proximity and the observers’ choice ratio were directly measured also supports this hypothesis (the results are shown in Figure B1 in Appendix B). However, the stone and wood samples (Figure 4a and b) exhibited negative correlations with the sub-band skewness, especially for the higher-spatial-frequency (>8 cpi) components. The negative correlation, in this case, means that the higher the (sub-band image) skewness, the lower the HLP choice ratio (i.e., the higher the LLP choice ratio).

The image statistics for the stone samples exhibited relatively weaker correlations with the observers’ choices (Figure 6a), and the wood samples exhibited negative correlations at higher-spatial-frequency sub-bands (Figure 6b). The images in these categories had periodical patterns on their surfaces. However, the luminance patterns for the stone and wood could have been more irregular than the case of the leather and fabric samples; i.e., they had irregular fluctuations in the repetition of the spatial patterns that appeared on the cut surface even as relatively periodic patterns. These regularities or irregularities in the periodic luminance patterns may have limited the effectiveness in the lower-order (up to third-order) image statistics. In fact, an informal attempt to apply orientation-selective filters to, e.g., the wood images did not improve the results in terms of the correlation with the observers’ choice ratios (not shown). If we employed more complex image statistics, such as the joint statistics for texture synthesis by Portilla and Simoncelli (2000), this may have resulted in the possible capture of image features that the observers used during the decision process in the experiments. The supplementary experiment (Appendix C) partially confirmed that the complex image statistics affected material impressions. In this experiment, phase-scrambled images generated from the stimulus images in the original experiment were used to disorganize complex image statistics while maintaining the simple image statistics. The results indicate that the observers’ responses for stone and wood samples were significantly altered (Figure C2), which implies that complex image statistics play a significant role in material quality judgment, at least for these classes of materials.

The differences in the choice ratios among material categories may be caused by several factors. The primary factor could be the characteristics of the image statistics. The correlations were stronger for the RMS contrast and skewness in sub-band images at higher spatial frequencies (>8 cpi) in fabric and leather samples. It is possible to consider that a finer spatial structure, which is inherent in each material category, could have been used as a clue to evaluate these material qualities. Giesel and Zaidi (2013) showed that the perceptions of human observers of different material features, such as the roughness and thickness of fabric samples, can be characterized by specific scales of luminance variations. They showed that the manipulation of luminance components of specific spatial scales could largely alter material impressions of the fabric images. Their report suggests a link between simple image features and material feature perception of fabric images as a kind of heuristic. Many fabric and leather sample images, especially those used in this study, have periodic luminance patterns on the surface. This may be the reason why a strong correlation was observed for even simple image statistics with the material perceptions for the fabric and leather samples (Figure 6c and d).

The correlation between the luminance contrast proximity and the observers’ choice (Figure 6a) and the wood samples exhibited negative correlations at higher-spatial-frequency sub-bands (Figure 6b). The images in these categories had periodical patterns on their surfaces. However, the luminance patterns for the stone and wood could have been more irregular than the case of the leather and fabric samples; i.e., they had irregular fluctuations in the repetition of the spatial patterns that appeared on the cut surface even as relatively periodic patterns. These regularities or irregularities in the periodic luminance patterns may have limited the effectiveness in the lower-order (up to third-order) image statistics. In fact, an informal attempt to apply orientation-selective filters to, e.g., the wood images did not improve the results in terms of the correlation with the observers’ choice ratios (not shown). If we employed more complex image statistics, such as the joint statistics for texture synthesis by Portilla and Simoncelli (2000), this may have resulted in the possible capture of image features that the observers used during the decision process in the experiments. The supplementary experiment (Appendix C) partially confirmed that the complex image statistics affect material impressions. In this experiment, phase-scrambled images generated from the stimulus images in the original experiment were used to disorganize complex image statistics while maintaining the simple image statistics. The results indicate that the observers’ responses for stone and wood samples were significantly altered (Figure C2), which implies that complex image statistics play a significant role in material quality judgment, at least for these classes of materials.

Another major factor could be the difference in observers’ strategy. The observers could have employed different strategies or paid attention to different material characteristics.

The cause of the difference in the choice ratios among material categories

The differences in the choice ratios among material categories may be caused by several factors. The primary factor could be the characteristics of the image statistics. The correlations were stronger for the RMS contrast and skewness in sub-band images at higher spatial frequencies (>8 cpi) in fabric and leather samples. It is possible to consider that a finer spatial structure, which is inherent in each material category, could have been used as a clue to evaluate these material qualities. Giesel and Zaidi (2013) showed that the perceptions of human observers of different material features, such as the roughness and thickness of fabric samples, can be characterized by specific scales of luminance variations. They showed that the manipulation of luminance components of specific spatial scales could largely alter material impressions of the fabric images. Their report suggests a link between simple image features and material feature perception of fabric images as a kind of heuristic. Many fabric and leather sample images, especially those used in this study, have periodic luminance patterns on the surface. This may be the reason why a strong correlation was observed for even simple image statistics with the material perceptions for the fabric and leather samples (Figure 6c and d).

The correlation between the luminance contrast proximity and the observers’ choice (Figure 6a) and the wood samples exhibited negative correlations at higher-spatial-frequency sub-bands (Figure 6b). The images in these categories had periodical patterns on their surfaces. However, the luminance patterns for the stone and wood could have been more irregular than the case of the leather and fabric samples; i.e., they had irregular fluctuations in the repetition of the spatial patterns that appeared on the cut surface even as relatively periodic patterns. These regularities or irregularities in the periodic luminance patterns may have limited the effectiveness in the lower-order (up to third-order) image statistics. In fact, an informal attempt to apply orientation-selective filters to, e.g., the wood images did not improve the results in terms of the correlation with the observers’ choice ratios (not shown). If we employed more complex image statistics, such as the joint statistics for texture synthesis by Portilla and Simoncelli (2000), this may have resulted in the possible capture of image features that the observers used during the decision process in the experiments. The supplementary experiment (Appendix C) partially confirmed that the complex image statistics affect material impressions. In this experiment, phase-scrambled images generated from the stimulus images in the original experiment were used to disorganize complex image statistics while maintaining the simple image statistics. The results indicate that the observers’ responses for stone and wood samples were significantly altered (Figure C2), which implies that complex image statistics play a significant role in material quality judgment, at least for these classes of materials.

Another major factor could be the difference in observers’ strategy. The observers could have employed different strategies or paid attention to different material characteristics.
features in the experimental task when evaluating images in different material categories. To investigate the observers’ strategy in performing the task, we asked some of them about which image features they relied on in a questionnaire after the experiment. The majority of the respondents indicated that they relied on the similarity of global or local texture impression or overall impression of images at a glance and not the specific material features, such as roughness, bumpiness, or glossiness. This questionnaire suggests that they did not pay close attention to specific material features, at least not consciously. Therefore, our experiments do not elucidate the precise relationship between image features and the perception of specific material features. To clarify which kind of clue is used most for evaluating material qualities and which of the lower- or higher-luminance components contributed to the observers’ strategies to a greater extent, it would be necessary to perform further experiments specifically designed to focus on the strategy of evaluating every material feature, such as bumpiness, hardness, and wood-like/stone-like/fabric-like/leather-like appearance. Through these experiments, some differences in the relative importance of high- and low-luminance components could be elucidated for each material quality. This is similar to a previous study that reported on the importance of high-luminance regions in glossiness perception (Marlow et al., 2012), but we avoided this judgment in an attempt to derive a general rule on the impact of lower-/higher-luminance components on material perception.

Limitations of our findings

Our suggestions have several limitations, especially in relation to image statistics, due to the experimental stimuli and procedures. One of the first limitations is that only flat-type samples were used as the stimuli; i.e., we excluded curved samples to avoid the influence of glossiness on observers’ judgments. However, the use of flat-type stimuli only may miss some important relationships between luminance range and material perception. For example, the wavy samples have some overall luminance gradient from shading. Considering that changes in perceived lightness from the luminance gradient affect the apparent contrast of a grating (Pamir & Boyaci, 2016), the luminance gradient from global surface shape could affect the quality of material perception. These possible effects of macro-shading patterns could be investigated using samples with macro-scale shapes but without strong specular reflections.

In addition, we employed only grayscale images as stimuli in this study, and therefore, possible chromaticity effects on material impression were completely ignored. For instance, in a concave corner of an object material surface, an increase in chromatic saturation and a decrease in luminance occurred simultaneously due to mutual reflections of light. This kind of clue would lead to the estimation of the relationships between the illuminant and surface reflectance from correlations between luminance and saturation (Golz & MacLeod, 2002; Kuriki, 2015). Although similar relationships exist for object shape and reflected light intensities in the grayscale images, it is possible that this relationship between chromatic saturation and luminance may serve as a strong clue to improve the quality of surface property (material) perception.

Finally, cues for macroscopic shape in a three-dimensional space were insufficient in our stimuli. We used stimulus images without object contours, which are known to affect material and macro-shape impression by interacting with luminance patterns on the surfaces (Marlow & Anderson, 2015, 2016). Some of our material samples had bumpy surfaces, and thus, object contours should contain such shape information. Therefore, material impression might be affected by the addition of object contours if the observers’ responses were based on material impression. Such higher-order effects of luminance patterns on material impression may provide us with additional clues about the underlying processing of the relative importance of low-/high-luminance components.

Keywords: material perception, image statistics, psychophysics

Acknowledgments

This study was supported by JSPS KAKENHI 22135005 and 15K00372 to TN and 16H01658 to IK. We thank Kowa Koida for providing us with the images of the material samples. The authors are also grateful for comments from attendees of the Shitsukan and Color Vision meeting in Sendai, which was supported by the Research Institute of Electrical Communication, Tohoku University, for the Cooperative Research Project H28/B04 and H28/A04.

Commercial relationships: none.
Corresponding author: Takehiro Nagai.
Email: nagai.t.aa@m.titech.ac.jp.
Address: Department of Information and Communications Engineering, School of Engineering, Tokyo Institute of Technology, Tokyo, Japan.

References

Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness.
Appendix A

All stimulus sample images are shown in Figures A1 and A2.

(A) Stone

(B) Wood

Figure A1. All stimulus samples in (a) stone and (b) wood categories used in the experiment. These samples were used as original images.
Figure A2. All stimulus samples in (c) leather and (d) fabric categories used in the experiment. These samples were used as original images.
Appendix B

The HLP/LLP images differ from the original image not only in luminance histogram, but also in the spatial frequency spectrum. Here, we examined the effects of similarity of the sub-band image contrast on the observers' judgments. We calculated the contrast difference index (CDI) as follows. First, the difference in the RMS contrast between the original images and corresponding LLP/HLP images were calculated for each sample and its sub-band images as absolute values. Then, the contrast difference for the HLP image was subtracted from the LLP image in each sample, and the resulting value was defined as the CDI. Namely, a CDI was defined by the following equation:

\[
CDI = \frac{|C_{llp} - C_{org}| - |C_{hlp} - C_{org}|}{C_{org}/C_{llp}/C_{hlp}},
\]

where \(C_{org}\), \(C_{llp}\), and \(C_{hlp}\) were RMS contrasts of the original, LLP, and HLP images, respectively. Thus, positive values of this index mean that the contrast in an original image was closer to the HLP image than the LLP image. After the derivation of CDIs, we further calculated the correlation coefficient between the observer's choice ratio (z score) and CDIs.

The results are shown in Figure B1. In these plots, the positive correlation indicates that the observers tended to choose the image from the LLP/HLP images whose pixel or sub-band contrast was closer to the original image. The correlation coefficients were positive only for the leather and the fabric samples, especially for the high-spatial-frequency components. This suggests possible impacts of sub-band contrast similarity on observers' judgments. In contrast, the correlations were almost zero or negative for the stone and wood samples, suggesting that the sub-band contrast (or spatial frequency spectrum) similarity cannot explain the observers' responses for the stone and wood samples. In summary, differences in spatial frequency components between the original and HLP/LLP images only partially explain our results.

![Figure B1. Results of an additional analysis on the basis of luminance contrast proximity between original and LLP/HLP images.](image-url)
Appendix C

Although we instructed the observers to judge the closeness of material impression, it is possible that the observers could have relied more on image similarities and not on material impression similarity. To examine this possibility, we performed a supplementary experiment using phase-scrambled images whose simple luminance statistics were the same as that of the images in the original experiment (original images hereafter), which were used as stimuli. For these phase-scrambled images, material impressions could be weakened significantly. If the observers’ responses completely relied on image similarities and not on material impression similarities, the results in the supplemental experiment should be nearly the same as those in the original experiment.

In this supplemental experiment, we created a new set of images by randomizing the phases for each spatial frequency in the original object images. The procedure was as follows. First, we performed a two-dimensional fast Fourier transformation (FFT) on the original object images. Then, after we randomized the phase components in the spectral domain (phase scramble), the images were again transformed back to the spatial domain with an inverse FFT process. Finally, we mapped the luminance histogram from the original image to the corresponding phase-scrambled image. According to this procedure, the new image had a completely identical pixel-based luminance histogram and nearly identical spatial frequency amplitudes to the original image. This procedure enabled us to weaken the material impression of images while keeping low-level image statistics unchanged. Examples of phase-scrambled images are shown in Figure C1. The effects of the phase scramble seem to differ across samples; especially in samples with regular luminance patterns, the effects seem to be minor. The task of the observers was almost the same as in the main experiment. However, the observers were allowed to judge on the

![Figure C1. Examples of phase-scrambled images used in the supplementary experiment. From top: (a) stone, (b) wood, (c) leather, and (d) fabric categories.](image-url)
basis of image similarities only when judgments based on material impression was difficult.

The results are shown in Figure C2. Each panel shows the relationship between the z-scored choice ratios in the original experiment (on the horizontal axis) and those in this supplemental experiment (on the vertical axis) in each category. The choice ratios for the fabric samples were strongly correlated with the experiments. This result suggests that the observers' judgments for the fabric samples relied mainly on simple image statistics. The relationship of material impressions of fabric samples with spatial frequency amplitudes has been reported in a previous study (Giesel & Zaidi, 2013). In contrast, for other category samples, the correlations between the experiments were weaker. In addition, the range in z scores of the choice ratios for the stone and wood samples were reduced compared with the results of the original experiment with one exception in each category. This indicates an increase of the task difficulty for the phase-scrambled images. Therefore, the observers' choice ratios for the stone and wood samples may not be fully accounted for by image similarity judgments. In association with the image similarity judgments, typical texture-discrimination models based on local orientation filters (e.g., Landy & Bergen, 1991; Regan & Hong, 1995) may not capture the differences in the observers' responses.
between the original and supplemental experiments at least for stone and wood samples. This is because the phase-scramble procedure hardly affected the spatial frequency and orientation content of the images.

In summary, the results of this supplemental experiment suggest that the results of our main experiment reflect not only in image similarity judgment, but also in other types of similarity judgments, presumably related to surface quality.