Image statistics for surface reflectance perception

Lavanya Sharan,[1,*] Yuanzhen Li,[2] Isamu Motoyoshi,[3] Shin'ya Nishida,[3] and Edward H. Adelson[2]

[1]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

[2]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

[3]*Human and Information Science Laboratory, NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3-1, Morinosato–Wakamiya, Atsugi, Kanagawa 243-0198, Japan*

[*]*Corresponding author: l_sharan@mit.edu*

Received August 3, 2007; revised November 27, 2007; accepted January 11, 2008; posted February 6, 2008 (Doc. ID 85917); published March 13, 2008

Human observers can distinguish the albedo of real-world surfaces even when the surfaces are viewed in isolation, contrary to the Gelb effect. We sought to measure this ability and to understand the cues that might underlie it. We took photographs of complex surfaces such as stucco and asked observers to judge their diffuse reflectance by comparing them to a physical Munsell scale. Their judgments, while imperfect, were highly correlated with the true reflectance. The judgments were also highly correlated with certain image statistics, such as moment and percentile statistics of the luminance and subband histograms. When we digitally manipulated these statistics in an image, human judgments were correspondingly altered. Moreover, linear combinations of such statistics allow a machine vision system (operating within the constrained world of single surfaces) to estimate albedo with an accuracy similar to that of human observers. Taken together, these results indicate that some simple image statistics have a strong influence on the judgment of surface reflectance.

OCIS codes: 330.0330, 330.4060, 330.5000, 330.5510, 330.7310.

1. INTRODUCTION

The albedo of a surface is a measure of its diffuse reflectivity. Perceived albedo is known as "lightness," and the ability to judge albedo with some accuracy, despite changing viewing conditions, is known as "lightness constancy." Lightness constancy is not perfect, especially in extreme conditions such as those arranged by Gelb [1]. When an ideal matte, planar surface is viewed in isolation, one cannot determine its albedo. A black surface may be seen as white, an illusion known as the Gelb effect. Here surface luminance is the only relevant stimulus parameter, since illumination and albedo are confounded (they multiply together to produce the observed luminance). Therefore, lightness constancy is poor due to the lack of any disambiguating information from the context.

With nonideal surfaces, the Gelb demonstration fails. It fails badly for complex surfaces such as stucco [2], as shown in Fig. 1. The two stucco images have the same mean luminance and are surrounded by the same dark background, yet one looks darker and glossier than the other. The interreflections and specularities in these surfaces seem to provide extra information, and observers are evidently able to utilize some of this information to achieve lightness constancy. Gilchrist and Jacobsen built boxes containing miniature rooms and painted them with either black or white matte paint [3,4]. Observers viewed the rooms, one at a time, through a small aperture, so they were immersed in a field of uniform reflectance. Observers could tell which room was which, presumably because of differences in interreflections.

Most research in surface perception has been dominated by the case of smooth, Lambertian patches and planar 3-D configurations, such as those shown in Figs. 2(a) and 2(b) [1,5–14]. Existing theories of lightness perception have no way of predicting the effects that occur with real-world surfaces such as stucco. However, several recent studies, including ours, have examined stimuli that incorporate some of the complexity of real-world conditions [Figs. 2(c) and 2(d)] [2–4,15–23].

In the fields of computer graphics and computer vision, there has been an interest in characterizing the bidirectional reflectance distribution function (BRDF), which is a full description of the reflectance properties of an opaque surface [24]. Estimating BRDF of surfaces from photographs is a challenging machine vision problem. Work in these fields has mainly followed the inverse optics approach, aiming to recover such a full model of 3-D layout and illumination of the scene as being consistent with a given 2-D image. This is an impossibly difficult problem given the many-to-one mapping from 3-D scenes to a 2-D image. Therefore, existing algorithms require additional constraints or assumptions that go far beyond those included in a single picture such as Fig. 1 [25–35]. The mechanism of human perception must be different.

We have taken up the lightness perception problem from several points of view. First, we ask how well human observers can judge albedo when viewing isolated surfaces using materials such as stucco. Second, we propose that simple statistics of the 2-D image of a surface can be used in a cue-based approach to lightness perception. Finally, we show that manipulating these candidate statistics in an image alters human judgments in a predictable manner. Our approach is to be contrasted with inverse optics approaches, which depend on the estimation of the

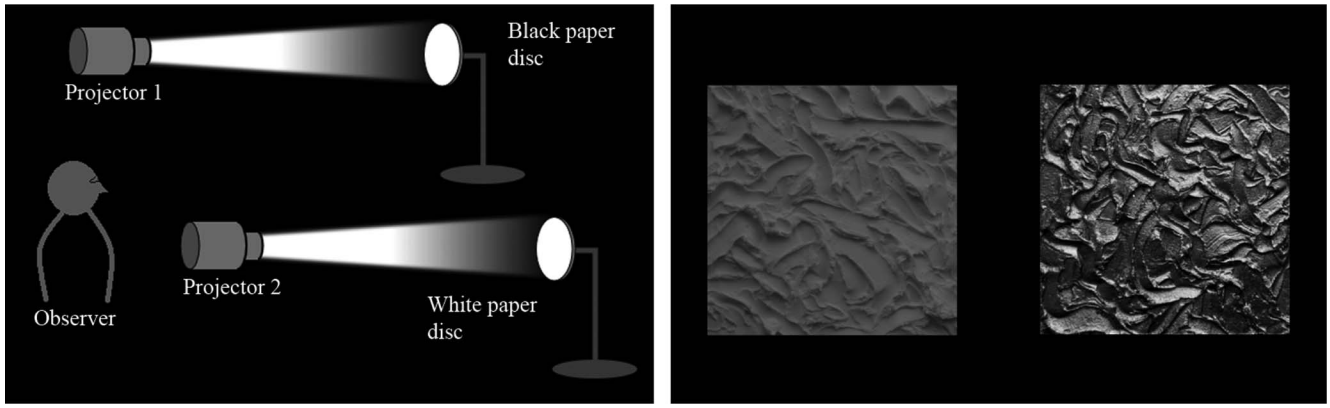


Fig. 1. (a) In the Gelb demonstration a smooth Lambertian black surface can look the same as a smooth Lambertian white surface. (b) Gelb effect fails for complex surfaces. The stucco samples have the same mean luminance, yet it is easy to tell the white stucco from the black.

parameters of an internal model that can explain the image data in detail. There is evidence that inverse optics models can be useful in understanding some human judgments. However, in real-world scenes, the surface geometry, illumination distribution, and BRDF are too complex and too uncertain for inverse optics to have much success. Therefore, it is reasonable that the visual system will use heuristics based on statistical cues when these cues are informative. The importance of image statistics was suggested in a study by Nishida and Shinya [15], which found that reflectance perception of non-Lambertian and nonsmooth surfaces was critically dependent on the luminance histogram of the 2-D image of a surface. Dror *et al.* [16,17] studied the appearance of spheres in real-world illumination and found that simple image statistics were useful in characterizing the reflectance properties of synthetic and natural spheres.

We consider lightness perception for photographs of opaque surfaces viewed in isolation, with the mean luminance scaled to a constant value for all surfaces. Our surfaces have significant mesostructure so that shading, interreflection, and specular highlights become significant components of the appearance. In our previous work [2] we reported that the skewness of the luminance histogram is correlated with the albedo and gloss of real-world surfaces. Human judgments of lightness and glossiness were also correlated with luminance skewness. We suggested that this statistic can be easily computed by early neural mechanisms and found an aftereffect that supports this hypothesis.

In the present work, we took a more computational approach to the problem, focusing in detail on the statistics that are associated with lightness. We evaluated the absolute effectiveness of a variety of image statistics and their combinations in estimating the physical albedo not only from correlations, but also from how well machine learning algorithms can tell light and dark surfaces based on those statistics. The results suggest that moment and percentile statistics of the luminance histogram and subband histograms are informative. Although learning algorithms cannot predict lightness perfectly, their performance is similar to that of human observers. In addition, the pattern of errors made by the algorithms was very similar to that of human errors. On changing these statistics of images, human judgments were affected accordingly. These findings suggest that human observers use histogram statistics for lightness estimation. Finally, in order to manipulate the subband histograms in addition to the luminance histogram [2] without introducing image artifacts, we developed a modification to the Heeger–Bergen texture synthesis algorithm.

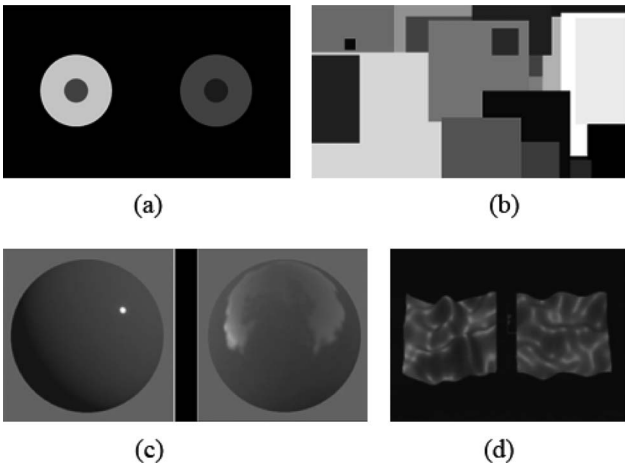


Fig. 2. Stimuli used for studying reflectance perception: (a) Wallach's disc annulus displays, (b) Mondrian-like displays with flat Lambertian surfaces, (c) Fleming *et al.*'s simulated spheres in complex real-world illumination [18], (d) simulated locally smooth bumpy surfaces used by Nishida and Shinya [15].

2. IMAGE STATISTICS AS CUES TO LIGHTNESS

In this section, from the viewpoint of ecological optics, we analyze how simple statistics of the 2-D image of a surface tell us about surface reflectance properties. Our approach is similar to that of Dror *et al.* [16,17], who considered images of smooth, shiny spheres rendered or photographed under complex, real-world illumination conditions. They measured moment (2nd, 3rd, and 4th) and percentile (10th, 50th, and 90th) statistics of pixel in-

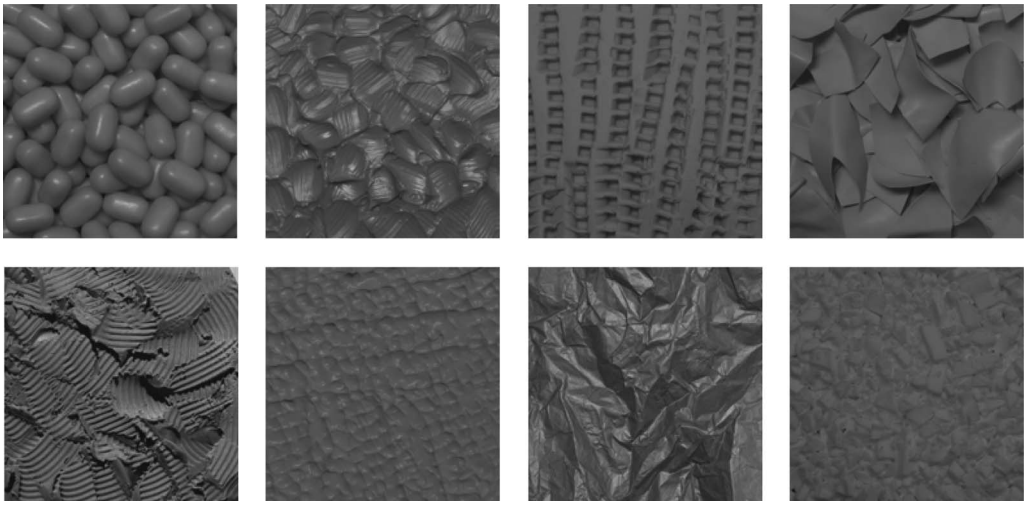


Fig. 3. Examples of surfaces in our data set. All surfaces shown here were photographed under an overhead fluorescent light.

tensities and wavelet coefficients on the surface of the spheres via a cylindrical projection of the 2-D image. These statistical measurements were found to be useful for classifying the spheres into shiny, matte, white, gray, chrome, etc.

Like Dror *et al.*, we are interested in identifying image statistics that are diagnostic of surface reflectance properties. However, we operate under a set of assumptions different from theirs. We do not assume a known surface geometry; rather we allow our surfaces to possess 3-D medium-scale structure. We consider surfaces in simple, artificial illumination conditions, and while some of our surfaces are glossy (non-Lambertian), we focus on statistics that are predictive of albedo, i.e., the diffuse reflectance component.

A. Image Data

We gathered high-dynamic-range color photographs of several real-world surfaces, such as paper, candies, cloth, stucco, etc. (Fig. 3). Opaque surfaces with spatially uniform reflectance properties were selected so that each surface is associated with a unique albedo value. We used planar surface samples with medium-scale surface structure or surface mesostructure [36]. While we allowed the surfaces to be specular, we studied only the diffuse component. Surfaces were photographed under three indoor lighting conditions [see Figs. 4(a)–4(c)]. The specifics of the camera and lighting setup are provided in Fig. 4(d). All images were acquired in RAW 12-bit format by a Canon EOS 10D camera. The RAW images were linearized using "dcraw" software [37]. The linearization process converts the pixel intensities in a RAW image to the measured luminance up to a multiplicative scaling factor. Appendix C contains details of the linearization procedure.

Our surfaces were orange, yellow, red, white, or black (Fig. 3). We used 30 surfaces of various shapes and reflectance properties. As we are interested in lightness, all color photographs were converted to gray scale by separating the color channels. For colored surfaces, individual color channels were treated as distinct gray-scale images. Figures 4(e) and 4(f) show an example of the color components of an orange surface. The blue channel looks like a black surface, while the red channel looks like a white surface. This happens because, for orange colored materials, the different colors of light are reflected in different

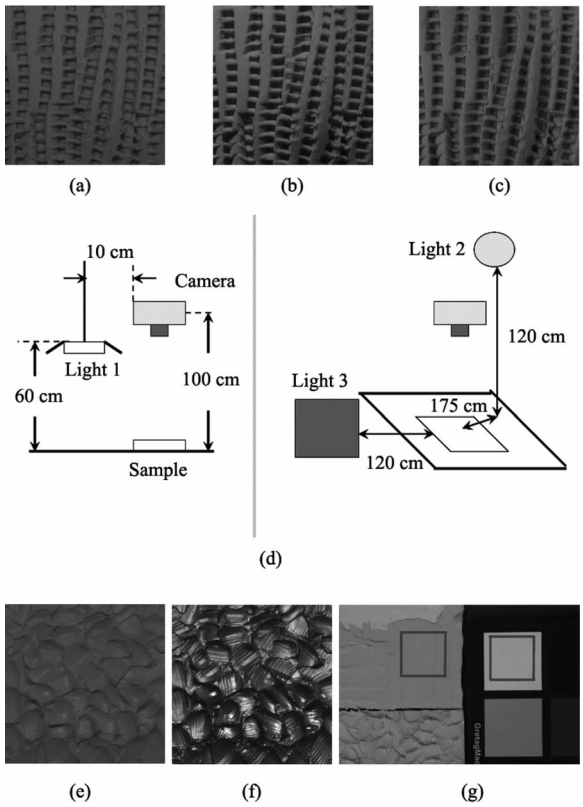


Fig. 4. Image data acquisition. Three indoor lighting conditions were used: (a) light 1, overhead fluorescent light source; (b) light 2, focused halogen spotlight; (c) light 3, diffuse tungsten halogen lamp. (d) Schematic layout of the setup. Two views, one from the side and one from front, are shown. (e), (f) Red and blue color components of an orange surface look like white and black surfaces, respectively. (g) Ground truth is acquired using a uniformly illuminated flat material sample and a standard white surface. A user clicks on two regions, one on the sample and one on the standard. The ratio of mean luminance of the two regions is used to calculate the albedo for each color channel.

ways. Thus, we can acquire photographs of surfaces that share identical geometry and illumination conditions but vary in their reflective properties.

Many of our materials exhibit a strong specular reflection component. One example is the crumpled black paper in Fig. 3. In order to capture such materials with a limited dynamic range camera, we used the technique of multiple exposure imaging. Multiple exposure photographs of the same scene were combined using HDRShop software to produce a single high-dynamic-range image [38]. As a final step, all images were multiplicatively normalized to have the same mean image luminance. This step is essential, because we want to know which statistics of an image, other than mean luminance, are useful for reflectance perception. In total, we had 30 materials $\times 3$ lighting conditions $\times 3$ color channels $=270$ images.

For all the surfaces in our data set, we acquired the ground truth for diffuse reflectance. A smooth, flat sample of each surface, devoid of any mesostructure, was selected and placed next to a standard white surface [see Fig. 4(g)]. For handmade surfaces, we prepared a flat sample by hand. For other surfaces, we used the flattest samples available. Both the standard and the sample were photographed under uniform illumination conditions. The diffuse reflectance of the sample was calculated by using the ratio of the linearized intensity in a region containing the sample to that of a region containing the standard. The regions in the photograph were selected carefully to avoid shadows and highlights.

B. Statistics of the Luminance Histogram

We studied the luminance histograms for the images in our data set and found that histograms of light (high-albedo) and dark (low-albedo) materials (most materials are non-Lambertian and nonsmooth) display characteristic differences. The luminance histograms for dark surfaces tend to have higher Michelson contrast (standard deviation divided by the mean) and have longer, positive tails. For lighter surfaces, the histograms have lower Michelson contrast and are usually symmetric (Fig. 5). These differences can be attributed to ways in which light and dark surfaces interact with light. Light surfaces have higher diffuse reflectance; therefore light bounces around filling up the shadows, leading to a lower contrast than dark surfaces. If a light and a dark surface have the same amount of specular reflection, the specular highlights are more visible in the darker surface owing to higher contrast. Therefore, contributions from interreflections and highlights lead to different shapes for the luminance histograms of light and dark surfaces.

These systematic differences in the luminance histograms can be captured by a host of statistical measures—

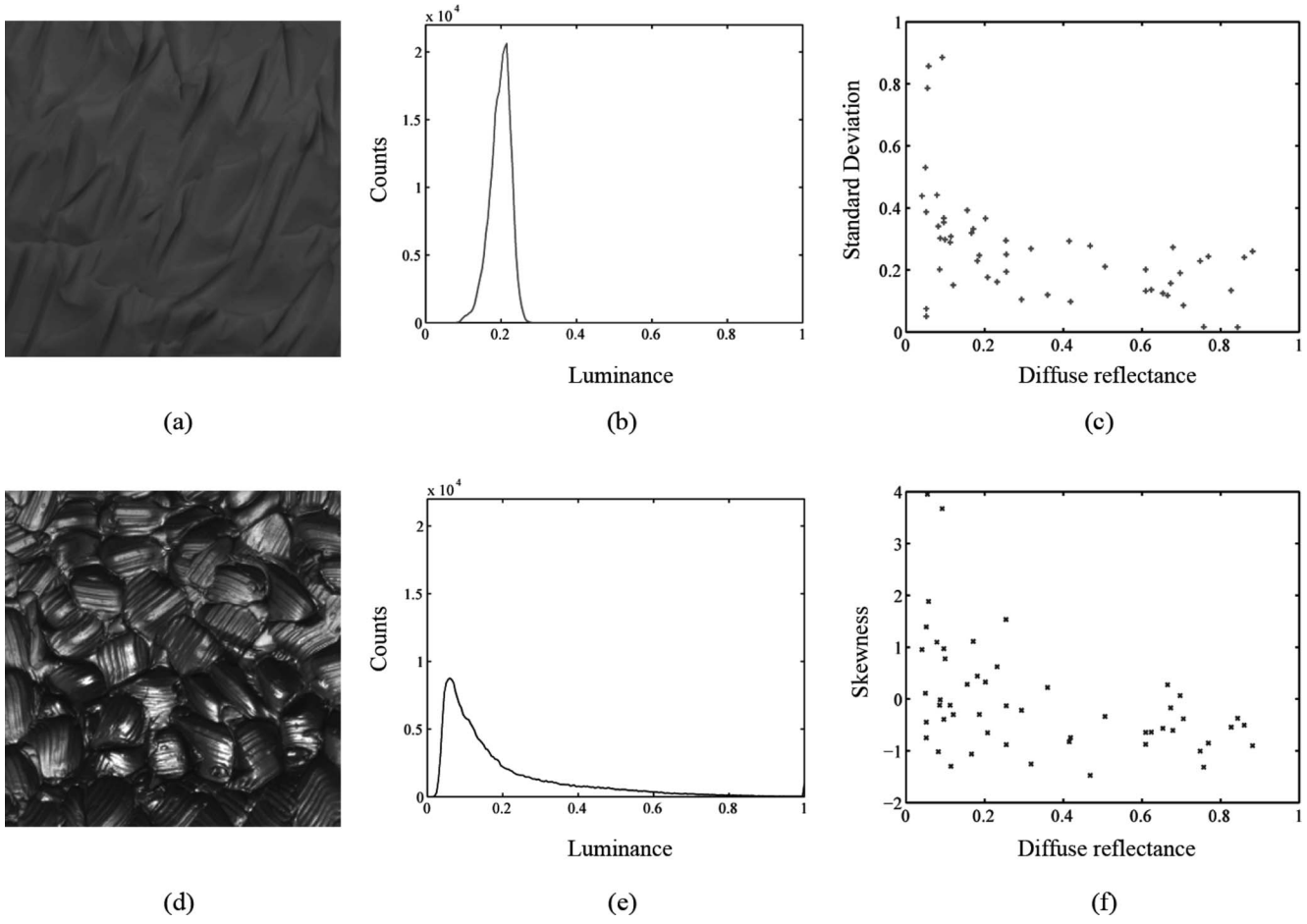


Fig. 5. Luminance histograms of light and dark materials exhibit systematic differences. (a) Light modeling clay; (d) dark stucco; and (b), (c) the respective luminance histograms. (c), (f) Standard deviation and skewness of the log-luminance histogram is plotted against the ground truth for albedo for all the surfaces in our data set. All plots pertain to the overhead fluorescent lighting condition.

moment (standard deviation or skewness) or percentile statistics (10th, 50th, or 90th). In our companion paper [2], we had focused on one statistic, skewness or the third standardized moment of the luminance histogram, and we showed that skewness is correlated with albedo as well as gloss. In the present study, from the viewpoint of ecological optics, we consider a wide range of statistics that are correlated with albedo, such as standard deviation and percentiles of the original and filtered images. We compute these statistics either directly on the luminance values or on the log of luminance values. We found that statistics of the log-luminance histogram are distinct for light and dark surfaces, just as for the luminance histogram.

In the rest of this paper, we will discuss the results of log-luminance analysis. In our experience, the conclusions remain the same for luminance and log-luminance statistics. This is trivially true for order statistics that are unchanged by a log transformation.

In Figs. 5(c) and 5(f), the standard deviation and skewness of the log-luminance histogram is plotted against the true albedo of a surface. The dependence of the moment statistics on the physical property can be seen in these plots. Receiver operating characteristic (ROC) analysis offers another way of visualizing this correlation. The ROC is a plot of the true positive rate versus the false alarm rate of a binary classifier. A perfect classifier achieves 100% classification accuracy with a 0% false alarm rate, and the area under the ROC curve (AUC) is 1.0. For the worst classifier (unbiased coin flip), the true positive rate equals the false alarm rate (AUC=0.5). In our case, if any of the statistics—90th percentile, standard deviation, or skewness of log-luminance—is used to classify surfaces as light (physical albedo <0.5) or dark (albedo >0.5), the ROC curves lie somewhere in between the ideal and the worst classifier curves (AUC ranges from 0.73 to 0.77) [Fig. 6(a)]. The performance of moment statistics is significantly above chance, implying that both statistics yield useful information about albedo. Other percentile statistics (10th and 50th) also have similar ROC curves (AUC=0.69 and 0.77, respectively).

While ROC analysis considers the utility of a statistic for binary reflectance classification, we can use regression analyses to see how well the statistics *estimate* reflectance. Figure 6(b) shows that a linear regression fit is an inadequate model for the relationship between skewness of log-luminance and albedo. Similar plots were obtained for other moment and percentile statistics of log-luminance. We also conducted nonlinear regression analyses in order to model the data in Fig. 6(b) better. We found that applying a log transformation to the axes of Fig. 6(b) leads to a somewhat improved linear fit [Fig. 6(c)]. As skewness and albedo are dimensionless quantities, applying a log transformation does not change the physical significance of our results. Figure 6(d) shows the log–log plot for the 90th percentile of log-luminance histogram and albedo. Visualizing the relationship between our statistics and albedo is easier after applying the log transformations. However, as the r^2 statistic in Figs. 6(c) and 6(d) indicates, these nonlinear transformations do not capture the dependence of statistics on albedo entirely. We did not use more complex models to nail down the behavior of our data because of the danger of overfitting. We have only 30 materials in our data set. While the linear fits in Figs. 6(b)–6(d) are not perfect, they are still statistically significant. Therefore, our statistics contain useful information, although the relationship between statistics and diffuse reflectance is not entirely straightforward.

It is important to emphasize that the statistics described thus far are predictive of albedo as long as the images on which they are computed look like surfaces. The same statistics are of no use when measured on arbitrary images that are not associated with a value of albedo [2]. Indeed if we pixel scramble our images, thereby destroying the perception of a surface, the luminance statistics just described remain unchanged.

C. Statistics of Subband Histograms

As luminance (or log-luminance) statistics are insensitive to spatial structure, we examined the statistics of filter outputs next. This is a reasonable thing to do because the visual system is more likely to have access to filtered values than to raw luminance values. We used center–surround and oriented edge detection filters in a multiscale decomposition [39,40]. In Fig. 7, one observes that pixel histograms of filtered images look different for light and dark surfaces. For dark surfaces, filter output histograms have heavier tails and, in the case of center–surround filtering, the outputs are also skewed.

The filters amplify the local contrast differences between white and black surfaces. The skewness of the center–surround filter outputs is presumably related to the asymmetry in the distribution of shadows and highlights in natural images. Unlike shadows, which tend to be spread over a larger image region, specular highlights tend to be small and concentrated. It is likely that filter output statistics are affected by these characteristic asymmetries. Figures 8(a) and 8(b) plot ROC curves for subband statistics—standard deviation and 90th percentile. Individual statistics fare much better than chance and hence are predictive of albedo. Figures 8(c) and 8(d) plot the log of the 90th percentile of filter outputs against the log of albedo. The linear regression fits in Figs. 8(c) and 8(d) demonstrate that the filter statistics are highly correlated with albedo.

In our experiments, we also found that the skewness statistic is sensitive to the choice of filters and filter parameters. Center–surround filters are somewhat better than oriented filters at the task of skewness detection. However, in terms of albedo prediction, all statistics, both of log-luminance and the two kinds of filters, perform about equally well.

D. Combining Statistics

Given that individual statistics can predict albedo fairly well, it is interesting to ask how the statistics perform relative to one another. In Figs. 6 and 8 we observe that all statistics—moments and percentiles derived from luminance or filter outputs—perform about equally well in ROC tests and regression analyses. Not only do all the statistics predict albedo with the same degree of success, but we found that they are also correlated with one another. Consider Fig. 9(a): At first glance, it is not clear

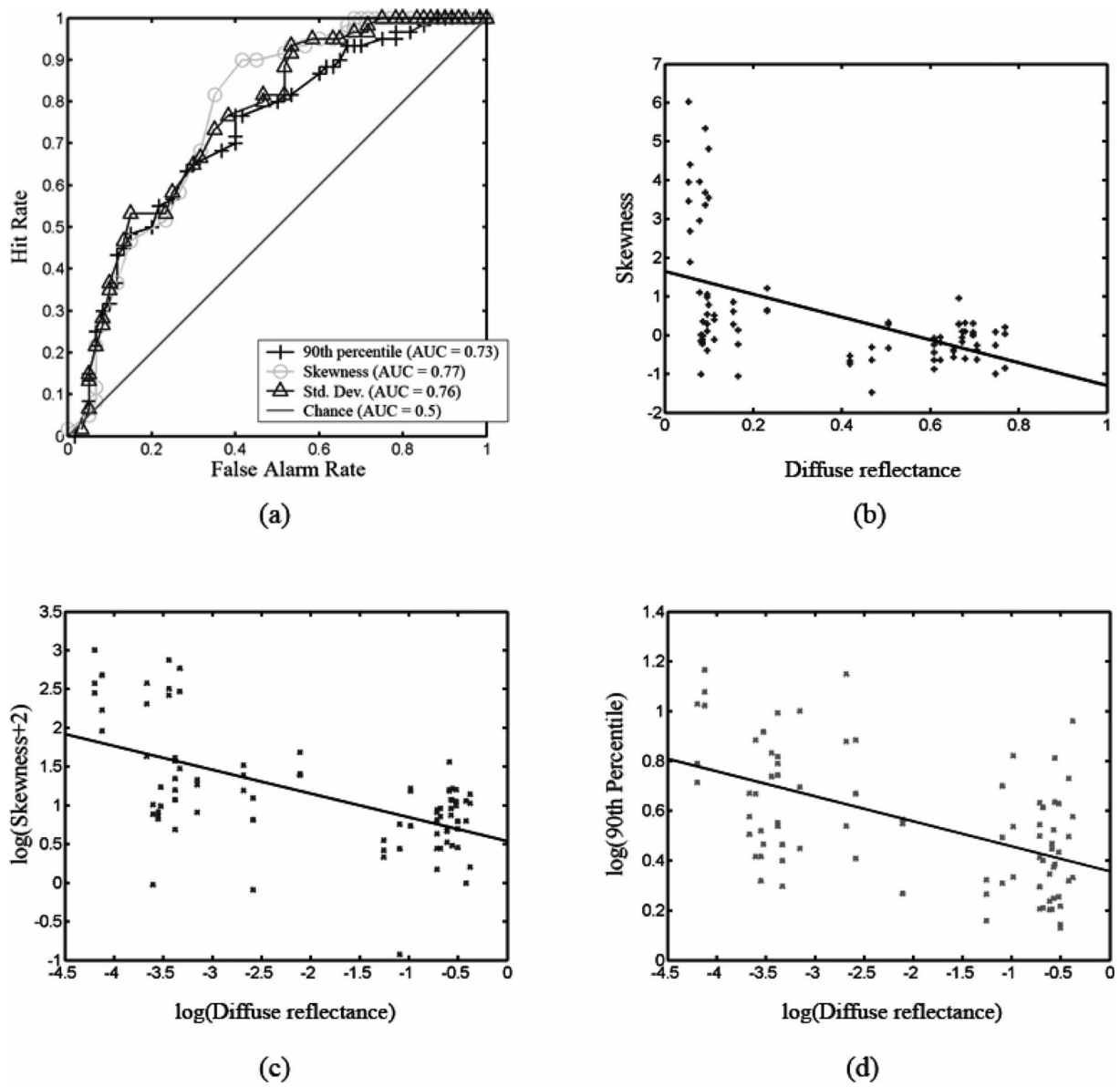


Fig. 6. (a) ROC curves for the 90th percentile, standard deviation, and skewness of log-luminance values. These statistics perform well above chance at the task of classifying surfaces as light or dark. (b) Skewness of the log-luminance histogram is plotted against the physical diffuse reflectance. A linear regression model is a poor fit ($p<0.05$, $r^2=0.27$). (c) Applying a log transformation to both axes of (b) improves the fit of the linear regression model ($p<0.05$, $r^2=0.34$). (c) Log of 90th percentile of the log-luminance histogram is plotted against the log of the diffuse reflectance. The linear fit is still not very good ($p<0.05$, $r^2=0.29$). Statistics were pooled across all lighting conditions for all plots in this figure.

why the standard deviation of log-luminance and the 90th percentile statistic of the center–surround filter outputs should covary. We performed chi-square independence tests and mutual information values to confirm these empirical correlations. In our previous work [2], we had observed that the skewness of filter outputs is highly, though not completely, correlated with the skewness of the luminance histogram for images like those in Fig. 3.

In a sense, it is not surprising that the statistics are dependent on one another. We noted earlier that shapes of the histograms of log-luminance and filter outputs have distinctly different shapes for light and dark surfaces. These characteristic differences can be captured in various ways by various statistics. The surfaces with lower albedo have higher local contrast as well as more structure at higher frequencies. Therefore, the statistics that measure contrast (e.g., standard deviation) covary with the statistics that measure energy in higher frequency (e.g., 90th percentiles) for each surface.

In spite of the high degree of correlation, we found that for purposes of albedo estimation, combining a few statistics is better than using just one. We used a support vector regression technique with a linear kernel to combine statistics [41]. The image data set was divided into three groups, one of which was chosen as the training set. The regression technique learned a linear relationship between a chosen set of statistics (features) measured on the training set and the ground truth values for albedo. A linear kernel was chosen for simplicity. The ε parameter of the regression was set to 0.1, and the penalty param-

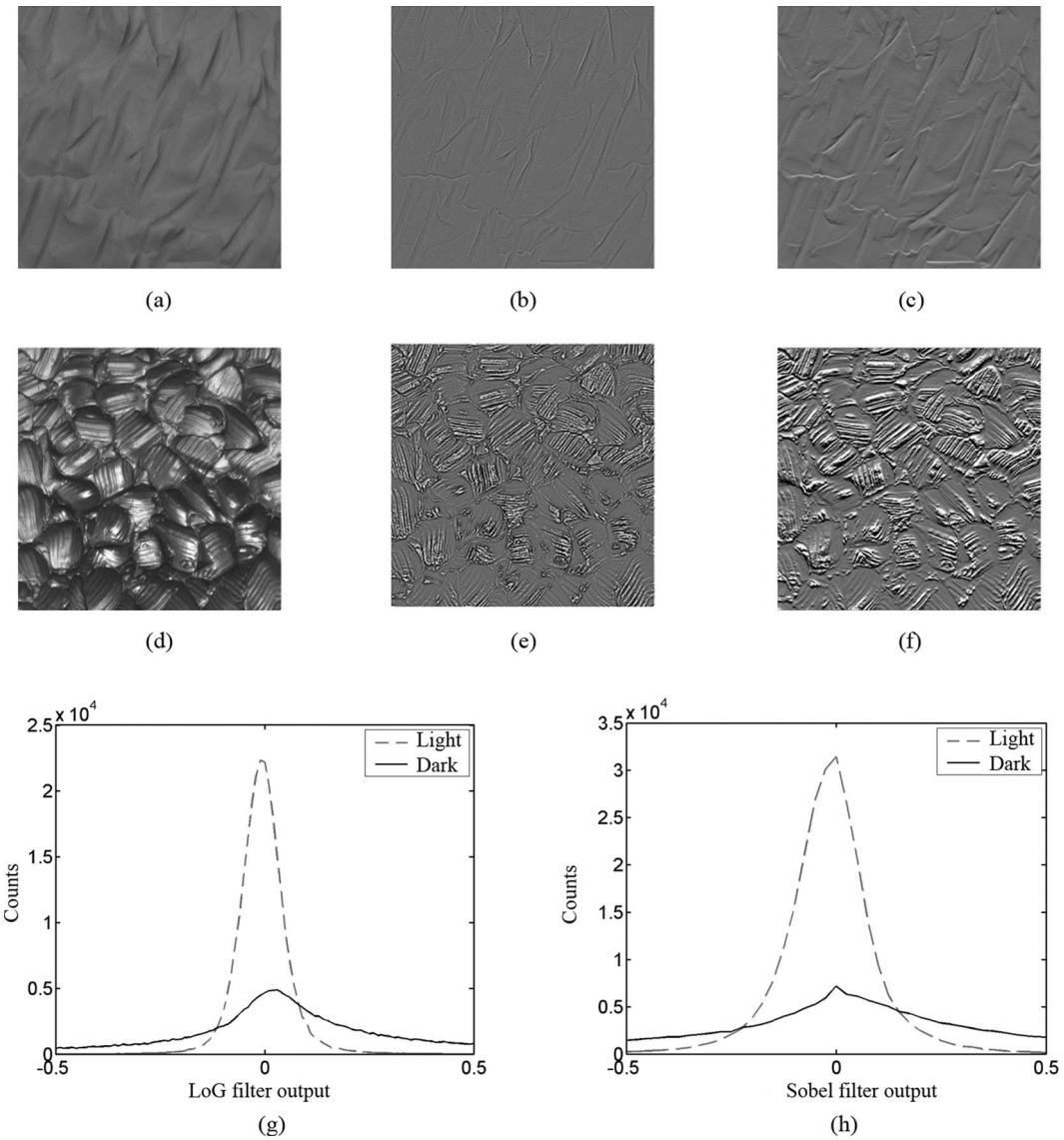


Fig. 7. Pixel histograms of filtered images look different for white and black surfaces. (a) Light modeling clay from Fig. 5; (b) output of Laplacian of Gaussian filter (σ=0.5 of size 5×5 pixels); (c) output of horizontal Sobel filter (3×3 pixels); (d) dark stucco from Fig. 5. (e), (f) LoG and Sobel filter outputs; (g) pixel histograms of images in (b) and (e); (h) pixel histograms of images in (c) and (f). All plots pertain to the overhead fluorescent lighting condition.

eter C was chosen by fivefold cross validation on the training set. Once the regression parameters are learned, the technique can provide an estimate of the albedo for any new image of a surface. It is important to clarify that the regression technique fits a linear model to our statistics to predict the albedo. Even though we noted earlier (Figs. 6 and 8) that a linear model can predict "log(albedo)" from "log(statistic)" somewhat better than albedo from statistic, we did not use log transformations in our regression model. This is because the increase in prediction performance is not enough to justify the added complexity of the regression model.

Figures 9(b)–9(d) show the outputs of three linear models that differ in the number and type of statistics they use to predict albedo. In these figures we see that the performance of a linear model improves by using two statistics instead of one. However, on adding any more statistics, the gains to be made are not significant. As we tried to incorporate more than two or three statistics as features in our models, the correlation among the statistics led to saturation in performance. We found that the precise choice of features (moments or percentiles, luminance or filter statistics) or the exact number of features (two, three, or four) is not too critical. So, for the rest of this work, we will use a fixed linear model (henceforth referred to as the "model" or "regression technique"). The model uses three statistics—standard deviation, 10th percentile, and 90th percentile of the center–surround filter output.

In Fig. 10, the output of the model is compared to the ground truth for albedo. We see that the regression technique is not perfect at estimating the physical albedo, but it does a fairly good job. Interestingly, we found that the technique makes larger errors on surfaces that are flat and nearly Lambertian than those with more complex geometry and reflectance properties. This performance is

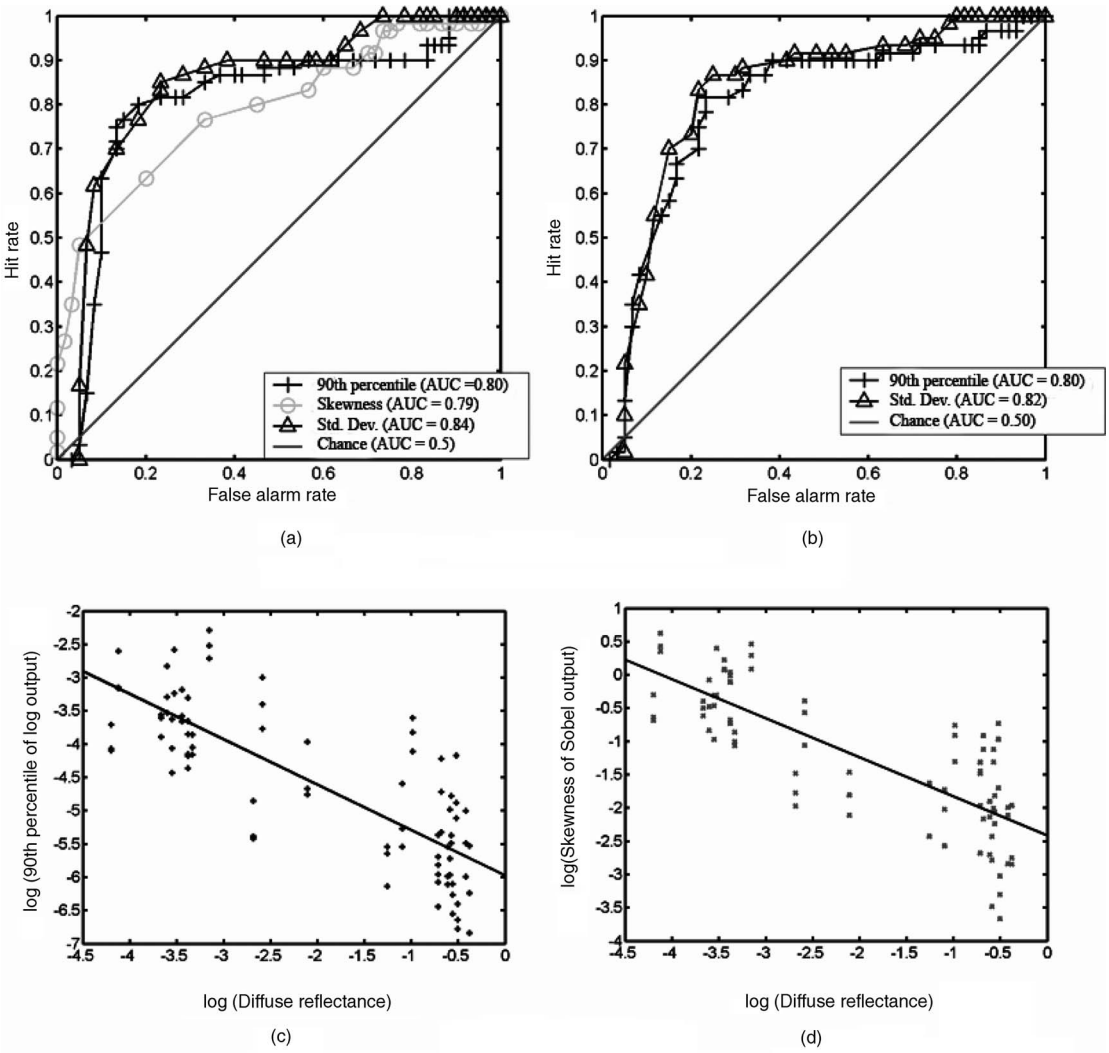


Fig. 8. ROC curves for statistics of filter outputs: (a) 90th percentile, skewness, and standard deviation of LoG filter output; (b) 90th percentile and standard deviation for Sobel filter output are significantly above chance; (c), (d) plot the log of the 90th percentile of the LoG and Sobel filter outputs, respectively, against the log of physical diffuse reflectance. The linear regression fits are shown as black lines ($p < 0.05$); r^2 values are 0.63 and 0.62 for (c) and (d), respectively.

similar to human observers (Fig. 1). To study the correlation between statistics and perception better, we conducted psychophysical experiments to measure human performance on our images.

3. EXPERIMENT I

We asked human observers to rate the lightness for all the photographs in our data set in a context similar to the Gelb conditions of Fig. 1. From our informal "anti-Gelb" observations in Fig. 2, we know that such a task is meaningful. It is easy to judge the lightness of rough non-Lambertian surfaces in isolation.

A. Stimuli

The image data of the previous section were used as stimuli for this experiment. The mean luminance equalized images were displayed on a gamma-corrected LCD monitor. The images were displayed at a resolution of 512×512 pixels against a middle gray background. The intensity of the background was set to the mean image luminance; thus both variables do not change throughout the experiment.

B. Apparatus

The three indoor light sources [see Figs. 4(a)–4(d)] that were used to photograph our surfaces were an overhead fluorescent lamp (Kino Flo Diva Lite 200) placed 60 cm above the sample surface, a halogen spotlight (LTM Pepper 300 W Quartz-Fresnel) placed 175 cm away and 120 cm above, and finally a light box (Lowel Rifa 66, 750 W tungsten halogen lamp) that produced diffuse, soft lighting 120 cm away from the surface. The LCD monitor was a Dell 20.1 in. flat panel (1 in. = 2.54 cm) at 1280×1024 resolution, 75 Hz frame rate, and 70 cd/m^2 mean luminance. To obtain the ground truth [see Fig. 4(g)] the standard white surface was chosen from the Gretag Macbeth Color Checker chart. Light meters Sekonic L-608 and Minolta CS-100 were used to ensure uniform illumination.

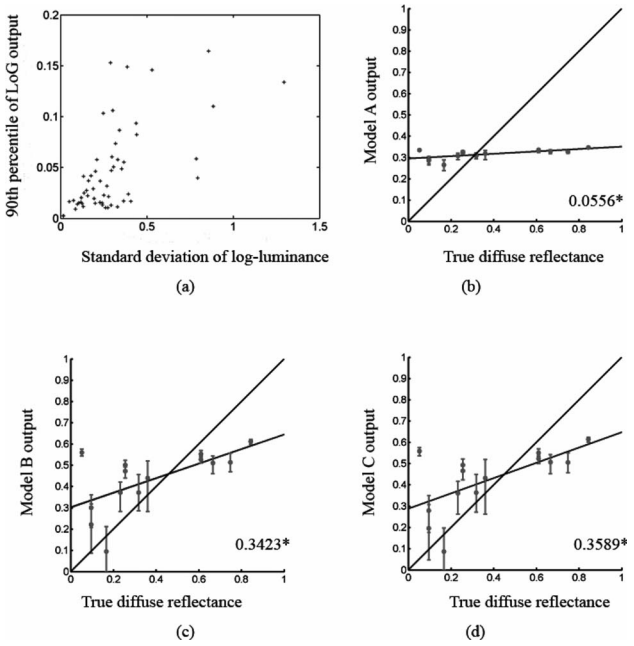


Fig. 9. (a) Standard deviation of log-luminance and the 90th percentile of LoG filter output are correlated ($r=0.6237$, $p<0.05$). (b), (c), (d) Outputs of three linear models are plotted against the true diffuse reflectance for a subset of our surfaces. Model A uses one statistic, the standard deviation of LoG filter outputs to predict the albedo of a surface. Model B uses two statistics—standard deviation and 10th percentile of LoG filter outputs. Model C uses three statistics—standard deviation and the 10th and 90th percentiles of LoG filter outputs. For all three cases, the model ratings were averaged over all three lighting conditions. The error bars indicate the minimum and maximum ratings. If the models were perfect at predicting physical albedo, all points would lie along the black line with slope=1. The slopes of the best fit lines are indicated in each plot. The asterisk denotes statistical significance ($p<0.05$). The r^2 statistic is similar for all plots—0.42 for (b), 0.38 for (c), and 0.40 for (d).

Observers viewed the LCD monitor in a dark room. A box enclosing the two light sources and a Munsell chart with standard surface patches served as the reference (see Fig. 11). The box was constructed from white foam board panels and covered with dark gray craft paper on the outside. One side of the box was left open to allow observers to view the Munsell chart. Compact fluorescent light bulbs of color temperature 5500 K (SunWave full spectrum CFL bulbs) were used to uniformly illuminate the chart. The Munsell chart comprised eight gray squares, numbered 1 to 8, on a random noise background [see Fig. 11(b)]. The gray squares were matched by eye to the Munsell standard reflectance, N2 through N9 (Gretag Macbeth 31-step neutral value scale) under the SunWave bulbs. The squares as well as the random noise background were printed on Epson enhanced matte paper using an Epson Stylus Photo R800 printer.

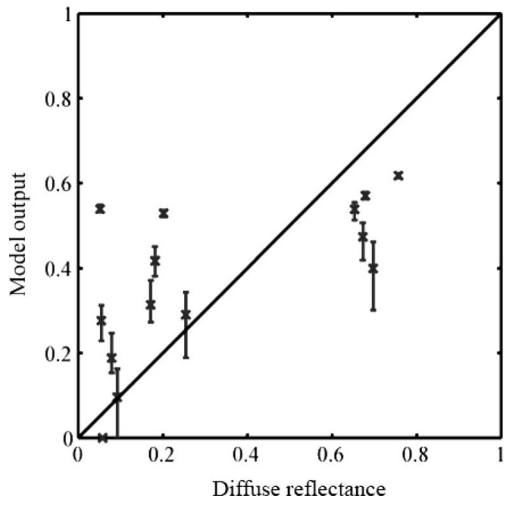


Fig. 10. Output of the regression technique is plotted versus ground truth for diffuse reflectance (albedo). Bars indicate maximum and minimum ratings. If the technique were perfect, all points would lie along the diagonal.

C. Procedure

Observers viewed the photographs, one after another, and provided ratings between 1 and 8 to indicate the standard Munsell patch that was closest in reflectance to the stimuli. Fractional ratings such as 4.5 were permitted to allow observers to express their answers at a finer resolution than the Munsell scale provided to them. However, most observers did not use fractional ratings. For the few who did, the fractional ratings were converted to the equivalent albedo value. Observers were divided into three groups. Each group viewed a different set of surfaces. The experiment was self-paced. For each surface, three repetitions were run for each lighting condition. The order of the images was randomized. The experiment lasted 30 min.

D. Observers

Twelve observers with normal or corrected-to-normal vision participated in the experiment. All observers were naive to the purpose of the experiment.

E. Results

We found that observers can, to some extent, estimate the albedo or diffuse surface reflectance under our experimental conditions. Figure 12 plots the perceived diffuse reflectance versus the ground truth for observers in one group. Observers are not perfect at estimating ground truth, but they perform reasonably well. In Fig. 12, we reject the null hypothesis ($p<0.05$) that there is no linear relationship between observer data and ground truth. Therefore, contrary to the predictions made by classical lightness theories, human observers can judge lightness in the absence of mean luminance information and context. Our observers tend to agree with one another (see Fig. 13). We analyzed the deviation of observer ratings from ground truth, i.e., the errors observers made. We found that the size of the error does not seem to be related to the physical reflectance of the surface. In other words, black materials, are not harder to judge than white materials, for example. Instead, it seems that the closer a surface is to the flat, smooth, purely matte ideal of Fig. 1(a), the harder it is to judge its diffuse reflectance (see Fig. 14).

In our data, the effect of lighting was not significant. In Fig. 15 we see that changing the illumination does not affect the perceived reflectance of a surface too much. This observation is consistent with the work of Fleming *et al.* because our illumination conditions did not vary as drastically as theirs [18]. Fleming *et al.* showed that reflec-

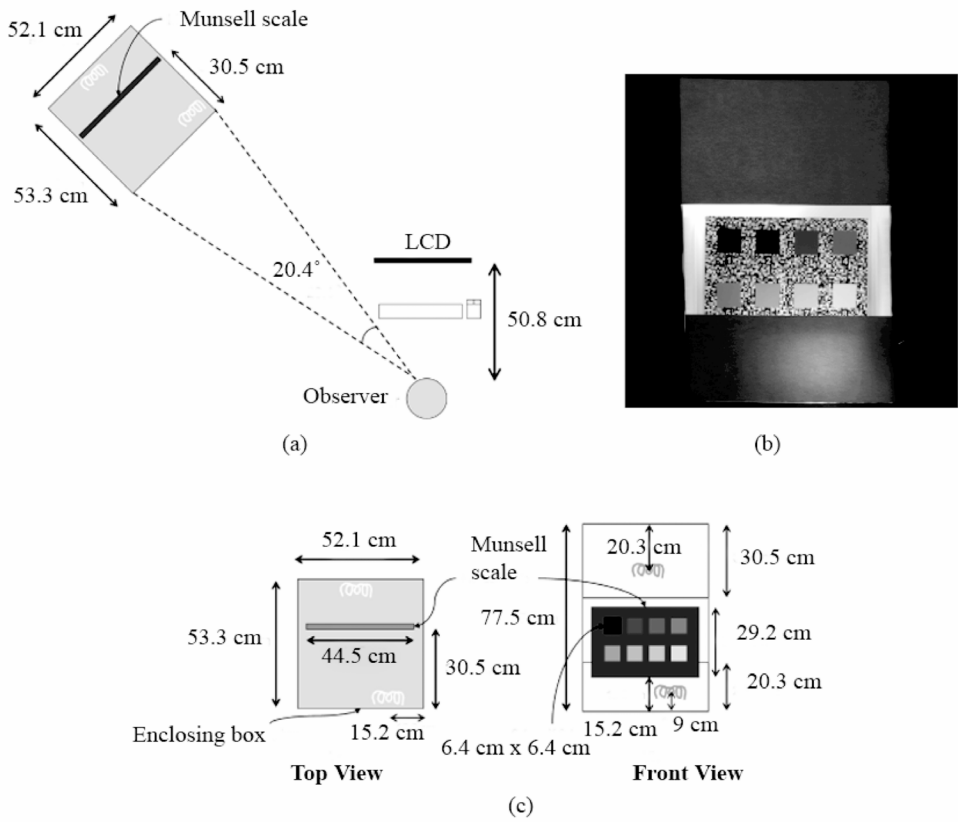


Fig. 11. Observers viewed stimuli on the LCD panel and matched reflectance properties of the surface on the screen to one of the standard surfaces on the Munsell chart. (a) Dimensions of experimental layout, (b) photograph of the reference box, (c) dimensions of the box.

tance perception is significantly poorer under atypical illumination (e.g., Gaussian noise) than under real-world illumination. Our lighting conditions fall in the real-world category and are all indoor laboratory setups. Therefore, it is not surprising that observer performance did not vary with illumination. It is important to note that in our data set, we used a fixed viewpoint and probed three illumination directions. Unlike databases acquired for BRDF measurements, such as the CURET database, we did not explore the space of lighting directions and viewpoints extensively [42]. One may ask, What effect do variations in lighting direction and camera viewpoint have on our results? In Appendix B we have addressed this issue in detail.

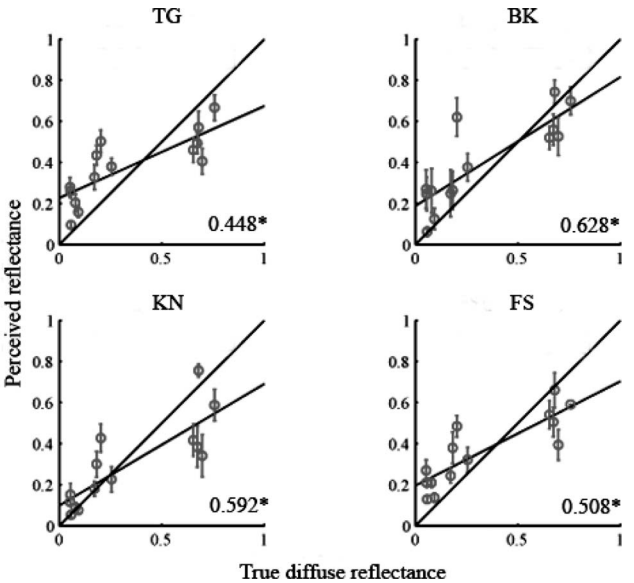


Fig. 12. Results of experiment I. Perceived albedo versus ground truth for four observers. Responses were pooled across all lighting conditions. Error bars indicate 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope=1. The gray line is the linear regression fit to each observer's data. The slope of the best fit line is indicated in each plot. The asterisk denotes significance (p-value <0.05).

Finally, as a consequence of using colored surfaces, we could study how observers rated the different color channels. Figure 16 displays observer ratings for the R, G, and B channels of a colored surface. The color channels differ only in the diffuse and specular reflectance properties; the illumination conditions and surface geometry remain the same. Observers rate each color channel differently, thereby establishing that interreflections and surface gloss influence lightness perception greatly.

Figure 17 allows a comparison of observer performance with that of diagnostic image statistics. We note that both observers and statistics cannot predict the ground truth perfectly, but both of them make similar mistakes. The correlation coefficient r^2 ranges from 0.6 to 0.78. This agreement is surprising since the statistics have no clue as to how observers estimate reflectance properties. The learning technique that employs statistics as features was trained to predict the physical diffuse reflectance, not human performance. These findings suggest that image

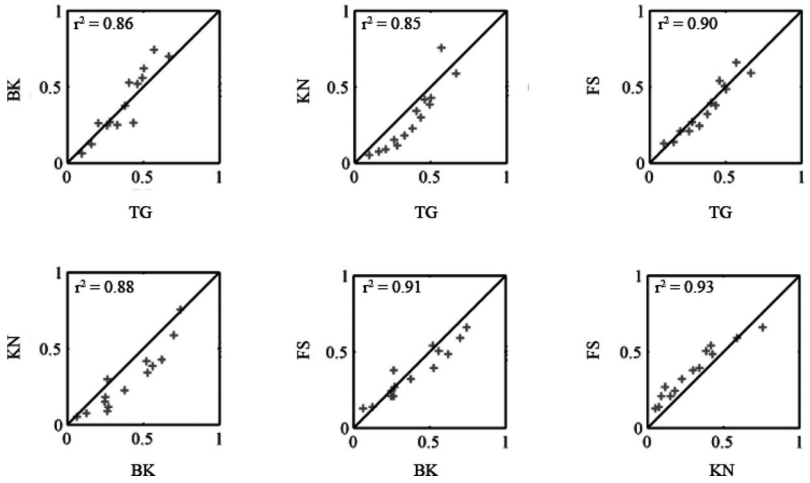


Fig. 13. Observers tend to agree with one another. Perceived reflectance ratings for every pair of observers from Fig. 12 are plotted here. If all observers behaved in the same way, all data points would lie on the black lines with slope=1; r^2 values indicate that there is a great deal of agreement among observers.

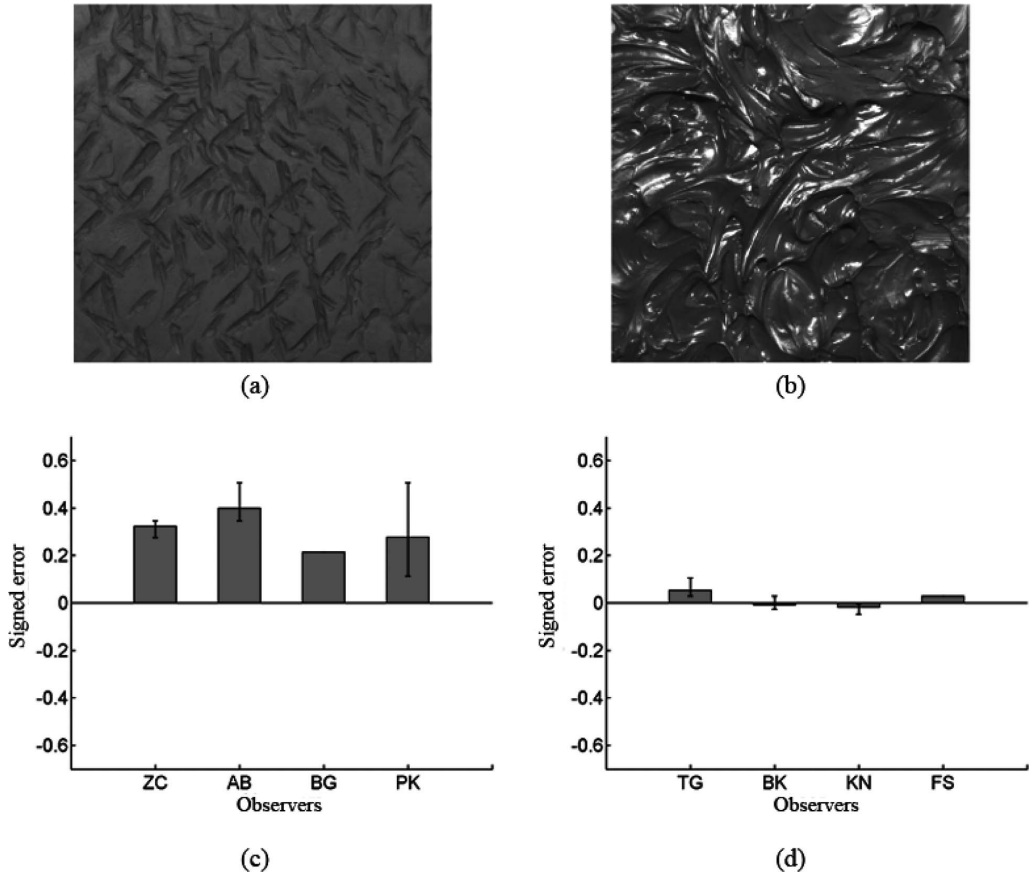


Fig. 14. Surfaces in (a) and (b) have similar diffuse reflectance values (0.085 and 0.092), respectively but dissimilar specular components and surface shapes. (c), (d) Errors in human judgments for the two surfaces seem to vary with the change in specular reflectance and mesostructure. Error bars indicate maximum and minimum errors. The data and photographs in this figure pertain to the overhead fluorescent lighting condition.

statistics of the kind we have considered must capture perceptually relevant information.

4. EXPERIMENT II

Given the strong correlation between perceived reflectance and the informative image statistics, we posed the following question: How do changes in the histogram statistics of an image affect reflectance perception? We know that as the physical reflectance of a surface changes, the histogram statistics change accordingly. If instead we manipulate the statistics of a given image, what happens? As discussed earlier, most of our statistics, moments, and percentiles of luminance and filtered outputs are corre-

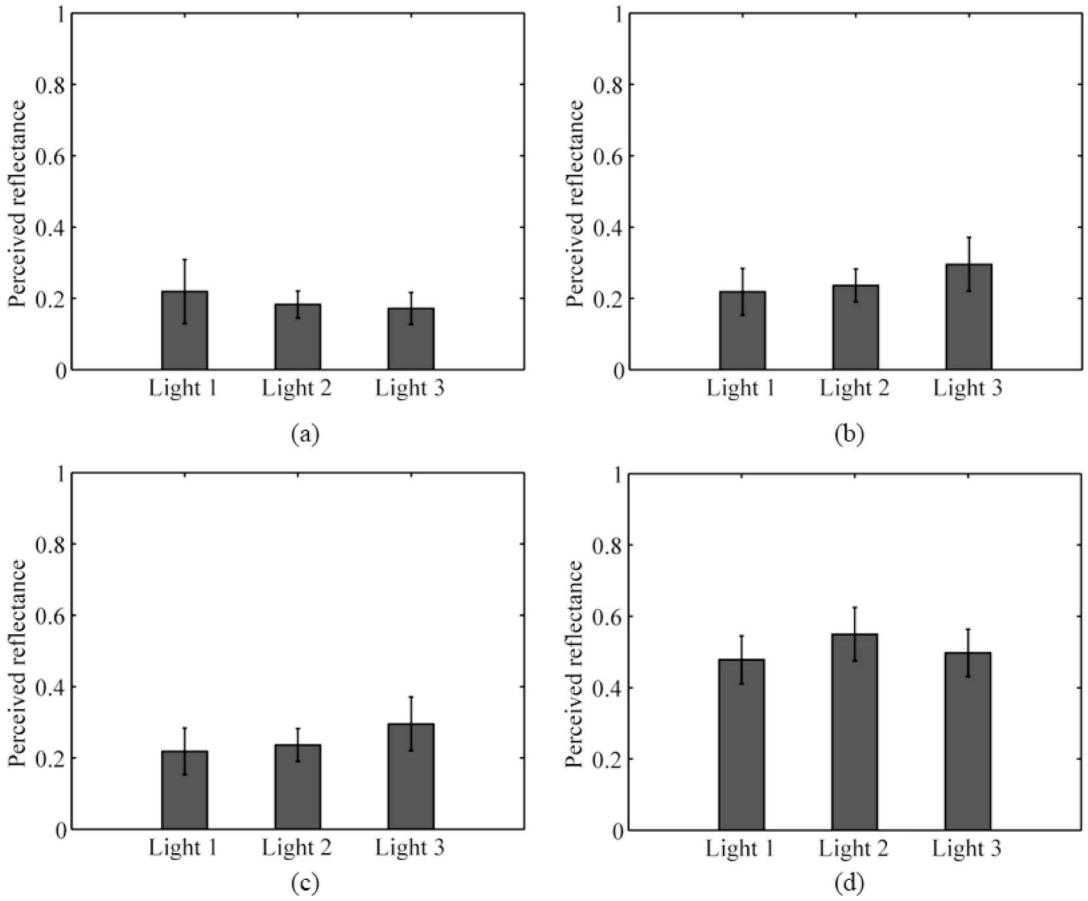


Fig. 15. Effect of lighting conditions. Light 1 is the overhead fluorescent light, light 2 is the halogen spotlight, and light 3 is the diffuse halogen source. (a)–(d) show the averaged observer ratings for four surfaces. Error bars indicate 95% confidence intervals.

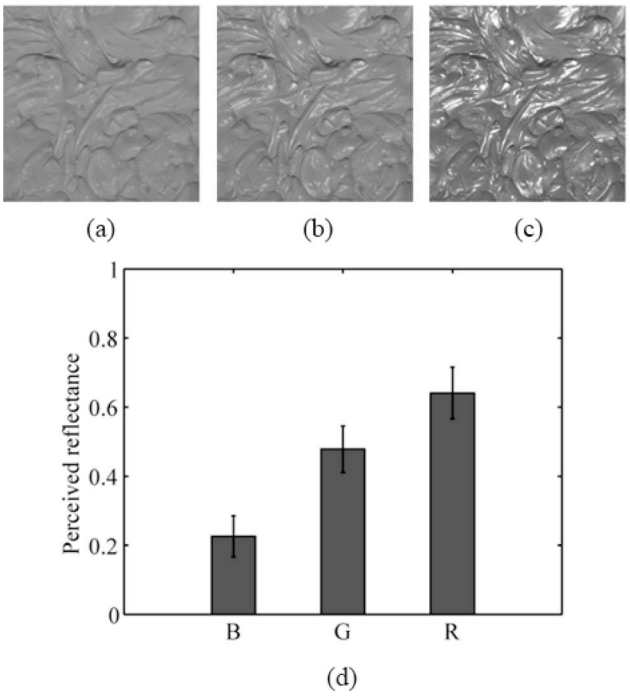


Fig. 16. Observers do not rate individual color channels in the same way. (a) Red channel, (b) green channel, and (c) B channel of an orange surface and the respective averaged observer ratings. (d) Even with the same mean image luminance, illumination conditions, and surface geometry, observers can extract useful information from the image to discern diffuse reflectance.

lated with one another in our data set. It is difficult to manipulate the statistics independently of one another by applying monotonic transformations on the images. In our previous study, we manipulated only the luminance histogram [2]. In the present work, we manipulate the histogram statistics all at once. As the luminance and subband statistics are correlated, an iterative technique such as the Heeger–Bergen texture synthesis algorithm, is required to simultaneously constrain both kinds of statistics [43]. The Heeger–Bergen algorithm iteratively matches the pixel histogram and the histogram of wavelet coefficients of a source image texture to a target image texture.

The algorithm converges in a few iterations to an image that has nearly the same pixel and wavelet coefficient histograms as the target image. The Heeger–Bergen algorithm may be applied in our case in the following manner: Let an image of a black surface be the source, and an image of a white surface be the target. The result of running the Heeger–Bergen algorithm on the black surface will be an image of a surface that has the same histogram statistics as the white surface. If histogram statistics capture anything of perceptual relevance, the resulting image should look lighter than the original black surface. We found that applying the Heeger–Bergen technique directly to our images in the way just described resulted in visible image artifacts.

To remove the artifacts, we introduce the following

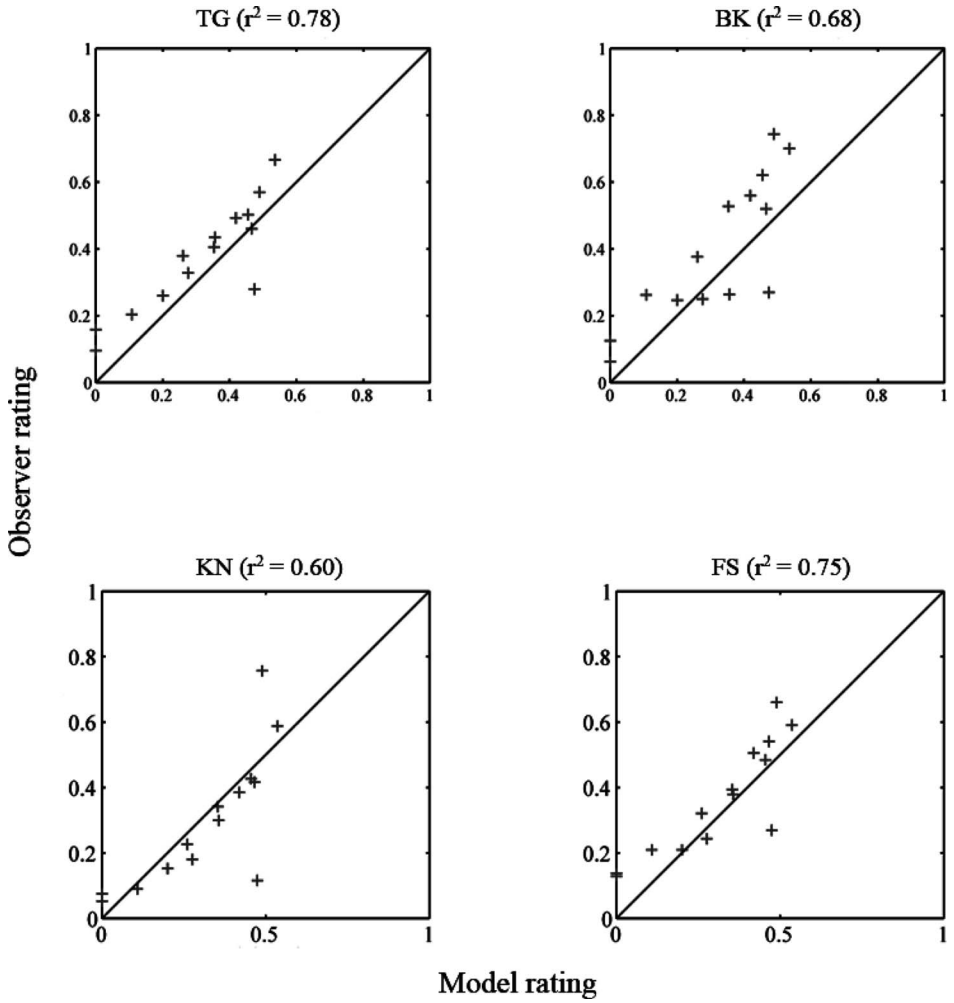


Fig. 17. Agreement between observers and the regression technique (the model) is fairly high. The r^2 values here are somewhat lower than the agreement between observers but are not very different (Fig. 13).

modification to the Heeger–Bergen algorithm: Instead of matching histograms of the filter subbands directly, we match histograms of activity maps instead. Activity maps are obtained by blurring the absolute value of filter subbands. A more detailed discussion of the modified Heeger–Bergen technique may be found in Appendix A. Figure 18(b) shows the effect of swapping the histogram statistics of a white and a black surface with the activity map Heeger–Bergen technique. We see that the perception of the surfaces is remarkably altered. The manipulated black surface looks much lighter, and the manipulated white surface looks much darker. To test this observation more conclusively, we ran an additional experiment with exactly the same conditions as before (experiment I). The only change was that we used histogram-manipulated images along with the original photographs of the surfaces. The same 12 observers who took part in experiment I also participated in experiment II. Observers were divided into three groups as before. Each group viewed a different set of surfaces. The partitioning of surfaces across groups was precisely the same as in experiment I. We ensured that for a given observer, stimuli presented during experiment I were not repeated in experiment II. Figure 19 plots observer data for this experiment. Observers consis-

Fig. 18. (a) Light surface and (b) dark surface. (c) Result of matching histogram statistics of (a) to those of (b) and (d) and vice versa.

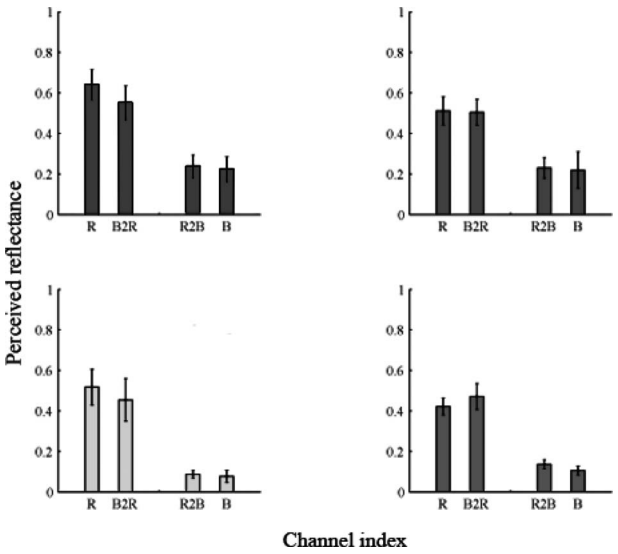


Fig. 19. Observer data for histogram-manipulated images (a) through (d) show four different surfaces in one group. The channels R and B refer to the red and blue color channels, respectively, of the color photographs of each surface. The R2B channel is the result of matching the histogram statistics of the R channel image to that of the B channel. B2R is the result of matching the histogram statistics of the B channel image to that of the R channel. The data are pooled across all observers in the group. All plots here are from the overhead fluorescent lighting condition. Error bars are 95% confidence intervals. We note that observers consistently rate R2B similar to B rather than to R and vice versa.

tently rate the manipulated images as more similar to the target image than the source image. In other words, images with similar histogram statistics are rated similarly. These results establish a two-way relationship between perceived reflectance and histogram statistics. Changing statistics affects reflectance perception, and knowing the statistics lets us predict the perception.

5. DISCUSSION

In this work, we considered a range of real-world surfaces. The interaction of light with our surfaces is more complex than with planar, Lambertian surfaces that have commonly been considered thus far. We found that human observers were not perfectly lightness constant, but their constancy improved as surfaces became more complex. In the absence of context and mean luminance information, the visual system relies on other cues present in the image. For the surfaces we considered, diffuse interreflections and specular highlights offer rich information about reflectance properties. We proposed quantitative image measurements that are correlated with the diffuse surface reflectance. Moment and percentile statistics of the luminance histogram and of filter output histograms were useful. A linear combination of these statistics was used to estimate the diffuse reflectance of our surfaces with an accuracy similar to that of human observers. Not only did the statistics achieve the same degree of success, they seemed to make the same mistakes as humans [10]. Moreover, manipulating these candidate statistics in an image of a surface altered the perception of lightness (experiment II). Therefore, it is conceivable that our diagnostic statistics are used as cues to lightness by the visual system, especially when other cues such as mean luminance or luminance ratios offer no information.

How do our results compare with other studies on lightness perception that have used complex real-world stimuli? Gilchrist and Jacobsen's black-and-white rooms were viewed in isolation, without a context with which to compare [4]. As the rooms were matte and mean luminance was accounted for, the only useful information available to the observers was the pattern of diffuse interreflections. It is likely that observers employed cues such as luminance histogram statistics to distinguish the rooms. Our findings confirm Nishida and Shinya's original observation that the luminance histogram influences lightness perception [15]. Rutherford and Brainard's finding that observers do not explicitly perceive the illuminant appears to favor our approach of using image cues to estimate lightness over inverse optics approaches [44]. Robilotto and Zaidi asked observers to judge the lightness of crumpled gray papers, some with patterns on them, in a 3-D setup under natural viewing conditions [19,20]. They found that observer performance could be explained in terms of low-level cues such as brightness and contrast. In their experiments, unlike ours, a surrounding context was always included. Therefore, cues from the surround compete with any information available from highlights and interreflections within the paper stimulus.

Several issues remain unanswered in the present work. We considered the restricted case of surfaces with spatially homogenous reflectance properties under simple artificial illumination. We know from our daily visual experience that we estimate lightness under more challenging illumination and surface conditions. Therefore, it is conceivable that there exist other cues or informative image measurements that apply to a less restricted setting than ours. Another issue that requires resolution is the interdependence of histogram statistics. We found a high degree of correlation in our set of statistics. Does this observation extend beyond our limited set of surfaces and illumination conditions? In fact, in the companion study, we manipulated the standard deviation and skewness of the luminance histogram independently of each other and found that skewness was a stronger cue [2]. Such a manipulation was not possible in the current work, given the number of dependent statistics we considered.

Our experiments with the application of the Heeger–Bergen texture synthesis technique to images of surfaces revealed a potential application—material transformation. One can imagine a Photoshop plugin that changes the appearance of a surface in an image region by manipulating the local image statistics. Our current modification to the Heeger–Bergen technique has had modest success in achieving this goal. Finally, a major question that remains unanswered is, What image statistics distinguish an image of a natural surface from an arbitrary image? Our statistics can predict perception only when we are given an image of a real-world surface. For any other image, the statistics and indeed even lightness judgments are not very meaningful. In our companion work, we addressed this question to some extent [2]. For pixel-scrambled images (which look like noise), when observers are asked to make lightness judgments, they

judge the overall brightness instead. Since luminance statistics are unaffected by pixel scrambling, they cannot explain perception. Clearly, in such cases, other perceptual mechanisms must be at work [45]. On phase scrambling our surface images, luminance and subband statistics can be retained, but spatial structure is lost. For such phase-scrambled images, lightness perception can be explained by our set of diagnostic image statistics, but not the perception of surface gloss. Gloss perception seems very sensitive to the interpretation of an image as a plausible real-world situation [18,23]. Further progress on these questions is required.

In the current work, as well as in the companion paper, we suggest a new perspective on natural image statistics. While variance and kurtosis, both even-order statistics, have been extensively studied, skewness, an odd statistic, has been largely neglected. We also demonstrated the utility of percentile statistics. These findings have implications for machine vision systems that perform material recognition as well as for psychophysicists and physiologists studying midlevel vision mechanisms and representations [46–48].

6. CONCLUSION

Humans use a variety of cues in judging the albedo of a surface. For an ideal planar surface, context is essential; when the surface is viewed in isolation, there is insufficient information to estimate albedo. However, surfaces in the real world are more complex. They have both diffuse and specular reflection, and they often have significant mesostructure. The resulting image has a visual texture that offers some cues about the surface's albedo. We have found that a variety of simple image statistics are correlated with albedo. Useful statistics, which may be computed on the luminance histogram or on subband histograms, include standard deviation, skewness, and percentiles. Linear combinations of these statistics allow a machine vision system, operating within the constrained world of single surfaces, to estimate the physical albedo reasonably well. The machine tends to make the same mistakes that humans do, suggesting that humans are using similar statistics in making their judgments. By manipulating these statistics, we can increase or decrease the apparent albedo of a given surface in a predictable way, giving further evidence that these image statistics play a major role in the surface judgments.

APPENDIX A: MODIFIED HEEGER–BERGEN ALGORITHM

In the Heeger–Bergen algorithm the luminance histogram as well the wavelet subband histograms of a target image are iteratively matched to those of a source image [43]. The histogram-matching procedure in the Heeger–Bergen algorithm involves applying a nonlinear pointwise gain; i.e., each pixel value in an image is mapped to a new value, independent of other pixels. A pointwise gain is not desirable in the overcomplete filter subbands because if the value of a pixel is manipulated independently of its neighbors, local distortions can occur in the final output. We propose the following solution: Instead of matching histograms of the source and target subbands directly, we will modify the target subband histograms via activity maps. An activity map is defined as the result obtained by taking the absolute value of a subband and then blurring it with a Gaussian kernel (Fig. 20).

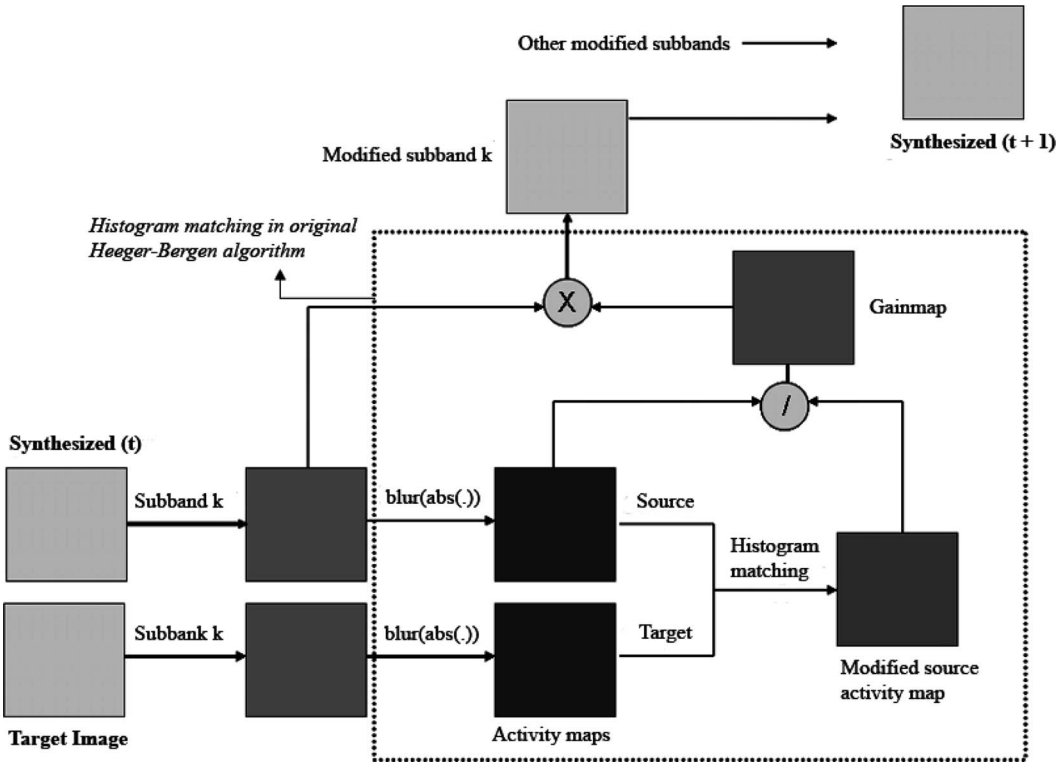


Fig. 20. Block diagram of the activity-map-based Heeger–Bergen technique.

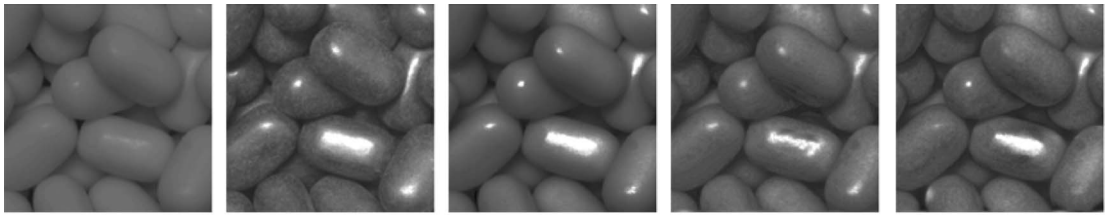


Fig. 21. Comparison of histogram-matching techniques: (a) source image, (b) target image, (c) luminance histogram matching, (d) Heeger–Bergen output, (e) activity-map-based Heeger–Bergen output.

The combination of absolute value and blurring transforms the subband image into a local energy map. When we match the histograms of the activity maps of the target and source images, then a pointwise gain is applied to the source activity map. As the activity map may be thought of as a local energy map, a pointwise gain on the source activity map is effectively a locally smooth gain on the original subband. Let the original source activity map be A_{orig} and the histogram matched source activity map be A_{modified}. Then the gain map G is calculated as

$$G = \frac{A_{modified}}{A_{orig}},$$

where G is multiplied to the original source subband to obtain the modified subband. Therefore, matching the histograms of the activity maps allows us to apply a spatially local gain, which results in fewer image artifacts and smoother-looking pictures. The local gain modifies the value at a pixel depending on the values of its neighbors; therefore the distortions introduced by histogram matching are reduced locally. Figure 21 shows the improvements obtained by using the modified Heeger–Bergen technique.

APPENDIX B: EFFECTS OF VARYING ILLUMINATION AND VIEWPOINT ON IMAGE STATISTICS AND PERCEPTION

In our image data, we used a fixed viewpoint and sampled three illumination directions with different light sources (Figs. 3 and 4). In our psychophysical experiments, we found that the perceived reflectance of a surface was not significantly affected by these changes in illumination conditions (Fig. 15). It may be argued that since we did not explore the space of viewing and illumination directions extensively, the results of Fig. 15 are somewhat preliminary. In order to address this concern, we performed some additional computational and psychophysical experiments.

Image data that are acquired for BRDF measurement are sampled densely along viewpoint and illumination directions. For example, the Columbia–Utrecht (CURET) database contains photographs of 61 real-world surfaces [42]. Each surface has been captured from 205 distinct combinations of viewing and illumination angles. Such a database would be useful for studying the effect of illumination and viewpoint variation on reflectance perception. However, there are two reasons why the CURET database is not ideal for our purposes. First, the resolution of all CURET images is 640×480 pixels, which includes pixels belonging to the dark background and the mounting equipment. So the effective pixel resolution of the material samples is much lower. It can be as low as 50×50 pixels at oblique viewing angles. Second, while the materials in the CURET database have wide-ranging reflectance properties, some of them have spatially varying reflectance functions (e.g., straw, peacock feather, and corn husk) or possess shallow mesostructure (e.g., leather, frosted glass, and corduroy). In the present study, we focused on materials with significant mesostructure that can be associated with a unique value of albedo. Thus, we are limited to only a quarter of the 61 CURET materials.

In spite of the concerns just outlined, the CURET images are very useful because they span a broad spectrum of illumination and viewing directions. We chose nine materials from the database (materials 10–12 and 16–21) that matched our criteria of spatially uniform albedo and nontrivial mesostructure. For each material, we selected 159 of the 205 photographs and cropped out the material sample to obtain 100×100 pixel patches. In the CURET images, there is an inherent trade-off between desired pixel resolution of a material and the number of viewpoints that can be used. In order to accommodate as many oblique views as possible without sacrificing too much resolution, we decided on 100×100 pixels [see Fig. 22(a)]. The CURET image data come calibrated to ensure linearity between pixels and radiance. For CURET materials, the ground truth for reflectance is known in terms of BRDF tables and Oren–Nayar model parameters [49]. We used the Oren–Nayar model parameterization, since the parameter ρ corresponds to albedo.

In addition to the CURET images, we acquired more image data for our own materials. The time and resource costs of reproducing something like the CURET database with our materials are immense. Therefore, we considered only two viewpoints and three illumination directions. The images were 512×512 pixel resolution [see Fig. 22(b)]. We chose nine materials from our collection (Fig. 3). Our selection included materials such as handmade stucco, modeling clay, and Tic Tacs. The images were acquired and linearized in the same manner as described before. The ground truth for albedo for these materials has already been measured earlier.

We analyzed the image statistics for both sets of images—CURET images and ours. Figure 23 plots the skewness of center–surround filtered images against the ground truth for albedo. Similar results were obtained for other moment and percentile statistics, both for luminance and filter outputs. We observed that the statistics

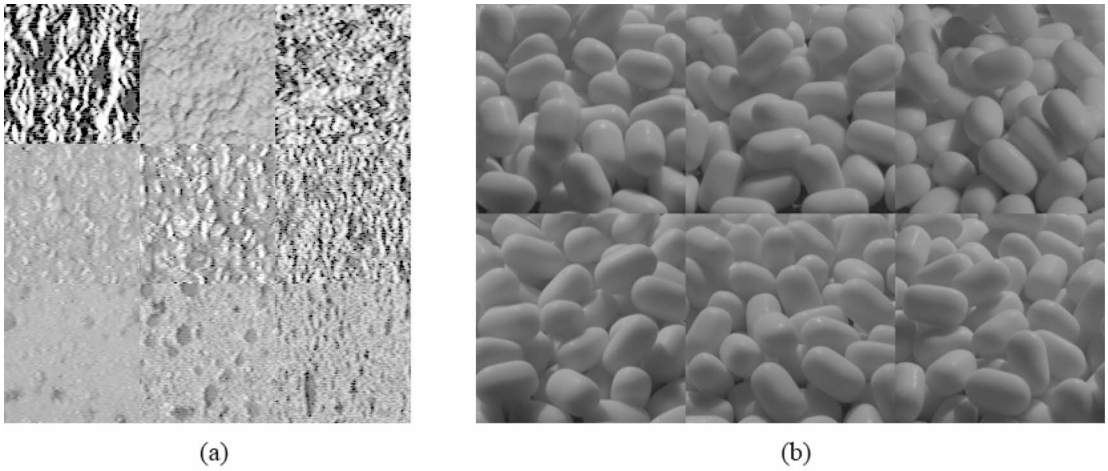


Fig. 22. (a) Some of the CURET images we used. The three rows correspond to materials 11 (plaster), 18 (rug), and 21 (sponge). Three different views are shown for each material. The images were multiplicatively normalized to have the same mean luminance. (b) One of the materials from our data set. The two rows correspond to two viewpoints and the columns to the three different illumination directions. All images were normalized to have the same mean.

varied enormously with viewpoint and illumination variation in the CURET data set but much less so for our images. These differences may be attributed to the fact that our images do not explore the space of viewpoint and illumination variation as aggressively as the CURET images. There is another, more subtle, reason for these results. Our images differ from the CURET images in a fundamental way. Our surfaces have deeper mesostructure than the CURET surfaces (Fig. 22). For shallower surfaces, the appearance changes more dramatically as viewpoint and illumination directions are varied. Consider the shadows in an image of a surface. For a shallow surface, as the illumination (or viewpoint) becomes grazing, dark shadows appear in the image. For a 3-D surface with deeper mesostructure, dark shadows are always present because no matter what the illumination (or viewpoint), self-shadowing occurs due to higher surface relief. Indeed, when one considers the CURET images of Fig. 22(a), particularly those in the first row, it is hard to believe that the same surface can be made to look so different by moving the lights and camera around. We conducted psychophysical experiments to study these effects further.

We asked three observers (two naïve subjects and one of the authors) to rate the lightness of mean luminance normalized images (both CURET and ours) in exactly the same experimental setup as before. For each of the nine CURET materials, six representative images were chosen [Fig. 22(a) show three such images for three materials]. For our materials, all six views were used for the nine materials. Images were viewed against a midgray background, one at a time, and observers indicated the Munsell chip that was closest in reflectance to the sample image on screen. Two repetitions were run for each image. Figure 24 plots the perceived diffuse reflectance against true reflectance for both sets of images. Observers were fairly accurate in their lightness judgments for our images but performed poorly on the CURET images. These results confirm the informal observations we made earlier

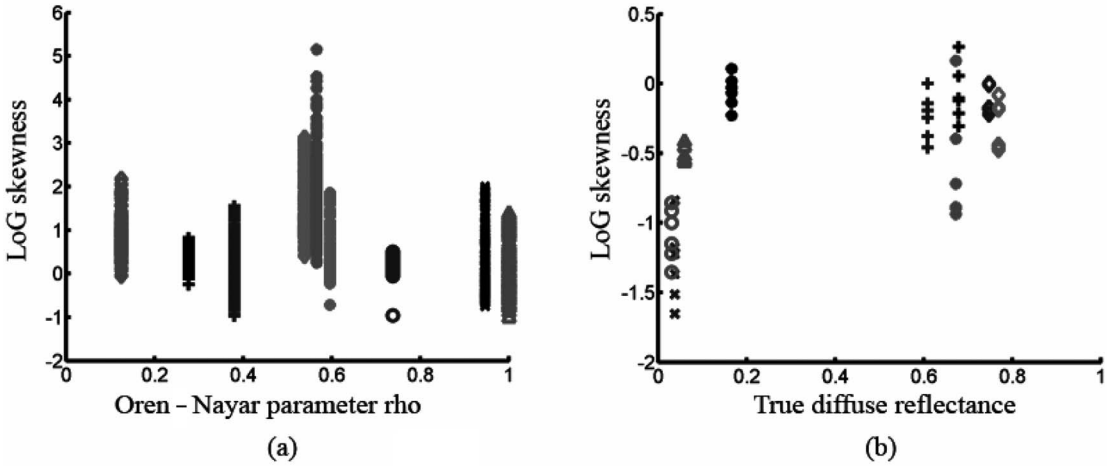


Fig. 23. Effect of varying illumination and viewpoint on image statistics. The skewness of center–surround filtered images is plotted against the ground truth for diffuse reflectance. A Laplacian of a Gaussian filter was used ($\sigma=0.5$, size 5×5 pixels). (a) Skewness statistics is plotted for all 159 images of the 9 CURET materials against the Oren–Nayar model parameter "rho" [49]. Note the vertical smear at each value of rho. (b) All six images of the nine materials in our data set were used. The x axis is the ground truth for diffuse reflectance for the materials.

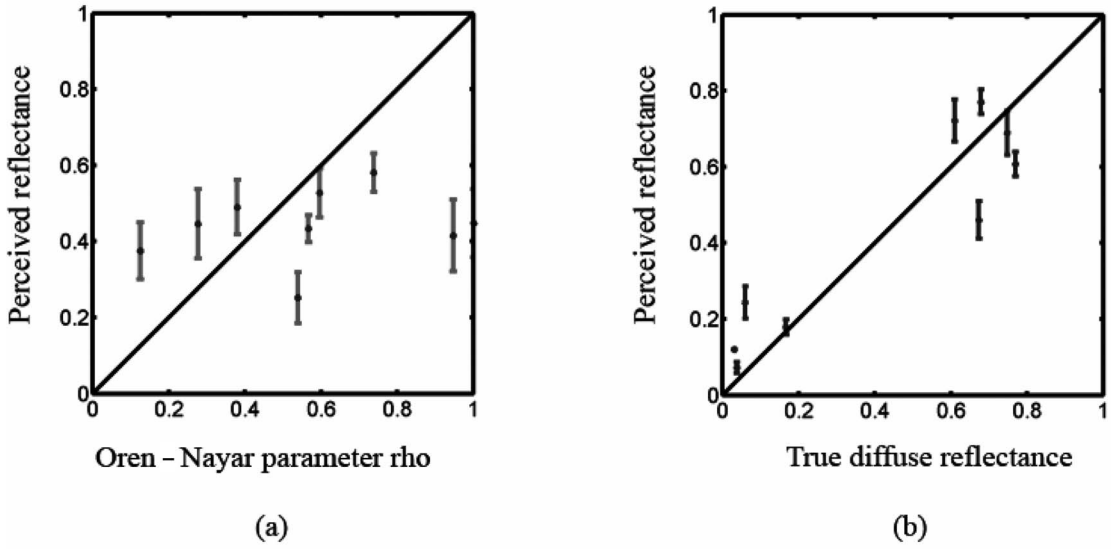


Fig. 24. Perceived reflectance for an observer is plotted against the ground truth. For each material, observer ratings were pooled across all viewing and illumination directions. Error bars are 95% confidence intervals. (a) CURET images: There is no linear relationship between perceived and true reflectance ($p > 0.05$). (b) Our images: Observer data can be explained by a linear model ($p < 0.05$, $r^2 = 0.85$). Similar trends were obtained for other observers.

pertaining to Fig. 22(a). The poorer resolution of the CURET images was an issue. Observers reported that some CURET images appeared to have wallpaperlike textures rather than 3-D surfaces of single albedo. Observers also complained about the grazing-angle CURET pictures that are formed of alternating stripes of dark shadows and bright highlights. Even when these pictures are seen as 3-D surfaces, it is hard to gauge the true albedo.

How well do image statistics account for these perceptual results? Figure 25 plots the perceived reflectance for the observer of Fig. 24 against the skewness of the image after center–surround filtering. Again, observer data for our images can be explained reasonably well by our chosen statistics. For the CURET images also, we see evidence for a correlation between perceived reflectance and our statistics. This is encouraging, given the dismal correlations in Figs. 23(a) and 24(a).

Taken together, these results lead us to conclude that the relationship among image statistics, physical reflectance, and reflectance perception holds for the images we acquired but breaks down for the CURET data set. As mentioned earlier, our images differ from the CURET images in many ways. It is reasonable to assume that these differences in computational and psychophysical results are correlated with the nature of the data sets. A more quantitative investigation of how factors such as surface

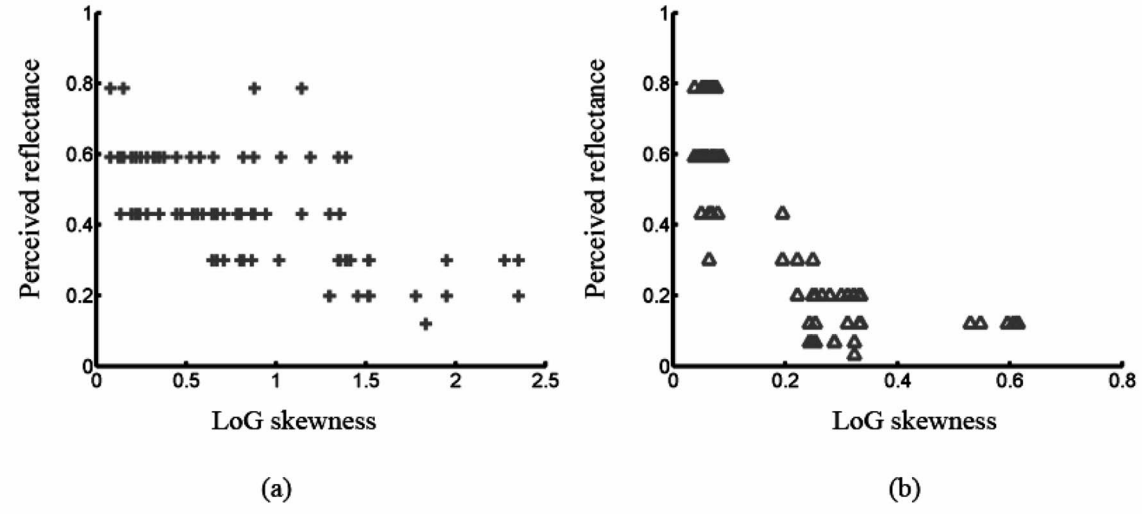


Fig. 25. Perceived reflectance of an image of a surface is plotted against the skewness of the center–surround filter output. A Laplacian of a Gaussian filter was used ($\sigma = 0.5$, size 5×5 pixels). (a) CURET images and (b) our images. A linear relationship can be observed in both plots. For (c), $p < 0.05$, $r^2 = 0.42$ and for (d), $p < 0.05$, $r^2 = 0.64$.

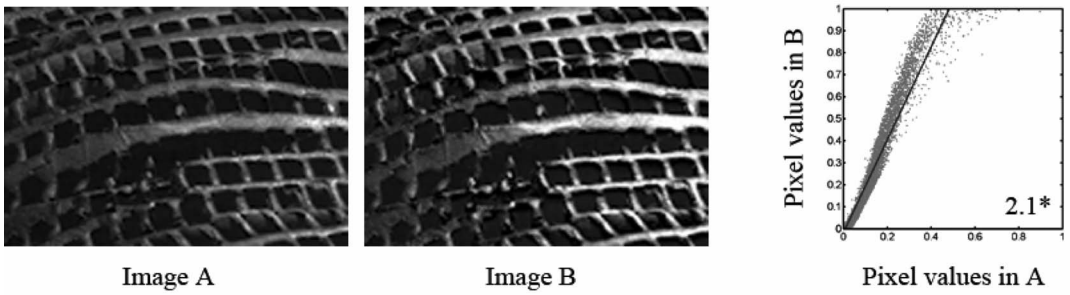


Fig. 26. Verifying linearity correction. Images A and B are the linear outputs of the dcraw program for two exposures of the same scene [37]. The original exposures were recorded in the Canon RAW format (CRW). The exposure time for B was twice that of A. All other camera parameters were the same for A and B. The pixel values of the linearized images are plotted against each other. If the combination of internal camera processing followed by dcraw was perfectly linear, all data points would lie along a line with slope=2. The actual fit is quite good: slope=2.1055±0.0084 ($p<0.05$, $r^2=0.94$).

structure, viewing and illumination conditions, and material properties influence image statistics (and reflectance perception) remains a direction for future research.

APPENDIX C: LINEARIZATION OF PIXEL VALUES

To ensure a linear relationship between pixel values and measured luminance, we calibrated our raw image data using dcraw software [37]; dcraw is an open-source program that decodes the raw images acquired by a camera in its native, manufacturer-specific format. The advantage of taking photographs in the raw format is that raw images have the most information and are processed the least by the camera. Almost all digital cameras can be set to produce a JPEG output instead of a raw output. Although the JPEG format is convenient, it is a lossy format and valuable data are lost as a result of the JPEG compression.

We acquired our image data in the Canon RAW format (CRW). The CRW images files were then converted by the dcraw program to 16-bit PSD (Adobe Photoshop) file format. In order to verify the correctness of the calibration procedure, we used the following methodology. Since we did not take photometric measurements at each point of our scenes, the precise luminance measurements that correspond to the pixels in an image are unknown. However, we did capture multiple exposure shots of the same scene. We used a fixed-focal-length (50 mm) lens. We applied manual settings for focus, aperture, and white balance and varied the shutter speeds in intervals of $1/2f$-stops and acquired between 9 and 12 consecutive exposures for each scene. We made use of the following principle: If the shutter is open for twice as long, the amount of light captured is twice as much. If the dcraw outputs are truly linear, the pixels values for two images of the same scene should be related to each other by exactly the same ratio as the lengths of their exposure times. Figure 26 illustrates this reasoning for a pair of example images. If the combination of internal camera processing followed by dcraw was perfectly linear, the data in Fig. 26 would fit the linear model perfectly with a slope equal to 2. The actual fit is quite good: slope=2.1055±0.0084 ($p<0.05$, $r^2=0.94$). We repeated this analysis for other image pairs for each of our scenes and verified that the accuracy of dcraw linearization was similar to Fig. 26.

ACKNOWLEDGMENTS

We thank our reviewers for their insightful comments and suggestions. We also thank R. W. Fleming and R. Rosenholtz for useful discussions. This research was supproted by a grant from the NTT-MIT Research Collaboration. It was also supported by the National Science Foundation under grant no. 03545805 to E. Adelson.

REFERENCES

1. A. Gelb, "Die 'Farbenkonstanz' dez Sehdinge [Color constancy of visual objects]," in *Handbuchder normalen und pathologischen Psychologie*, W. A. von Bethe, ed. (Springer, 1929), pp. 594–678.
2. I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," Nature **447**, 206–209 (2007).
3. A. L. Gilchrist, "The perception of surface blacks and whites," Sci. Am. **240**, 112–124 (1979).
4. A. L. Gilchrist and A. Jacobsen, "Perception of lightness and illumination in a world of one reflectance," Perception **13**, 5–19 (1984).
5. E. H. Land and J. J. McCann, "Lightness and retinex theory," J. Opt. Soc. Am. **61**, 1–11 (1971).
6. A. L. Gilchrist, "Perceived lightness depends on perceived spatial arrangement," Science **195**, 185–187 (1977).
7. E. H. Adelson, "Perceptual organization and judgment of brightness," Science **262**, 2042–2044 (1993).
8. M. D'Zmura and G. Iverson, "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces," J. Opt. Soc. Am. A **10**, 2148–2165 (1993).
9. D. H. Brainard, "Color constancy in the nearly natural image. 2. Achromatic loci," J. Opt. Soc. Am. A **15**, 307–325 (1998).
10. A. L. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou, "An anchoring theory of lightness perception," Psychol. Rev. **106**, 795–834 (1999).

11. M. G. Bloj, D. Kersten, and A. C. Hurlbert, "Perception of three-dimensional shape influences color perception through mutual illumination," Nature **402**, 877–879 (1999).
12. H. Boyaci, L. T. Maloney, and S. Hersh, "The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes," J. Vision **3**, 541–553 (2003).
13. D. H. Brainard, J. M. Kraft, and P. Longere, "Color constancy: developing empirical tests of computational models," in *Color Perception: From Light to Object*, R. Mausfeld and D. Heyer, eds. (Oxford U. Press, 2003), pp. 307–334.
14. L. T. Maloney and J. N. Yang, "The illumination estimation hypothesis and surface color perception," in *Color Vision: Connecting the Mind to the Physical World*, R. Mausfeld and D. Heyer, eds. (Oxford U. Press, 2003), pp. 335–358.
15. S. Nishida and M. Shinya, "Use of image-based information in judgments of surface reflectance properties," J. Opt. Soc. Am. A **15**, 2951–2965 (1998).
16. R. O. Dror, E. H. Adelson, and A. S. Willsky, "Recognition of surface reflectance properties from a single image under unknown real-world illumination," in *Proceedings of the IEEE Workshop on Identifying Objects across Variation in Lighting: Psychophysics and Computation* (IEEE, 2001), available at http://web.mit.edu/persci/people/adelson/pub_pdfs/dror_cvpr01_goem.pdf.
17. R. O. Dror, "Surface reflectance recognition and real-world illumination statistics," Ph.D. dissertation (Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2002).
18. R. W. Fleming, R. O. Dror, and E. H. Adelson, "Real-world illumination and the perception of surface reflectance properties," J. Vision **3**, 347–368 (2003).
19. R. Robilotto and Q. Zaidi, "Limits of lightness identification of real objects under natural viewing conditions," J. Vision **4**, 779–797 (2004).
20. R. Robilotto and Q. Zaidi, "Lightness identification of patterned three-dimensional, real objects," J. Vision **6**, 18–36 (2006).
21. J. Beck and S. Prazdny, "Highlights and the perception of glossiness," Percept. Psychophys. **30**, 401–410 (1981).
22. F. Pellacini, J. A. Ferwerda, and D. P. Greenberg, "Towards a psychophysically-based light reflection model for image synthesis," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, J. R. Brown and K. Akeley, eds. (ACM Press, 2000), pp. 55–64.
23. J. T. Todd, J. F. Norman, and E. Mingolla, "Lightness constancy in the presence of specular highlights," Psychol. Sci. **15**, 33–39 (2004).
24. B.-T. Phong, "Illumination for computer generated pictures," Commun. ACM **18**, 311–317 (1975).
25. Y. Sato, M. D. Wheeler, and K. Ikeuchi, "Object shape and reflectance modeling from observation," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, G. O. Owen, T. Whitted, and B. Mones-Hattal, eds. (ACM Press, 1997), pp. 379–387.
26. Y. Yu and J. Malik, "Recovering photometric properties of architectural scenes from photographs," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, S. Cunningham, W. Bransford, and M. F. Cohen, eds. (ACM Press, 1998), pp. 207–217.
27. S. R. Marschner, S. H. Westin, E. P. F. Lafortune, K. E. Torrance, and D. P. Green, "Image-based BRDF measurement including human skin," in *Proceedings of the 10th Eurographics Workshop on Rendering*, D. Lischinski and G. W. Larson, eds. (Springer, 1999), pp. 139–152.
28. Y. Yu, P. Debevec, J. Malik, and T. Hawkins, "Inverse global illumination: recovering reflectance models of real scenes from photographs," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, W. Waggenspack, ed. (ACM Press, 1999), pp. 215–224.
29. S. Boivin and A. Gagalowicz, "Image based rendering of diffuse, specular and glossy surfaces from a single image," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, L. Pocock, ed. (ACM Press, 2001), pp. 107–116.
30. S. Tominaga and N. Tanaka, "Estimating reflection parameters from a single color image," IEEE Comput. Graphics Appl. **20**, 58–66 (2000).
31. K. Nishino, Z. Zhang, and K. Ikeuchi, "Determining reflectance parameters and illumination distributions from a sparse set of images for view-dependent image synthesis," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2001), pp. 599–601.
32. R. Ramamoorthi and P. Hanrahan, "A signal processing framework for inverse rendering," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, L. Pocock, ed. (ACM Press, 2001), pp. 117–128.
33. P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, J. R. Brown and K. Akeley, eds. (ACM Press, 2000), pp. 145–156.
34. P. Debevec, C. Tchou, A. Gardner, T. Hawkins, C. Poullis, J. Stumpfel, A. Jones, N. Yun, P. Einarsson, T. Lundgren, M. Fajardo, and P. Martinez, "Estimating surface reflectance properties of a complex scene under captured natural illumination," Tech. Rep. ICT-TR-06 (University of Southern California Institute for Creative Technologies Graphics Laboratory, 2004).
35. J. J. Koenderink, A. J. Van Doorn, K. J. Dana, and S. Nayar, "Bidirectional reflectance distribution function of thoroughly pitted surfaces," Int. J. Comput. Vis. **31**, 129–144 (1999).
36. J. J. Koenderink and A. J. van Doorn, "Illuminance texture due to surface mesostructure," J. Opt. Soc. Am. A **13**, 452–463 (1996).
37. D. Coffin, "Raw digital photo decoding in Linux," http://www.cybercom.net/~dcoffin/dcraw.
38. P. Debevec, C. Tchou, and T. Hawkins, "HDRShop: high dynamic range image processing and manipulation," http://www.hdrshop.com.
39. P. Burt and E. H. Adelson, "Laplacian pyramid as a compact image code," IEEE Trans. Commun. **31**, 532–540 (1983).
40. E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Proceedings of IEEE Conference on Image Processing* (IEEE, 1995), pp. 444–447.
41. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," http://www.csie.ntu.edu.tw/~cjlin/libsvm.
42. Columbia–Utrecht Reflectance and Texture (CURET) database, http://www.cs.columbia.edu/CAVE/curet.
43. D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, R. Cook, ed. (ACM Press, 1995), pp. 229–238.
44. M. D. Rutherford and D. H. Brainard, "Lightness constancy: a direct test of the illumination estimation hypothesis," Psychol. Sci. **13**, 142–149 (2002).
45. C. Chubb, M. S. Landy, and J. Econopouly, "A visual mechanism tuned to black," Vision Res. **44**, 3223–3232 (2004).
46. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature **381**, 607–609 (1996).
47. E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1216 (2001).
48. M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," Int. J. Comput. Vis. **62**, 61–81 (2005).
49. M. Oren and S. K. Nayar, "Generalization of the Lambertian model and implications for machine vision," Int. J. Comput. Vis. **14**, 227–251 (1995).

Vol 447|10 May 2007|doi:10.1038/nature05724

LETTERS

Image statistics and the perception of surface qualities

Isamu Motoyoshi[1], Shin'ya Nishida[1], Lavanya Sharan[2] & Edward H. Adelson[2]

The world is full of surfaces, and by looking at them we can judge their material qualities. Properties such as colour or glossiness can help us decide whether a pancake is cooked, or a patch of pavement is icy. Most studies of surface appearance have emphasized textureless matte surfaces[1–3], but real-world surfaces, which may have gloss and complex mesostructure, are now receiving increased attention[4–7]. Their appearance results from a complex interplay of illumination, reflectance and surface geometry, which are difficult to tease apart given an image. If there were simple image statistics that were diagnostic of surface properties it would be sensible to use them[8–11]. Here we show that the skewness of the luminance histogram and the skewness of sub-band filter outputs are correlated with surface gloss and inversely correlated with surface albedo (diffuse reflectance). We find evidence that human observers use skewness, or a similar measure of histogram asymmetry, in making judgements about surfaces. When the image of a surface has positively skewed statistics, it tends to appear darker and glossier than a similar surface with lower skewness, and this is true whether the skewness is inherent to the original image or is introduced by digital manipulation. We also find a visual aftereffect based on skewness: adaptation to patterns with skewed statistics can alter the apparent lightness and glossiness of surfaces that are subsequently viewed. We suggest that there are neural mechanisms sensitive to skewed statistics, and that their outputs can be used in estimating surface properties.

Figure 1 shows two renderings of a three-dimensional model of Michelangelo's sculpture of St Matthew[12]. The version on the left appears darker and glossier than the one on the right. This is true even though the two images have been scaled to have the same mean luminance. We are unaware of any theories that will predict the changes in lightness or gloss that we observe.

The image of a surface arises from the combination of the surface geometry, the surrounding illumination, and the surface optics. Each of these components can be complex (for example, the reflectance at each point is characterized by a four-dimensional function known as the bidirectional reflectance distribution function[13]). Each is typically unknown, and estimating any one using 'inverse optics' requires knowing the others. To bypass this problem, we have looked for simple statistical image measurements that can provide information that is useful even if not complete. Any two-dimensional image measurements that are statistically related to properties of the three-dimensional scene are potentially useful[8–11].

We made a set of patches of stucco-like material. The values of albedo and glossiness were uniform within each patch, but they were varied systematically from one patch to another by changing paint pigmentation and acrylic media coating, respectively. We photographed these objects, linearized the pixel values and normalized the mean luminance by multiplicative scaling. We found that changes in albedo and glossiness were accompanied by characteristic changes in the luminance histogram. Consider the two stucco patches of

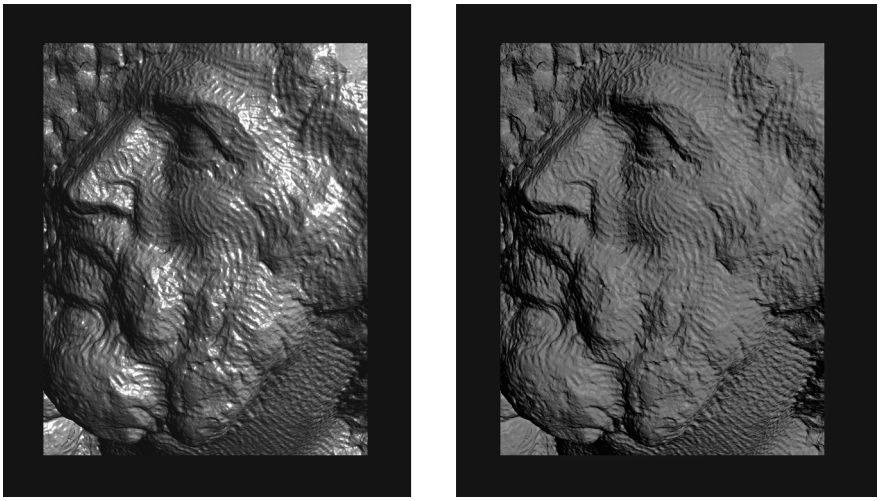


Figure 1 | These two synthetic images of Michelangelo's St Matthew sculpture have the same mean luminance. The one on the left looks darker and glossier than the one on the right.

[1]Human and Information Science Lab, NTT Communication Science Labs, Nippon Telegraph and Telephone Corporation, 3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan. [2]Department of Brain and Cognitive Sciences and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 43 Vassar Street, 46-4115, Cambridge, Massachusetts 02139, USA.

Fig. 2a. In comparison with a light matte surface (left), a dark glossy surface (right) has a long positive tail. In general, as the albedo of glossy surfaces is decreased, or as the glossiness is increased regardless of the albedo, the histogram's skewness tends to increase (Fig. 2b; black circles). These changes make sense given the influence of specular and diffuse reflectance on the appearance of specular highlights. Highlights are stronger and sharper on glossy surfaces, and they have higher contrast when viewed on darker surfaces, because they are seen against a body surface that has a lower luminance.

Having observed this physical relationship, we next looked for a corresponding psychophysical relationship. We showed these stucco images, one by one, to human observers, presenting them against a dark background on a monitor at constant mean luminance, and asked the observers to rate the lightness (perceived diffuse reflectance) or glossiness of each surface. The judgments were well correlated with the corresponding physical properties, as shown in Fig. 2b (red circles). Both the lightness and glossiness ratings were also well correlated with the skewness of the luminance histogram (Fig. 2c) to a degree comparable with, or even higher than, the correlations with corresponding physical properties ($r = -0.87$ for correlation with skewness of lightness ratings, and 0.89 for glossiness ratings, respectively).

We next chose a set of images of three materials (stucco, black cotton fabric and crumpled white paper, all of which were surfaces of uniform albedo and glossiness) and used a lookup table to force the luminance histograms to have specific skewness values. As expected, the lightness rating showed a strong negative dependency on skewness, whereas the glossiness rating showed a strong positive dependency. This was true for each image class (Fig. 2d). Further tests of a wide variety of materials gave similar results, described in Supplementary Data A.

In addition to the effects of skewness, we found a minor effect of the standard deviation of the luminance histogram on both lightness and glossiness. The mean luminance had a significant effect on lightness[1–3], but not on glossiness. We found little, if any, effect of kurtosis (Supplementary Data B).

The above results indicate that skewness or a similar measure of histogram asymmetry is useful in estimating surface qualities, and that humans may indeed use it. How might such statistics be computed at the neural level? The early stages of vision are dominated by neurons that represent luminance variation in certain sub-bands of spatial frequency. These cells do not have direct access to raw luminance, but there is a strong correlation between sub-band skewness and luminance skewness for the uniform albedo surfaces we used ($r \geq 0.86$ for sub-bands at spatial frequencies ranging from 4 to 64 cycles per image, obtained with two-octave gaussian bandpass filters). We note that sub-band statistics also have certain advantages over luminance statistics, because they reflect spatial image structure whereas luminance statistics do not (see also Supplementary Data D).

Skewness is a measure of the asymmetry of a distribution; it indicates the balance between the positive and negative tails. Various definitions have been used[14], the most popular being based on the third standardized moment. If X is a random variable with zero mean and unit variance, then skewness is the expected value of X^3. If X is the output of an array of neurons that act as sub-band filters, then a neural implementation needs normalization, cubing, and summation over a region; these are easy to implement in neural hardware. A bandpass neuron's output already has zero mean, and local gain control[15] will tend to normalize the response variance over a given region. Cubing and summing over a region are straightforward.

In more concrete terms, we suggest the flow diagram of Fig. 3. As input, we use an image constructed from the two St Matthew images. The image is filtered with on-centre and off-centre receptive fields and split into separate (all positive) streams. The on-centre and off-centre responses each pass through an accelerating nonlinearity. These responses are summed over a region, and these summed responses are subtracted. The difference signal is an estimate of local

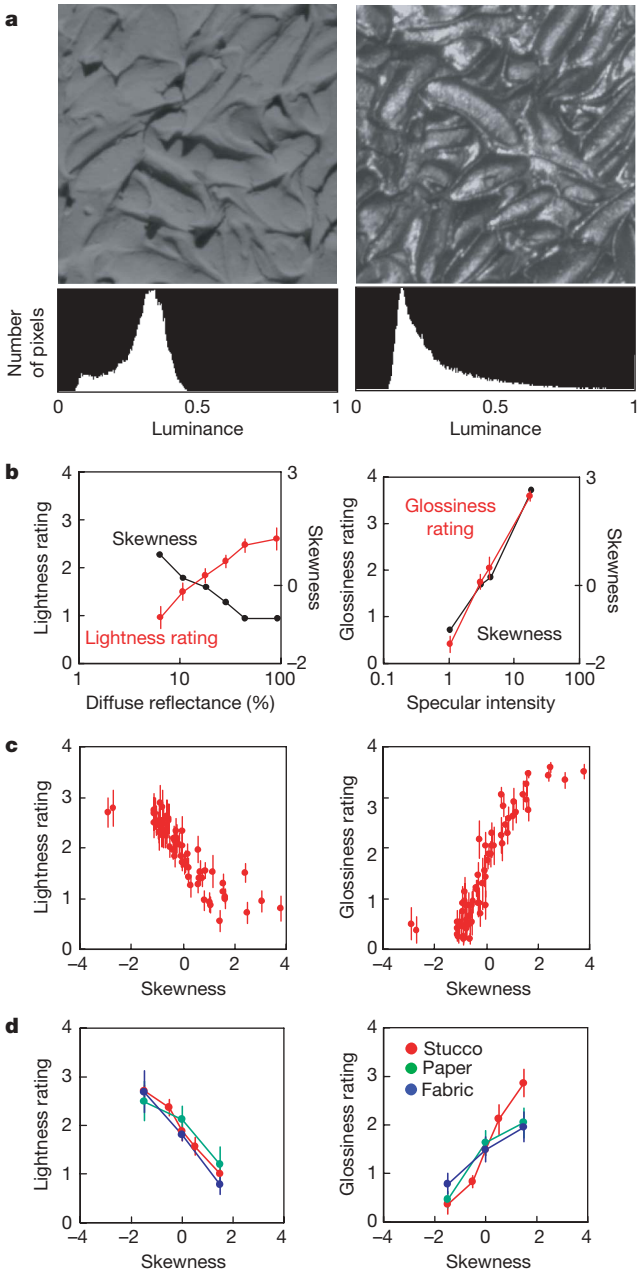


Figure 2 | Perceived lightness and glossiness may be based on the skewness of the luminance histograms. a, The stucco-like surface on the right looks darker and glossier than one on the left even though the mean luminance of both images is equal. We note that the luminance histogram is negatively skewed for the left image (skewness = −1.34), and positively skewed (skewness = 2.40) for the right image. **b**, In the left panel, as the diffuse reflectance increases, the lightness ratings given by human observers increase (red circles, *y*-axis scale on the left side) and the skewness of the image histogram decreases (black circles, *y*-axis scale on the right side). The data were obtained with medium-glossy surfaces. In the right panel, we see that as the specular reflectance increases, both the rated glossiness (red circles) and the histogram skewness (black circles) increase. The data were obtained with dark-grey surfaces. **c**, The rated lightness (left panel) and glossiness (right panel) of the 63 surface images of varying diffuse and specular reflectance under different illumination conditions. The human ratings correlate well with the skewness of the image histogram. Error bars represent ± 1 s.e.m across six observers. **d**, The effect of histogram manipulation on perceived lightness and glossiness. For images of stucco, crumpled paper, and fabric, the luminance histogram was matched to a positively or negatively skewed beta distribution. The rated lightness (left panel) and glossiness (right panel) varied according to the final skewness of the manipulated image (mean = 8.2 cd m^{-2}, s.d./mean = 0.1, See Supplementary Data B for data obtained with other parameters). Error bars represent ± 1 s.e.m. across seven observers.

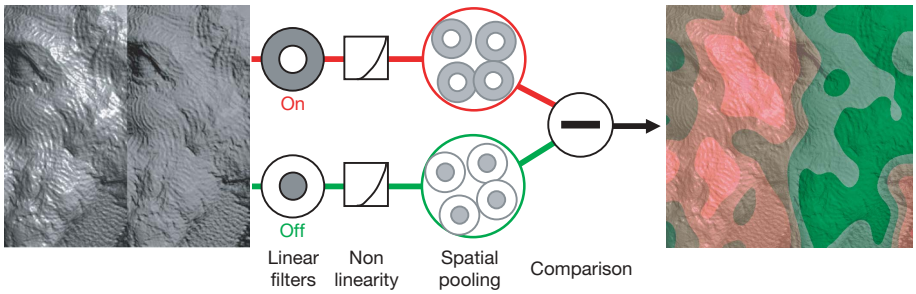


Figure 3 | A proposed neural mechanism for encoding the sub-band skewness by early visual units. The image is analysed by on-centre and off-centre filters followed by an accelerating nonlinearity (square, cube and so on). The outputs are then separately pooled over space. The difference between the pooled outputs of the on- and off-centre channels provides an estimate of the sub-band skewness.

skewness. See Supplementary Discussion for a more formal description of the model.

Could this computation be supported with known physiology? The on-centre and off-centre cells of the brain's lateral geniculate nucleus (LGN) would work as the initial stage, as would even-symmetric ('bar detector') simple cells. Next we require a cortical cell that pools and compares the outputs of a set of such cells after a nonlinearity. The classic complex cell will not suffice, because it is insensitive to contrast sign. Our putative skewness cells would be selective for contrast sign, but not for position and not necessarily for orientation. Such cells would be excited by bright (or dark) dots or lines anywhere within their receptive fields, but not by ones of the opposite sign. There are various reports of cells in areas V1 and V2 of the brain that are selective for contrast sign[16–20], and these could participate in the processing chain we are proposing.

The notion of 'skewness detectors' suggested a psychophysical experiment. Suppose we adapt to a pattern with positive skewness, and thereby shift the balance of sensitivities in the positive and negative skewness mechanisms. Would that shift our judgements of lightness and glossiness of a surface subsequently viewed?

We had subjects adapt to the patterns shown in Fig. 4a, which consisted of quasi-randomly placed blurred spots. Subjects fixated at a position between the two images. The adaptor with bright spots had positive skewness and the one with dark spots had negative skewness. After adaptation, two images of the same stucco surface that differed only in the sign of their skewness values were presented, side by side, as shown in Fig. 4b, and we asked subjects to judge their relative lightness and glossiness. Both judgments were shifted in the expected direction, as shown in Fig. 4d. We also used stucco images with positive or negative skewness as adapters, as in Fig. 4c, and found a similar shift. Thus, the effect is similar whether the adapting stimulus looks like a surface or merely a set of random spots.

We determined that the skewness adaptation did not affect the apparent brightness of the uniform grey test field of various luminance values. This ruled out the possibility that the lightness after-effect was due to a simple change in the luminance transducer function of the visual system.

We tested the inter-ocular transfer of the after-effects and found that the after-effect caused by the different-eye adaptor was 76% ± 6.1% (95% confidence interval) of that by the same-eye adaptor (estimated from the averaged results of three observers). Such incomplete transfer points to a partial involvement of monocular

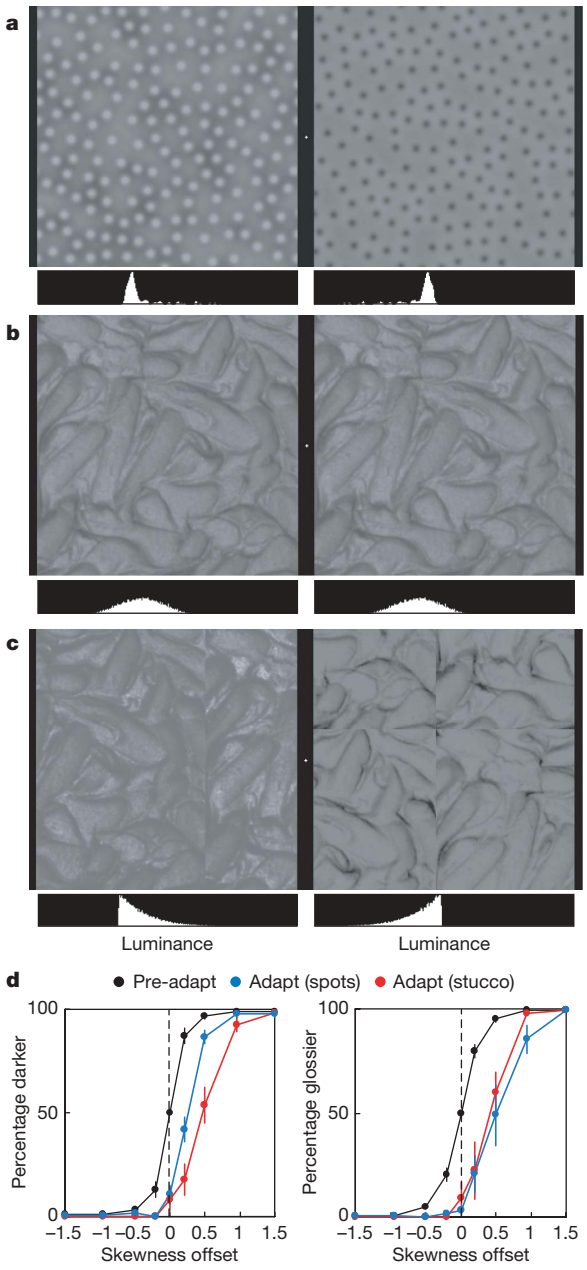


Figure 4 | After-effects of perceived lightness and glossiness. Observers adapted to the artificial textures shown in **a**. These textures consist of on-centre (left) and off-centre (right) difference-of-gaussian elements. After the prolonged observation of images in **a** observers were shown a pair of test images (**b**). Although the test surfaces shown in **b** are physically the same, observers saw the right surface as darker and glossier than the left one. Similar after-effects were obtained after adaptation to natural surfaces with skewed histograms, as shown in **c**. (See Supplementary Movies.) **d**, The probability that the subjects judged one test image (on the side that adapted to positive skewness), relative to the other test surface (on the side that adapted to negative skewness) as darker (lightness judgement, left panel) or glossier (glossiness judgement, right panel). Error bars represent ±1 s.e.m. across six observers. There were three adaptation conditions: no adaptation (black circles), adaptation to artificial textures (difference-of-gaussian patterns, blue circles), and adaptation to natural surfaces (stucco images, red circles). The horizontal axis indicates the difference in skewness of the luminance histogram of the two test images. Both plots in **d** correspond to the case of adaptation to positive skewness. The shift of the psychometric function indicates the after-effect.

sensors that are only evident in peripheral processing stages such as retina, LGN and V1. Thus, some of the adaptable processing stages might occur quite early in the visual system.

While skewness is predictive of perceived surface qualities, it can of course be computed on arbitrary images, whether or not they look like surfaces. A picture of fireworks against the night sky will be positively skewed, but one cannot meaningfully judge its albedo or gloss; the same is true of the adapting stimulus of Fig. 4a. Our findings were made in the case where the image is perceived as a surface of uniform albedo with some highlights. We do not know what aspects of image structure determine 'surfaceness' or 'highlights'. When our images are phase-scrambled so as to retain sub-band power, but not phase structure, they are typically seen as plausible but not convincing surfaces. The lightness effects are retained, but glossiness is lost. When the images are pixel-scrambled they are seen as two-dimensional noise patterns without a unitary albedo or gloss. These manipulations and the effects of spatial structure are discussed further in Supplementary Data C and D.

Malik and Perona[21] proposed that even-symmetric filters underlie human sensitivity to contrast sign in texture discrimination. Chubb *et al.*[22], working with unstructured random noise textures, found evidence for a 'blackshot' mechanism sensitive to dark outliers. Our stimuli (which are seen as surfaces rather than two-dimensional random patterns) and our task (judging surface quality rather than discriminating textured regions) are quite different, but the processing could involve similar computations.

The present study gives an interesting perspective on neural computation and natural image statistics. Variance and kurtosis, which are even-order statistics, have been vigorously studied[23–25]. Skewness, an odd-order statistic, has been largely ignored. Even-order statistics are always the same for an image and its negative, so that they are blind to any asymmetries in light and dark (such as those that occur with highlights and shadows). Skewness is specifically sensitive to these asymmetries. It is easily computed, and we find psychophysical evidence that it is used in human vision.

METHODS

Photographs of real surfaces (24 handmade stuccos, fabric, and crumpled paper) were taken by a 16-bit linear camera (Bitran BS-42N). The standard deviation (s.d.) and skewness of the luminance histogram were defined as:

$$\text{s.d.} = \sqrt{\frac{\sum (I(x,y)-m)^2}{N}}$$
$$\text{skewness} = \frac{\sum (I(x,y)-m)^3}{N(\text{s.d.})^3} \quad (1)$$

where $I(x,y)$ is the luminance of a pixel, m the mean luminance, and N the number of pixels (256×256). Surface images were presented on a CRT monitor (Sony GDM-F500R, refresh rate 100 Hz, luminance range of 0.1–82 cd m^{-2}) through a graphics card (Cambridge Research System, VSG2/5), with 8-bit luminance resolution for the luminance range of each image. The mean luminance of all images was normalized to 16.3 cd m^{-2}, and the background luminance was kept below 0.1 cd m^{-2}. In the first experiment (Fig. 2a–c), various stucco images were presented in random order, and subjects rated the lightness or the glossiness using a five-level physical scale (0 to 4). Physical samples (Optical Society of America patches or stucco patches) were shown to the subject as reference. In the second experiment (Fig. 2d), the skewness of the luminance histogram of surface images was varied by the procedure of histogram matching to a beta distribution, given by:

$$f(l) = \frac{1}{B(p,q)} l^{p-1}(1-l)^{q-1}$$
$$B(p,q) = \int_0^1 l^{p-1}(1-l)^{q-1}\,\mathrm{d}l \quad (2)$$

where $q = 10 - p$, l is the luminance and p is the parameter that controls skewness. In the after-effect experiment, subjects viewed an adaptation image pair (Fig. 4a or c) for 100 s at the beginning of a session. In each trial, after 4 s of top-up adaptation, a pair of oppositely skewed stucco images with a given skewness magnitude was presented for 0.5 s, and the subjects indicated which of the two surfaces appeared darker (or glossier). Methods are described more in detail in Supplementary Methods.

Received 21 December 2006; accepted 26 February 2007.
Published online 18 April 2007.

1. Land, E. H. & McCann, J. J. Lightness and retinex theory. *J. Opt. Soc. Am.* **61,** 1–11 (1971).
2. Gilchrist, A. *et al.* An anchoring theory of lightness perception. *Psych. Rev.* **106,** 795–834 (1999).
3. Brainard, D. H. Color constancy in the nearly natural image. 2. Achromatic loci. *J. Opt. Soc. Am. A* **15,** 307–325 (1998).
4. Dana, K. J. *et al.* Reflectance and texture of real-world surfaces. *ACM Trans. Graph.* **18,** 1–34 (1999).
5. Todd, J. T., Norman, J. F. & Mingolla, E. Lightness constancy in the presence of specular highlights. *Psych. Sci.* **15,** 33–39 (2004).
6. Pont, S. C. & Koenderink, J. J. Bidirectional texture contrast function. *Int. J. Comp. Vis.* **62,** 17–34 (2005).
7. Robilotto, R. & Zaidi, Q. Lightness identification of patterned three-dimensional, real objects. *J. Vis.* **6,** 18–36 (2006).
8. Nishida, S. & Shinya, M. Use of image-based information in judgments of surface-reflectance properties. *J. Opt. Soc. Am. A* **15,** 2951–2965 (1998).
9. Fleming, R. W., Dror, R. O. & Adelson, E. H. Real-world illumination and the perception of surface reflectance properties. *J. Vis.* **3,** 347–368 (2003).
10. Dror, R. O., Willsky, A. S. & Adelson, E. H. Statistical characterization of real-world illumination. *J. Vis.* **4,** 821–837 (2004).
11. Fleming, R. W. & Bülthoff, H. H. Low-level image cues in the perception of translucent materials. *ACM Trans. Appl. Percept.* **2,** 346–382 (2005).
12. Levoy, M, *et al.* The Digital Michelangelo Project. ⟨http://graphics.stanford.edu/projects/mich⟩ (2004).
13. Nicodemus, F. Directional reflectance and emissivity of an opaque surface. *Appl. Opt.* **4,** 767–773 (1965).
14. MacGillivray, H. L. Skewness and asymmetry: measures and orderings. *Ann. Stat.* **14,** 994–1011 (1986).
15. Heeger, D. J. Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiol.* **70,** 1885–1898 (1993).
16. Schiller, P. H., Finlay, B. L. & Volman, S. F. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.* **39,** 1288–1319 (1976).
17. Baizer, J. S., Robinson, D. L. & Dow, B. M. Visual responses of area 18 neurons in awake, behaving monkey. *J. Neurophysiol.* **40,** 1024–1037 (1977).
18. Shipp, S. & Zeki, S. The functional organization of area V2, I: specialization across stripes and layers. *Vis. Neurosci.* **19,** 187–210 (2002).
19. Kagan, I., Gur, M. & Snodderly, D. M. Spatial organization of receptive fields of V1 neurons of alert monkeys: comparison with responses to gratings. *J. Neurophysiol.* **88,** 2257–2274 (2002).
20. Mata, M. L. & Ringach, D. L. Spatial overlap of ON and OFF subregions and its relation to response modulation ratio in macaque primary visual cortex. *J. Neurophysiol.* **93,** 919–928 (2005).
21. Malik, J. & Perona, P. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A* **5,** 923–932 (1990).
22. Chubb, C., Landy, M. S. & Econopouly, J. A visual mechanism tuned to black. *Vision Res.* **44,** 3223–3232 (2004).
23. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381,** 607–609 (1996).
24. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24,** 1193–1216 (2001).
25. Kingdom, F. A. A., Hayes, A. & Field, D. J. Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Res.* **41,** 585–598 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Li for discussions. L.S. and E.H.A. were supported by NTT and by a grant from the National Science Foundation to E.H.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.M. (motoyosi@apollo3.brl.ntt.co.jp).

Exploring Features in a Bayesian Framework for Material Recognition

Ce Liu[1,3] Lavanya Sharan[2,3] Edward H. Adelson[3] Ruth Rosenholtz[3]
[1]Microsoft Research New England [2]Disney Research Pittsburgh [3]Massachusetts Institute of Technology
celiu@microsoft.com {lavanya,adelson,rruth}@csail.mit.edu

Abstract

We are interested in identifying the material category, e.g. glass, metal, fabric, plastic or wood, from a single image of a surface. Unlike other visual recognition tasks in computer vision, it is difficult to find good, reliable features that can tell material categories apart. Our strategy is to use a rich set of low and mid-level features that capture various aspects of material appearance. We propose an augmented Latent Dirichlet Allocation (aLDA) model to combine these features under a Bayesian generative framework and learn an optimal combination of features. Experimental results show that our system performs material recognition reasonably well on a challenging material database, outperforming state-of-the-art material/texture recognition systems.

1. Introduction

Material recognition is an important aspect of visual recognition. We interact with a variety of materials on a daily basis and we constantly assess their appearance. For example, when judging where to step on an icy sidewalk or buying fresh produce at a farmers' market or deciding whether a rash requires a trip to the doctor, material qualities influence our decisions. Therefore, it is valuable to build a visual recognition system that can infer material properties from images.

The problem of recognizing materials from photographs has been addressed mainly in the context of reflectance estimation. The visual appearance of a surface depends on several factors – the illumination conditions, the geometric structure of the surface sample at several spatial scales, and the surface reflectance properties, often characterized by the bidirectional reflectance distribution function (BRDF) [24] and its variants [9, 16, 26]. A number of techniques have been developed that can estimate the parameters of a BRDF model from a set of photographs, under restrictive assumptions of illumination, geometry and material properties [10, 11].

In this paper, we focus on recognizing high-level material categories, such as glass, metal, fabric, plastic or wood, instead of explicitly estimating reflectance properties. The reflectance properties of a material are often correlated with its high-level category (e.g. glass is usually translucent and wood is often brown), and in this work, we will exploit these correlations. However, it is important to point out that knowing only the reflectance properties of a surface is not sufficient for determining the material category. For example, the fact that a surface is translucent does not tell us if it is made of plastic, wax or glass.

Figure 1. **Material recognition in the wild.** The goal of this paper is to learn to recognize material categories from a single image. For this purpose, we will use our Flickr Materials Database [28] that captures a range of appearances within each material category.

Unlike other visual recognition tasks such as object or texture recognition, it is challenging to find good features that can distinguish different material categories because of the wide variations in appearance that a material can display. Our strategy is to design several low-level and middle-level features to characterize various aspects of material appearance. In addition to well-established features such as color, jet and SIFT [17, 21], we introduce several new features, such as curvature of edges, histogram of oriented gradient (HOG) feature along edges, and HOG perpendicular

Figure 2. **Material recognition vs. object recognition**. These vehicles are made of different materials (from left to right): *metal*, *plastic* and *wood*.

Figure 3. **Material recognition vs. texture recognition**. These checkerboard patterns are made of different materials (left to right): *fabric*, *plastic* and *paper*.

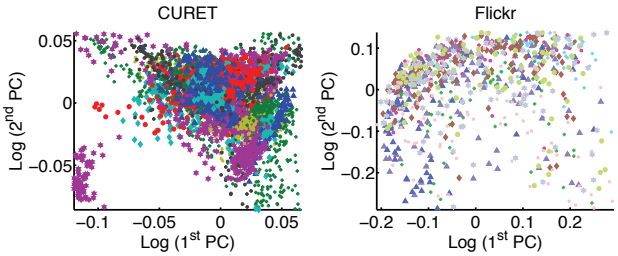


Figure 4. The projection on the first two principal components (PCs) of the texton histograms are shown for all images in the (left) 61 classes in the CURET database [9] and (right) 10 classes in the Flickr Materials Database [28]. The textons were derived from 5×5 pixel patches, as described in [30]. The colors indicate the various texture/material categories. CURET samples are more separable than Flickr.

to edges. After quantizing these features into dictionaries, we convert an image into a bag of words and use latent Dirichlet allocation (LDA) [3] to model the distribution of the words. By allowing topics to be shared amongst material categories, LDA is able to learn clusters of visual words that characterize different materials. We call our model augmented LDA (aLDA) as we concatenate dictionaries from various features and learn the optimal combination of the features by maximizing the recognition rate.

It is crucial to choose the right image database to evaluate our system. Most existing material/texture image databases fail to capture the complexity of real world materials, because they are either instance databases, such as CURET [9], or texture category databases with very few samples per class, such as KTH-TIPS2 [5]. The high recognition rates achieved on these databases ($> 95\%$ on CURET [30]) suggests a need for challenging, real world material databases.

In this work, we use the Flickr Materials Database [28] created by us, originally for studying the visual perception of materials. This database contains 10 common material categories - *fabric*, *foliage*, *glass*, *leather*, *metal*, *paper*, *plastic*, *stone*, *water* and *wood* (see Figure 1). We acquired 100 color photographs from Flickr.com for each category, including 50 close-ups and 50 object-level views. All images have 512×384 pixel resolution and contain a single material category in the foreground. These images capture a wide range of appearances within each material category. We show that although material categorization can be a very challenging problem, especially on a database like ours, our system performs reasonably well, outperforming state-of-the-art systems such as [30].

2. Related Work

Recognizing high-level material categories in images is distinct from the well-studied problem of object recognition. Although object identity is sometimes predictive of material category, a given class of objects can be made of different materials (see Figure 2) and different classes of objects can be made of the same material (see Figure 1). Therefore, many recent advances in object recognition such as shape context [2], object detectors [7] and label transfer [19] may not be applicable for material recognition. In fact, most object recognition systems rely on material-invariant features and tend to ignore material information altogether.

Material recognition is closely related to, but different from, texture recognition. Texture has been defined in terms of dimensions like periodicity, orientedness, and randomness [20]. It can be an important component of material appearance, *e.g.* wood tends to have textures distinct from those of polished metal. However, as illustrated in Figure 3, surfaces made of different materials can share the same texture patterns and as a consequence, mechanisms designed for texture recognition [18, 30] may not be ideal for material recognition.

Material recognition is also different from BRDF estimation. In computer graphics, there is great interest in capturing the appearance of real world materials. The visual appearance of materials like wood or skin, has been modeled in terms of the bidirectional reflectance distribution function (BRDF) [10, 22] and related representations such as BTF [9] and BSSRDF [16]. Material recognition might seem trivial if the BRDF is known, but in general, it is nearly impossible to estimate the BRDF from a single image without simplifying assumptions [10, 11].

A number of low-level image features have been developed for identifying materials. The shape of the luminance histogram of images was found to correlate with human judgement of surface albedo [25], and was used to classify images of spheres as shiny, matte, white, grey *etc*. [11]. Similar statistics were used to estimate the albedo and gloss of stucco-like surfaces [27]. Several techniques have been developed to search for specific materials in real world photographs such as glass [15, 23] or skin [14].

The choice of databases is, often, the key to success in vi-

sual recognition. The CURET database [9] that consists of images of 61 different texture samples under 205 different viewing and lighting conditions, has become the standard for evaluating 3-D texture classification algorithms. A variety of methods based on texton representations [6, 18, 29], bidirectional histograms [8] and image patches [30] have been successful at classifying CURET surfaces ($> 95\%$ accuracy). The KTH-TIPS2 database [5] consisting of 11 texture categories, 4 samples per category, and each photographed under a variety of conditions, was introduced to increase the intra-class variation. It was shown that a SVM-based classifier achieves 98.5% accuracy on this database [5]. Our Flickr Materials Database [28] contains 10 material categories and 100 diverse samples in category. On inspecting the images in Figure 1 and the plots in Figure 4, it is apparent that the Flickr Materials Database is more challenging than the CURET database, and for this reason we chose the Flickr Materials Database to develop and evaluate our material recognition system.

3. Features for Material Recognition

In order to build a material recognition system, it is important to identify features that can distinguish material categories from one another. What makes metal look like metal and wood look like wood? Is it color (neutral vs. browns), textures (smooth vs. grainy) or reflectance properties (shiny vs. matte)? Since little is known about which features are suited for material recognition, our approach is to try a variety of features, some borrowed from the fields of object and texture recognition, and some new ones developed specifically for material recognition. From a rendering point of view, once the camera and the object are fixed, the image of the object can be determined by (i) the BRDF of the surface, (ii) surface structures, (iii) object shape and (iv) environment lighting. Given the diversity of appearance in the Flickr Materials Database, we will attempt to incorporate all these factors in our features.

(a) Color and Texture

Color is an important attribute of surfaces and can be a cue for material recognition: wooden objects tend to be brown, leaves are green, fabrics and plastics tend to be saturated with vivid color, whereas stones tend to be less saturated. We extract 3×3 pixel patches from an RGB image as our color feature.

Texture, both of the wallpaper and 3-D kind [26], can be useful for distinguishing materials. For example, wood and stone have signature textures that can easily tell them apart. We use two sets of features to measure texture. The first set comprises the filter responses of an image through a set of multi-scale, multi-orientation Gabor filters, often called filter banks or **jet** [17]. Jet features have been used to recognize 3-D textures [18, 30] by clustering to form *textons* and using the distribution of textons as a feature. The second set

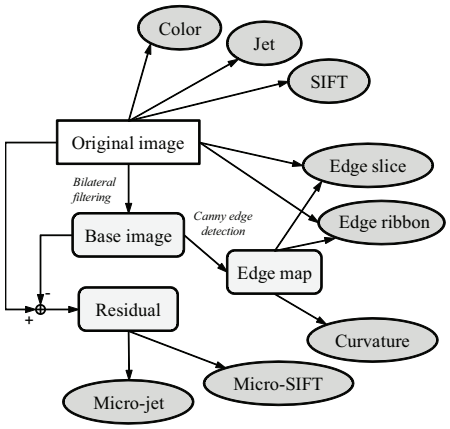
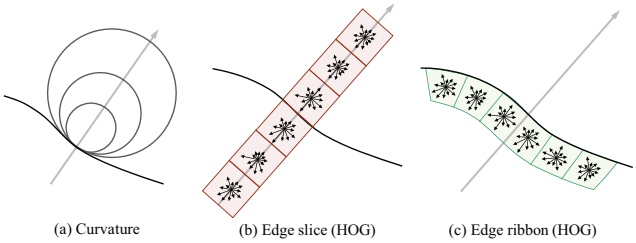


Figure 5. Illustration of how our system generates features.



(a) Curvature (b) Edge slice (HOG) (c) Edge ribbon (HOG)

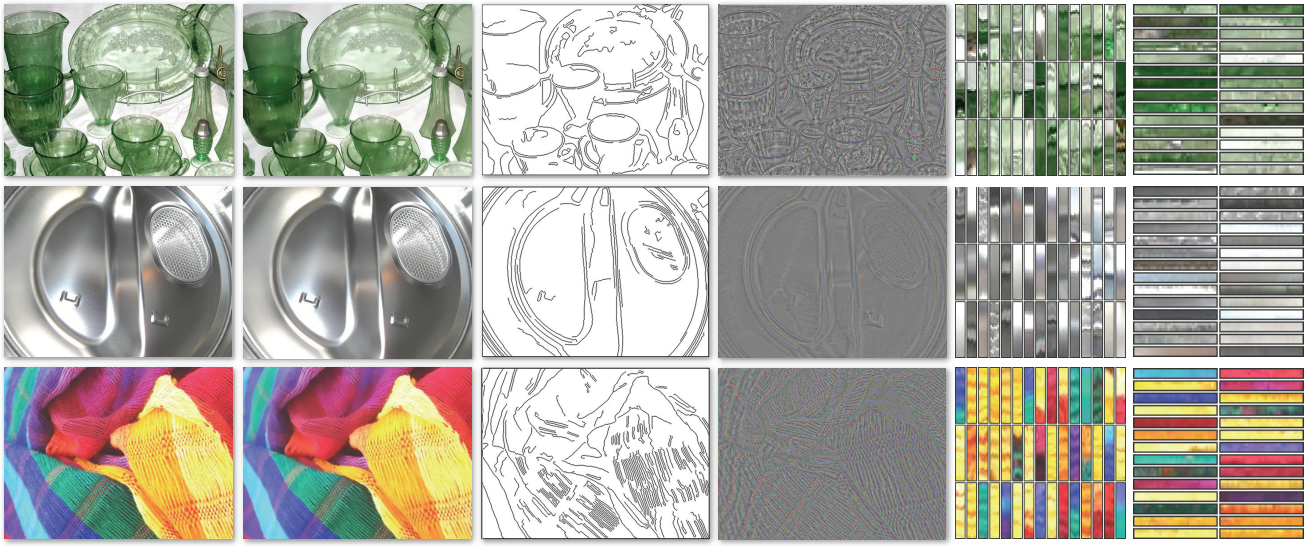
Figure 6. We extract *curvature* at three scales, *edge slice* in 6 cells, and *edge ribbon* in 6 cells at edges.

of features we use is **SIFT** [21]. SIFT features have been widely used in scene and object recognition to characterize the spatial and orientational distribution of local gradients [13].

(b) Micro-texture

Two surfaces sharing the same BRDF can look different if they have different surface structures, *e.g.* if one is smooth and the other is rough. In practice, we usually touch a surface to sense how rough (or smooth) it is. However, our visual system is able to perceive these properties even without a haptic input. For example, we can see tiny hairs on fabric, smooth surfaces in glass objects, crinkles in leather and grains in paper.

In order to extract information about surface structure, we followed the idea in [1], of smoothing an image by bilateral filtering [12] and then using the residual image for further analysis. The process is illustrated in Figure 7. We choose three images from material categories (a) - glass, metal and fabric - and perform bilateral filtering to obtain base image in (b) and display the residual in (d). The residual of bilateral filtering reveals variations in pixel intensity at a finer scale. For the fabric and metal example in Figure 7, the residual is due to surface structure whereas for glass, these variations are related to translucency. Although it is hard to cleanly separate the contributions of surface structure from those of the BRDF, the residual contains useful information about material category. We apply the same approach for characterizing the residual as we did for texture.



(a) Original image (b) Base image (bilateral filtering) (c) Canny edges on (b) (d) Residual image: (a)-(b) (e) Edge slices (f) Edge ribbon

Figure 7. **Some features for material recognition**. From top to bottom is *glass*, *metal* and *fabric*. For an image (a) we apply bilateral filtering [1] to obtain the base image (b). We run Canny edge detector [4] on the base image and obtain edge maps (c). *Curvatures* of the edges are extracted as features. Subtracting (b) from (a), we get the residual image (d) that shows micro structures of the material. We extract *micro-jet* and *micro-SIFT* features on (d) to characterize material micro-surface structure. In (e), we also show some random samples of edge slices along the normal directions of the Canny edges. These samples reveal lighting-dependent features such as specular highlights. The edge ribbon samples are shown in (f). Arrays of HOG's [7] are extracted from (e) and (f) to form *edge-slice* and *edge-ribbon* features.

We compute the jet and SIFT features of the residual image, and name them **micro-jet** and **micro-SIFT** for clarity.

(c) Outline Shape

Though a material can be cast into any arbitrary shape, the outline shape of a surface and its material category are often related *e.g*. fabrics and glass have long, curved edges, while metals have straight lines and sharp corners. The outline shape of a surface can be captured by an edge map. We run the Canny edge detector [4] on the base image, trim out short edges, and obtain the edge map shown in Figure 7 (c). To characterize the variations in the edge maps across material categories, we measured the **curvature** on the edge map at three different scales as a feature (see Figure 6).

(d) Reflectance-based features

Glossiness and transparency are important cues for material recognition. Metals are mostly shiny, whereas wooden surfaces are usually dull. Glass and water are translucent, while stones are often opaque. These reflectance properties sometimes manifest as distinctive intensity changes at the edges in an image. To measure these changes, as shown in Figure 6 (b), we extract histogram of oriented gradients (HOG) [7] features along the *normal* direction of edges. We take a slice of pixels with a certain width along the normal direction, compute the gradient at each pixel, divide the slice into 6 cells, and quantize the oriented gradients in to 12 angular bins. This feature is called **edge-slice**. We also measure how the images change along the *tangent* direction of the edges in a similar manner, as suggested in Figure 6 (c). This feature is called **edge-ribbon**, which is also quantized by 6 cells and 12 angular bins for each cell.

We have described a pool of features that can be potentially useful for material recognition: **color**, **SIFT**, **jet**, **micro-SIFT**, **micro-jet**, **curvature**, **edge-slice** and **edge-ribbon**. The flowchart of how our system generates these features is shown in Figure 5. Amongst these features, color, SIFT and jet are *low-level* features directly computed from the original image and they are often used for texture analysis. The rest of the features, micro-SIFT, micro-jet, curvature, edge-slice and edge-ribbon are *mid-level* features that rely on estimations of base images and edge maps (Figures 7 (b) & (c)). A priori, we do not know which of these features will perform well. Hence, we designed a Bayesian learning framework to select best combination of features.

4. A Bayesian Computational Framework

Now that we have a pool of features, we want to combine them to build an effective material recognition system. We quantize the features into visual words and extend the LDA [3] framework to select good features and learn per-class distributions for recognition.

4.1. Feature quantization and concatenation

We use the standard k-means algorithm to cluster the instances of each feature to form dictionaries and map image

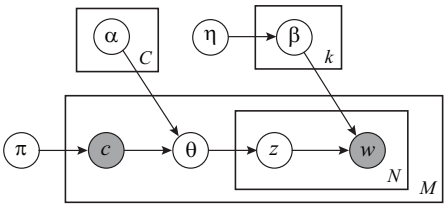


Figure 8. The graphical model of LDA [3]. Notice that our categorization shares both the topics and codewords. Different from [13], we impose a prior on β to account for insufficient data.

features into visual words. Suppose there are m features in the feature pool and m corresponding dictionaries $\{D_i\}_{i=1}^m$. Each dictionary has V_i codewords, *i.e.* $|D_i|=V_i$. Since features are quantized separately, the words generated by the ith feature are $\{w_1^{(i)},\cdots,w_{N_i}^{(i)}\}$, $w_j^{(i)}\in\{1,2,\cdots,V_i\}$ and N_i is the number of words. In order to put a set of different words together, a document of m sets of words

$$\{w_1^{(1)},\cdots,w_{N_1}^{(1)}\},\{w_1^{(2)},\cdots,w_{N_2}^{(2)}\},\cdots,\\ \{w_1^{(m)},\cdots,w_{N_m}^{(m)}\} \quad (1)$$

can be augmented to one set

$$\{w_1^{(1)},\cdots,w_{N_1}^{(1)},w_1^{(2)}+V_1,\cdots,w_{N_2}^{(2)}+V_1,\cdots,\\ w_1^{(m)}+\textstyle\sum_{i=1}^{m-1}V_i,\cdots,w_{N_m}^{(m)}+\sum_{i=1}^{m-1}V_i\} \quad (2)$$

with a joint dictionary $\mathbb{D}=\cup_i D_i,\ |\mathbb{D}|=\sum_{i=1}^m V_i$. In this way, we reduced the multi-dictionary problem to a single-dictionary one.

4.2. Latent Dirichlet Allocation

The latent Dirichlet allocation (LDA) [3] was invented to model the hierarchical structures of words. Details of the model can be found in [3, 13]. In order to be self-contained, we will briefly describe the model in the context of material recognition. As depicted in the graphical model in Figure 8, we first randomly draw the material category $c\sim\mathrm{Mult}(c|\pi)$ where $\mathrm{Mult}(\cdot|\pi)$ is a multinomial distribution with parameter π. Based on c, we select a hyper-parameter α_c, based on which we draw $\theta\sim\mathrm{Dir}(\theta|\alpha_c)$ where $\mathrm{Dir}(\cdot|\alpha_c)$ is a Dirichlet distribution with parameter α_c. θ has the following property: $\sum_{i=1}^k\theta_i=1$ where k is the number of elements in θ. From θ we can draw a series of topics $z_n\sim\mathrm{Mult}(z|\theta), n=1,\cdots,N$. The topic $z_n(=1,\cdots,k)$ selects a multinomial distribution β_{z_n} from which we draw a word $w_n\sim\mathrm{Mult}(w_n|\beta_{z_n})$, which corresponds to a quantization cluster of the features. Unlike [13] where β is assumed to be a parameter, we impose a conjugate prior η upon β to account for insufficient data as suggested by [3].

Since it is intractable to compute the log likelihood $\log p(w|\alpha_c,\eta)$, we instead maximize the lower bound $\mathcal{L}(\alpha_c,\eta)$ estimated through the variational distributions over $\theta,\{z_d\},\beta$. Please refer to [3] for details on deriving the variational lower-bound and parameter learning for α and η. Once we have learned α_c and η, we can use Bayesian MAP criterion to choose the material category

$$c^*=\arg\max_c\mathcal{L}(\alpha_c,\eta)+\lambda_c. \quad (3)$$

where $\lambda_c=\log\pi_c$.

4.3. Prior learning

A uniform distribution is often assumed for the prior $p(c)$, *i.e.* each material category will appear equally. However, since we learn the LDA model for each category independently (only sharing the same β), the learning procedure may not converge in finite iterations. Therefore, the probability density functions (pdfs) should be grounded for a fair comparison. We designed the following greedy algorithm to learn λ by maximizing the recognition rate (or minimizing the error).

Suppose $\{\lambda_i\}_{i\neq c}$ is fixed and we want to optimize λ_c to maximize the rate. Let y_d be the label for document d. Let $q_{d,i}=\mathcal{L}_d(\alpha_i,\eta)+\lambda_i$ be the "log posterior" for document d to belong to category i. Let $f_d=\max_i q_{d,i}$ be the maximum posterior for document d. We define two sets:

$$\begin{aligned}\Omega_c &= \{d|y_d=c,\ f_d>q_{d,c}\},\\ \Phi_c &= \{d|y_d\neq c,\ f_d=q_{d,y_d}\}.\end{aligned} \quad (4)$$

Set Ω_c includes the documents that are labeled as c and misclassified. Set Φ_c includes the documents that are not labeled as c and correctly classified. Our goal is to choose λ_c to make $|\Omega_c|$ as small as possible and $|\Phi_c|$ as large as possible. Notice that if we increase λ_c, then $|\Omega_c|\downarrow$ and $|\Phi_c|\downarrow$, therefore the optimal λ_c exists. We define the set of correctly classified documents with λ_c's

$$\begin{aligned}\Psi_c &= \{d|d\in\Omega_c,\ f_d<q_{d,c}+\lambda_c'-\lambda_c\}\cup\\ &\quad\{d|d\in\Phi_c,\ f_d>q_{d,c}+\lambda_c'-\lambda_c\},\end{aligned} \quad (5)$$

and choose the new λ_c that maximizes the size of Ψ_c:

$$\lambda_c\leftarrow\arg\max_{\lambda_c'}|\Psi_c|. \quad (6)$$

We iterate this procedure for each c repeatedly until each λ_c does not change.

4.4. Augmented LDA (aLDA)

Shall we use all the features in our predefined feature pool? Do more features imply better performance? Unfortunately, this is not true as we have limited training data. The more features we use, the more likely that the model overfits the training data and the performance decreases on test set. We designed a greedy algorithm in Figure 9 to select an optimal subset of our feature pool. The main idea is to select the best feature, one at a time, that maximizes the recognition rate on an evaluation set. The algorithm stops when adding more features will decrease the recognition rate. Note that we randomly split the training set H into L for parameter learning and E for cross evaluation. After $\mathbb{D}$ is learned, we use the entire training set H to relearn the parameters for $\mathbb{D}$.

Input: dictionary pool $\{D_1, \cdots, D_m\}$, training set H

- Initialize: $\mathbb{D} = \emptyset$, recognition rate $r = 0$
- Randomly split $H = L \cup E$

  for $l = 1$ to m

    for $D_i \notin \mathbb{D}$

      - Augment dictionary $\mathbb{D}' = \mathbb{D} \cup \{D_i\}$
      - Concatenate words according to $\mathbb{D}'$ using Eqn. (2)
      - Train LDA on L for each category (sharing β)
      - Learn prior λ using Eqn. (5) and (6)
      - r_i = recognition rate on E using Eqn. (3)

    end

    if $\max r_i > r$

      - $j = \arg\max_i r_i, \mathbb{D} = \mathbb{D} \cup \{D_j\}, \ r = r_j$

    else

      break

    end

  end

- Train LDA and learn prior λ on H
- r = recognition rate on H

Output: $\mathbb{D}, \ r$

Figure 9. **The augmented LDA (aLDA) algorithm**.

5. Experimental Results

We used the Flickr Materials Database [28] for all experiments described in this paper. There are ten material categories in the database: *fabric*, *foliage*, *glass*, *leather*, *metal*, *paper*, *plastic*, *stone*, *water* and *wood*. Each category contains 100 images, 50 of which are close-up views and the rest 50 are of views at object-scale (see Figure 1). There is a binary, human-labeled mask associated with each image describing the location of the object. We only consider pixels inside this binary mask for material recognition and disregard all the background pixels. For each category, we randomly chose 50 images for training and 50 images for test. All the experimental results reported in this paper are based on the same split of training and test.

We extract features for each image according to Figure 5. Mindful of computational costs, we sampled *color*, *jet*, *SIFT*, *micro-jet* and *micro-SIFT* features on a coarse grid (every 5^{th} pixel in both horizontal and vertical directions). Because there are far fewer pixels in edge maps than in the original images, we sampled every other edge pixel for *curvature*, *edge-slice* and *edge-ribbon*. Once features are extracted, they are clustered separately using k-means according to the number of clusters in Table 1. We specified the number of clusters for each feature, considering both dimensionality and the number of instances per feature.

After forming the dictionaries for each feature, we run the aLDA algorithm to select features incrementally. When learning the optimal feature set, we randomly split the 50 training images per category (set H) to 30 for estimating parameters (set L) and 20 for evaluation (set E). After the feature set is learned, we re-learn the parameters using the 50 training images per category and report the training/test rate. In the LDA learning step, we vary the number of topics from 50 to 250 with step size 50 and pick the best one. The learning procedure is shown in Figure 10, where for each material category we plot the training rate on the left in a darker color and test rate on the right in a lighter color. In Figure 10, the recognition rate is computed on the entire training/test set, not just on the learning/evaluation set. First, the system tries every single feature and discovers that amongst all features, SIFT produces the highest evaluation rate. In the next iteration, the system picks up color from the remaining features, and then edge-slice. Including more features causes the performance to drop and the algorithm in Figure 9 stops. For this final feature set "color + SIFT + edge-slice", the training rate is 49.4% and the test rate is 44.6%. The recognition rate of random guesses is 10%.

The boost in performance from the single best feature (SIFT, 35.4%) to the best feature set (color + SIFT + edge-slice, 44.6%) is due to our aLDA model that augments visual words. Interestingly, augmenting more features decreases the overall performance. When we use all the features, the test rate is 38.8%, lower than using fewer features. More features creates room for overfitting, and one solution to combat overfitting is to increase the size of the database. The fact that SIFT is the best-performing single feature in-

Feature name	Dim	average # per image	# of clusters
Color	27	6326.0	150
Jet	64	6370.0	200
SIFT	128	6033.4	250
Micro-jet	64	6370.0	200
Micro-SIFT	128	6033.4	250
Curvature	3	3759.8	100
Edge-slice	72	2461.3	200
Edge-ribbon	72	3068.6	200

Table 1. The dimension, number of clusters and average number per image for each feature.

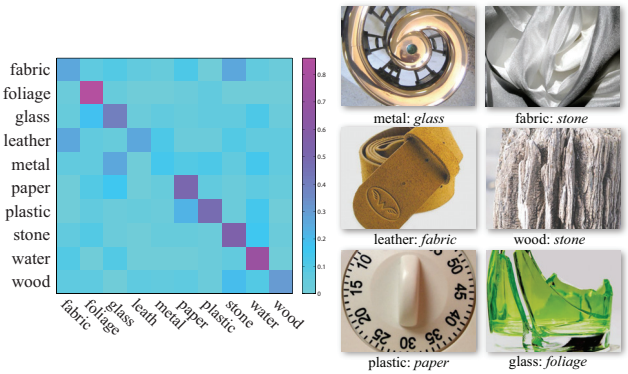


Figure 12. Left: the confusion matrix of our material recognition system using color + SIFT + edge-slice feature set. Row k is the probability distribution of class k being classified to each category. Right: some misclassification examples. Label "metal: *glass*" means that a metal material is misclassified as glass.

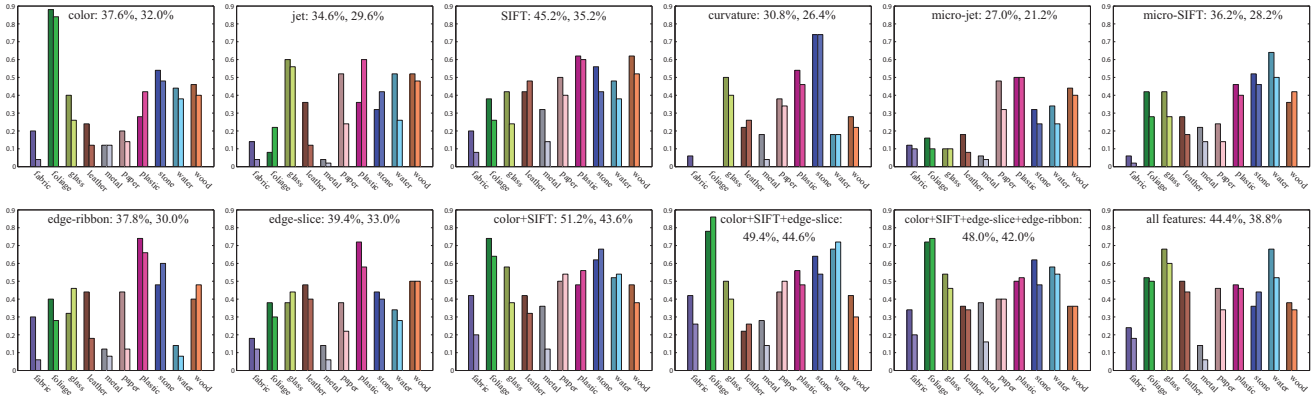


Figure 10. **The per-class recognition rate (both training and test) with different sets of features for the Flickr database** [28]. In each plot, the left, darker bar means training, the right, lighter bar means test. For the two numbers right after the feature set label are the recognition rate on the entire training set and the rate on the entire test set. For example, "color: 37.6%, 32.0%" means that the training rate is 37.6% and the test rate is 32.0%. Our aLDA algorithm finds "color + SIFT + edge-slice" to be the optimal feature set on the Flickr Materials Database.

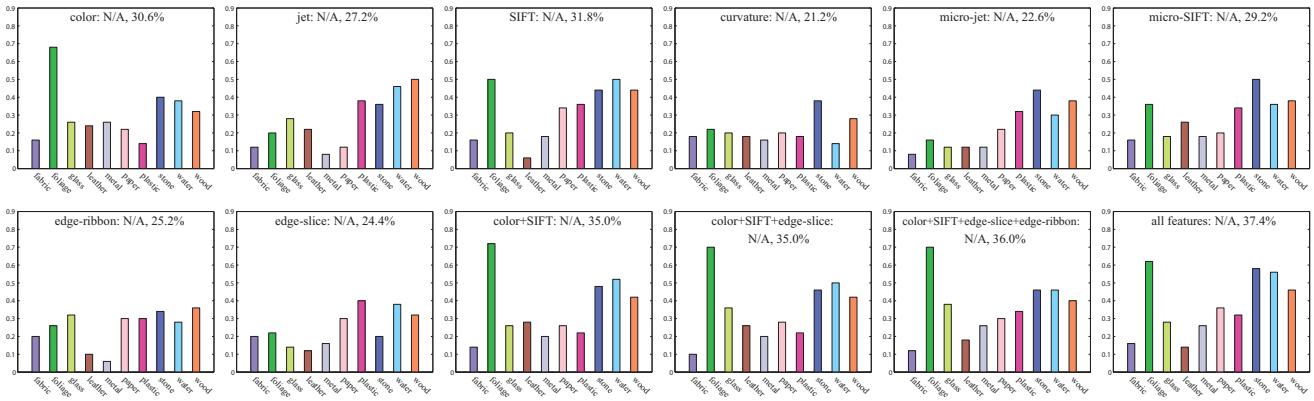


Figure 11. For comparison, we run Varma-Zisserman's system [30] (nearest neighbor classifiers using histograms of visual words) on our feature sets. Because of the nearest neighbor classifier, the training rate is always 100%, so we simply put it as N/A.

dicates the importance of texture in material recognition. In addition, SIFT also encapsulates some of the information captured by micro-SIFT. Edge-slice, which measures reflectance features, is also useful.

For comparison, we implemented and tested Varma-Zisserman's (VZ) algorithm [30] on the Flickr Materials Database. The VZ algorithm clusters 5×5 pixel gray-scale patches as codewords, obtains a histogram of the codewords for each image, and performs recognition using a nearest neighbor classifier. As a sanity check, we ran our implementation of VZ on the CURET database and obtained 96.1% test rate (their numbers are $95 \sim 98\%$, [30]). Next, we ran the exact VZ system tested on CURET on the Flickr Materials Database. The VZ test rate is 23.8%. This supports the conclusions from Figure 4 that the Flickr Materials Database is much harder than the CURET texture database.

As the VZ system uses features tailored for the CURET database (5×5 pixel patches), we ran VZ's algorithm using our features on Flickr Materials Database. The results of running VZ's system on exactly the same feature sets as in Figure 10 are listed in Figure 11. Since VZ uses a nearest neighbor classifier, it is meaningless to report the training rate as it is always 100%, so we only report the test rate. It is obvious why many of our features outperform fixed size gray-scale patch features on Flickr Materials Datatbase. In fact, the VZ system running on SIFT features has test rate of 31.8%, close to our system using SIFT alone (35.2%). However, combining features under the VZ's framework only slightly increases the performance to a maximum of 37.4%. Clearly, the aLDA model contributes to the boost in performance from 37.4% to 44.6%.

The confusion matrix of our system (color + SIFT + edge-slice, test rate 44.6%) in Figure 12 tells us how often each category is misclassified as another. For example, *fabric* is often misclassified as *stone*, *leather* misclassified as *fabric*, *plastic* misclassified as *paper*. The category *metal* is more likely to be classified as *glass* than itself. Some misclassification examples are shown in Figure 12. These results are not surprising because there are certain commonalities between *leather* and *fabric*, *plastic* and *paper*, as well

as *metal* and glass, as shown in Figure 12.

6. Discussion and Conclusion

Although the recognition rate achieved by our system 44.6% is lower than the rates reported in object recognition (*e.g.* [19]), it is significantly higher than the state of the art (23.8%, [30]). As illustrated in Figures 1 and 4, the sheer diversity and range of the Flickr Materials Database makes it a challenging benchmark for material recognition. We believe that material recognition is an important problem to study, and in this paper, we are merely taking one of the first steps towards understanding the problem.

To conclude, we have presented a set of features and a Bayesian computational framework for material category recognition. Our features were chosen to capture various aspects of material appearance in the real world. An augmented LDA (aLDA) framework was designed to select an optimal set of features by maximizing the recognition rate on the training set. We have demonstrated a significant improvement in performance when using our system over the state of the art on the challenging Flickr Materials Database [28]. We have also analyzed the contribution of each feature in our system to the performance gain. Our feature set and computational framework constitute the first attempt at recognizing high-level material categories "in the wild".

References

[1] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. In *ACM SIGGRAPH*, 2006.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *TPAMI*, 24(4):509–522, 2002.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3):993–1022, 2003.

[4] J. Canny. A computational approach to edge detection. *TPAMI*, 8(6):679–698, Nov 1986.

[5] B. Caputo, E. Hayman, M. Fritz, and J.-O. Eklundh. Classifying materials in the real world. *Image and Vision Computing*, 28(1):150 – 163, 2010.

[6] O. G. Cula and K. J. Dana. Recognition methods for 3d textured surfaces. In *SPIE, Human Vision and Electronic Imaging VI*, pages 209–220, 2001.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 2, pages 886–893, 2005.

[8] K. J. Dana and S. Nayar. Histogram model for 3d textures. In *CVPR*, pages 618–624, 1998.

[9] K. J. Dana, B. Van-Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, 1999.

[10] P. Debevec, T. Hawkins, C. Tchou, H. P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *ACM SIGGRAPH*, pages 145–156, 2000.

[11] R. Dror, E. H. Adelson, and A. S. Willsky. Recognition of surface reflectance properties from a single image under unknown real-world illumination. In *IEEE Workshop on identifying objects across variation in lighting*, 2001.

[12] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *ACM SIGGRAPH*, 2002.

[13] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, volume 2, pages 524–531, 2005.

[14] D. Forsyth and M. M. Fleck. Automatic detection of human nudes. *IJCV*, 32(1):63–77.

[15] M. Fritz, M. Black, G. Bradski, and T. Darrell. An additive latent feature model for transparent object recognition. In *NIPS*, 2009.

[16] H. W. Jensen, S. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *ACM SIGGRAPH*, pages 511–518.

[17] J. Koenderink and A. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 545:367–375.

[18] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1):29–44.

[19] C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing: Label transfer via dense scene alignment. In *CVPR*, 2009.

[20] F. Liu and W. Picard. Periodicity, directionality and randomness: Wold features for image modeling and retrieval. *TPAMI*, 18:722–733.

[21] D. G. Lowe. Distinctive image-features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[22] S. Marschner, S. H. Westin, A. Arbree, and J. T. Moon. Measuring and modeling the appearance of finished wood. In *ACM SIGGRAPH*, pages 727–734, 2005.

[23] K. McHenry and J. Ponce. A geodesic active contour framework for finding glass. In *CVPR*, volume 1, pages 1038–1044.

[24] F. Nicodemus. Directional reflectance and emissivity of an opaque surface. *Applied Optics*, 4(7):767–775, 1965.

[25] S. Nishida and M. Shinya. Use of image-based information in judgments of surface reflectance properties. *Journal of the Optical Society of America A*, 15:2951–2965.

[26] S. C. Pont and J. J. Koenderink. Bidirectional texture contrast function. *IJCV*, 62(1).

[27] L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson. Image statistics for surface reflectance perception. *Journal of the Optical Society of America A*, 25(4):846–865, 2008.

[28] L. Sharan, R. Rosenholtz, and E. Adelson. Material perception: What can you see in a brief glance? [Abstract]. Journal of Vision, 9(8):784, 2009.

[29] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *IJCV*, 62(1–2):61–81, 2005.

[30] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *TPAMI*, 31(11):2032–2047, 2009.