

# Image statistics for surface reflectance perception

Lavanya Sharan,<sup>1,\*</sup> Yuanzhen Li,<sup>2</sup> Isamu Motoyoshi,<sup>3</sup> Shin'ya Nishida,<sup>3</sup> and Edward H. Adelson<sup>2</sup>

<sup>1</sup>*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

<sup>2</sup>*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue,  
Cambridge, Massachusetts 02139, USA*

<sup>3</sup>*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](mailto: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. © 2008 Optical Society of America

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 pla-

nar 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

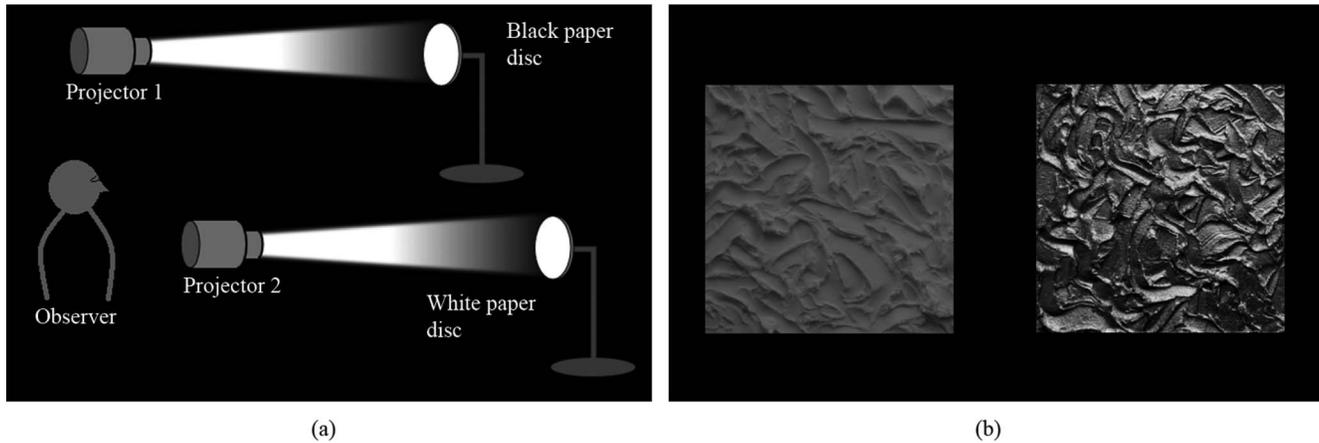


Fig. 1. (a) In the Gelb demonstration a smooth Lambertian black surface can look the same as a smooth Lambertian white surface. (b) Gelb effect fails for complex surfaces. The stucco samples have the same mean luminance, yet it is easy to tell the white stucco from the black.

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, in-

terreflection, 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.

## 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-

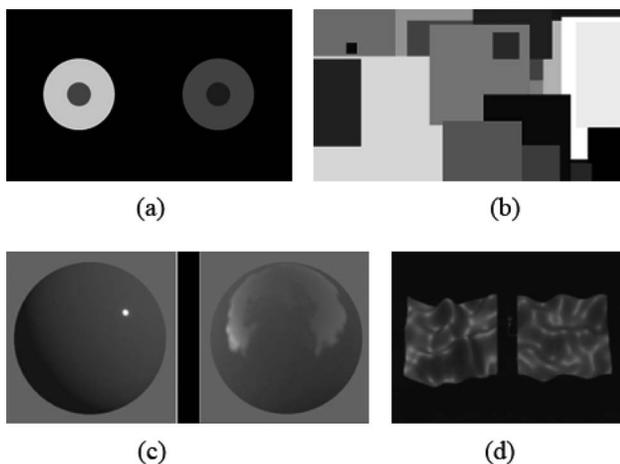


Fig. 2. Stimuli used for studying reflectance perception: (a) Wallach's disc annulus displays, (b) Mondrian-like displays with flat Lambertian surfaces, (c) Fleming *et al.*'s simulated spheres in complex real-world illumination [18], (d) simulated locally smooth bumpy surfaces used by Nishida and Shinya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

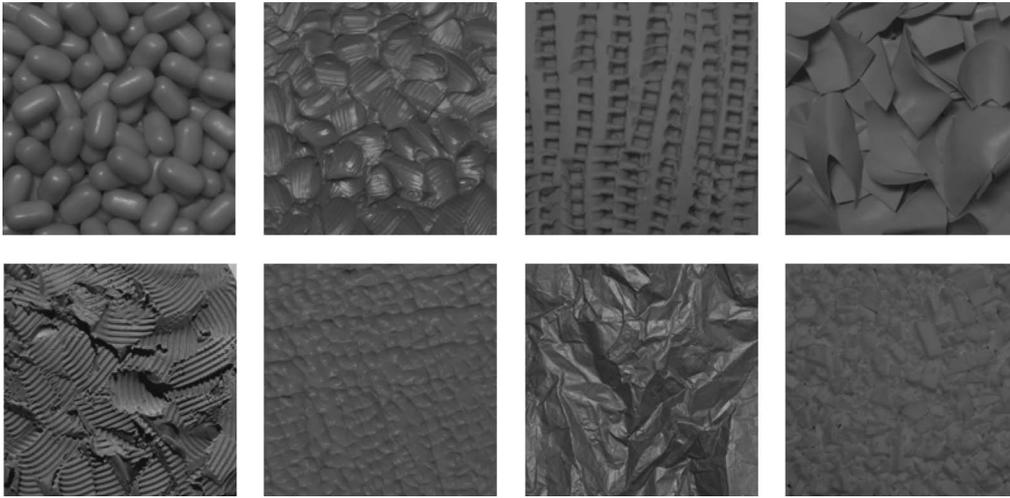


Fig. 3. Examples of surfaces in our data set. All surfaces shown here were photographed under an overhead fluorescent light.

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 “dcrw” 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 compo-

nents 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

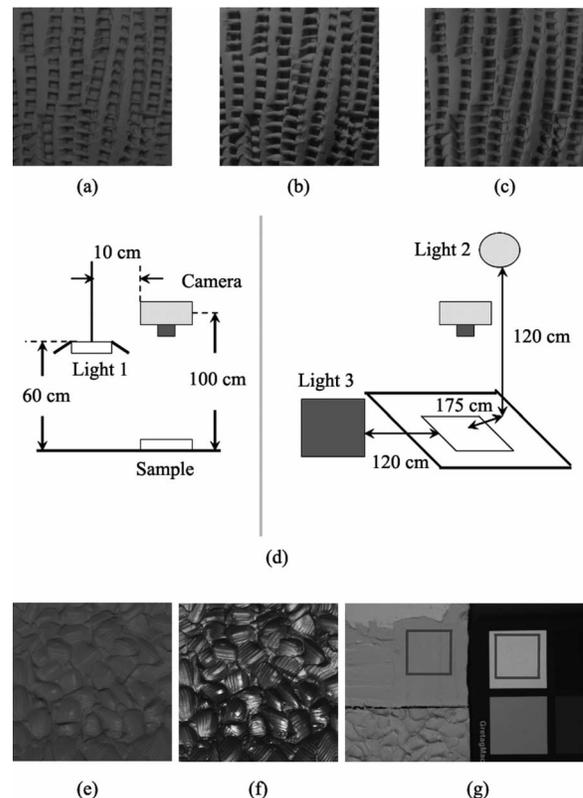


Fig. 4. Image data acquisition. Three indoor lighting conditions were used: (a) light 1, overhead fluorescent light source; (b) light 2, focused halogen spotlight; (c) light 3, diffuse tungsten halogen lamp. (d) Schematic layout of the setup. Two views, one from the side and one from front, are shown. (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—

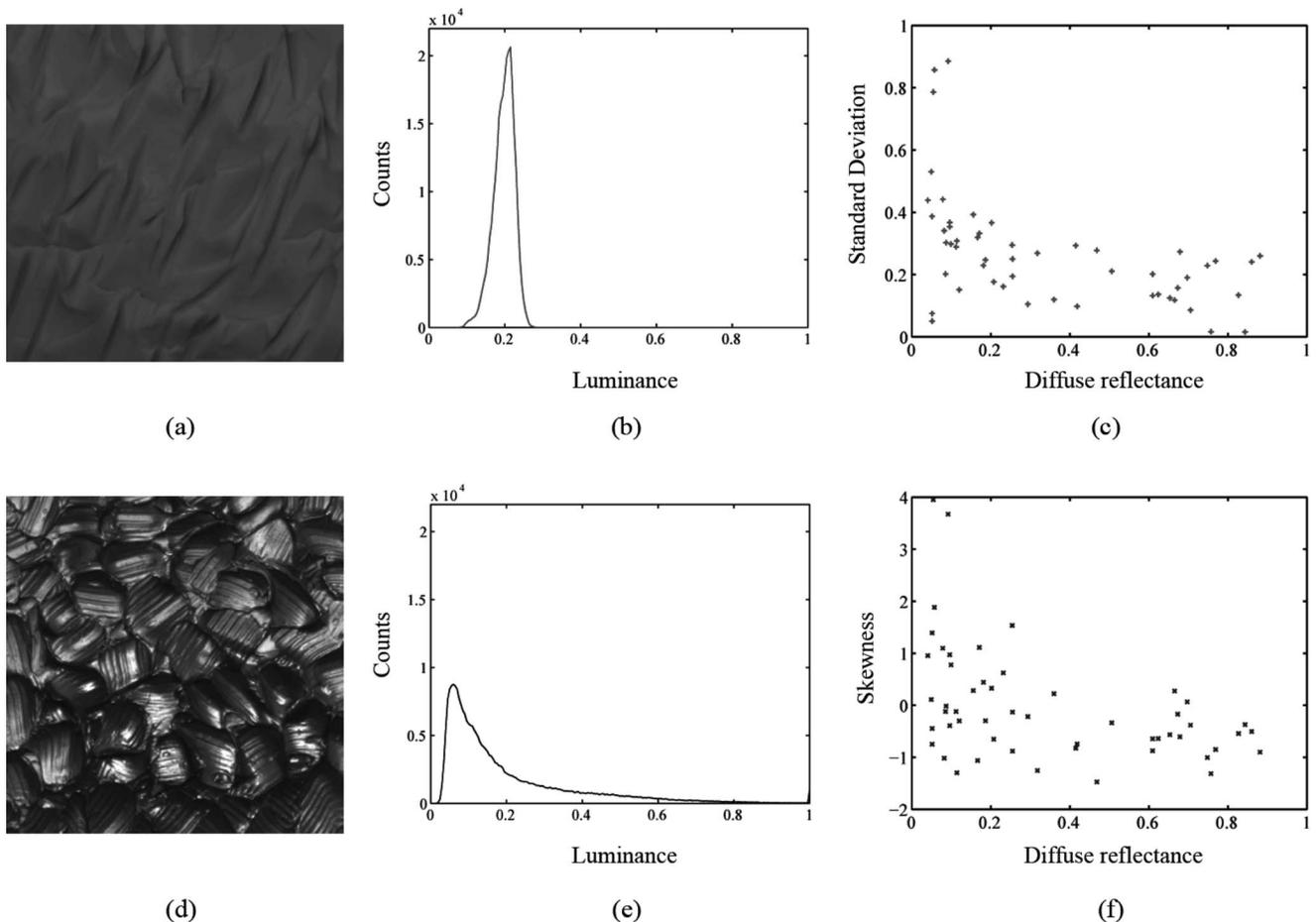


Fig. 5. Luminance histograms of light and dark materials exhibit systematic differences. (a) Light modeling clay; (d) dark stucco; and (b), (c) the respective luminance histograms. (c), (f) Standard deviation and skewness of the log-luminance histogram is plotted against the ground truth for albedo for 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 multi-scale 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

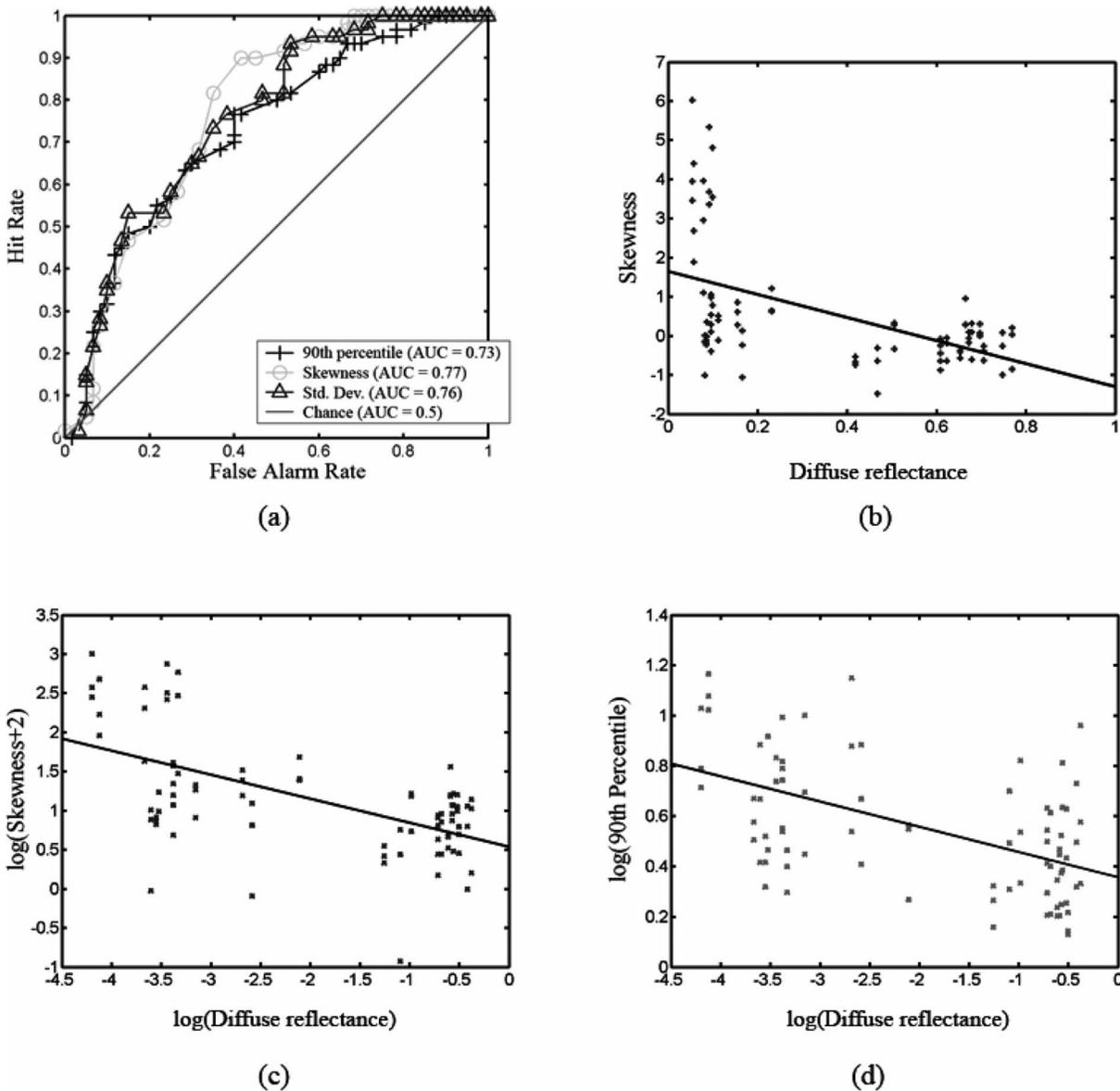


Fig. 6. (a) ROC curves for the 90th percentile, standard deviation, and skewness of log-luminance values. These statistics perform well above chance at the task of classifying surfaces as light or dark. (b) Skewness of the log-luminance histogram is plotted against the physical diffuse reflectance. A 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 ( $p < 0.05$ ,  $r^2 = 0.34$ ). (d) 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  $\epsilon$  parameter of the regression was set to 0.1, and the penalty param-

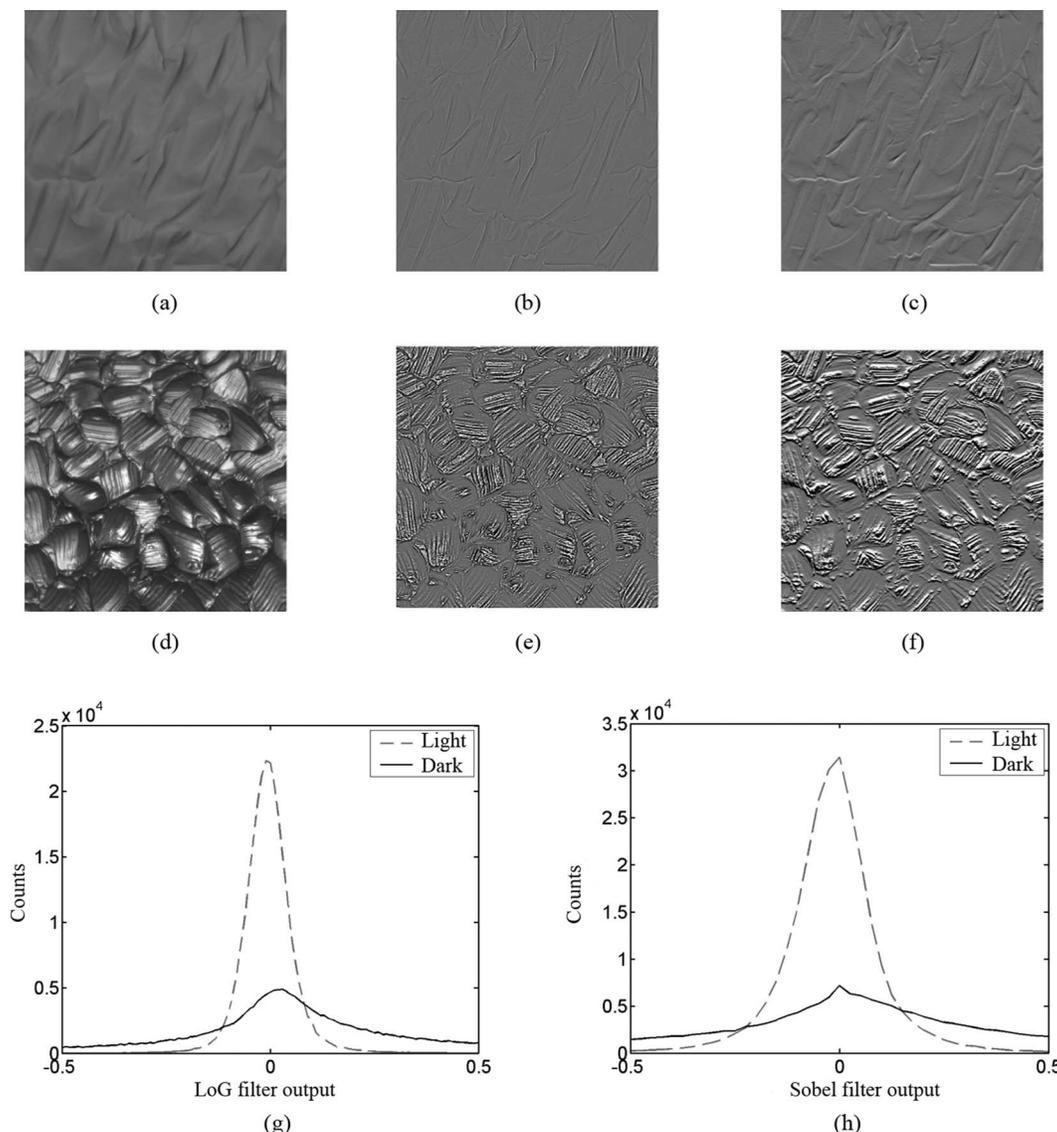


Fig. 7. Pixel histograms of filtered images look different for white and black surfaces. (a) Light modeling clay from Fig. 5; (b) output of Laplacian of Gaussian filter ( $\sigma=0.5$  of size  $5 \times 5$  pixels); (c) output of horizontal Sobel filter ( $3 \times 3$  pixels); (d) dark stucco from 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

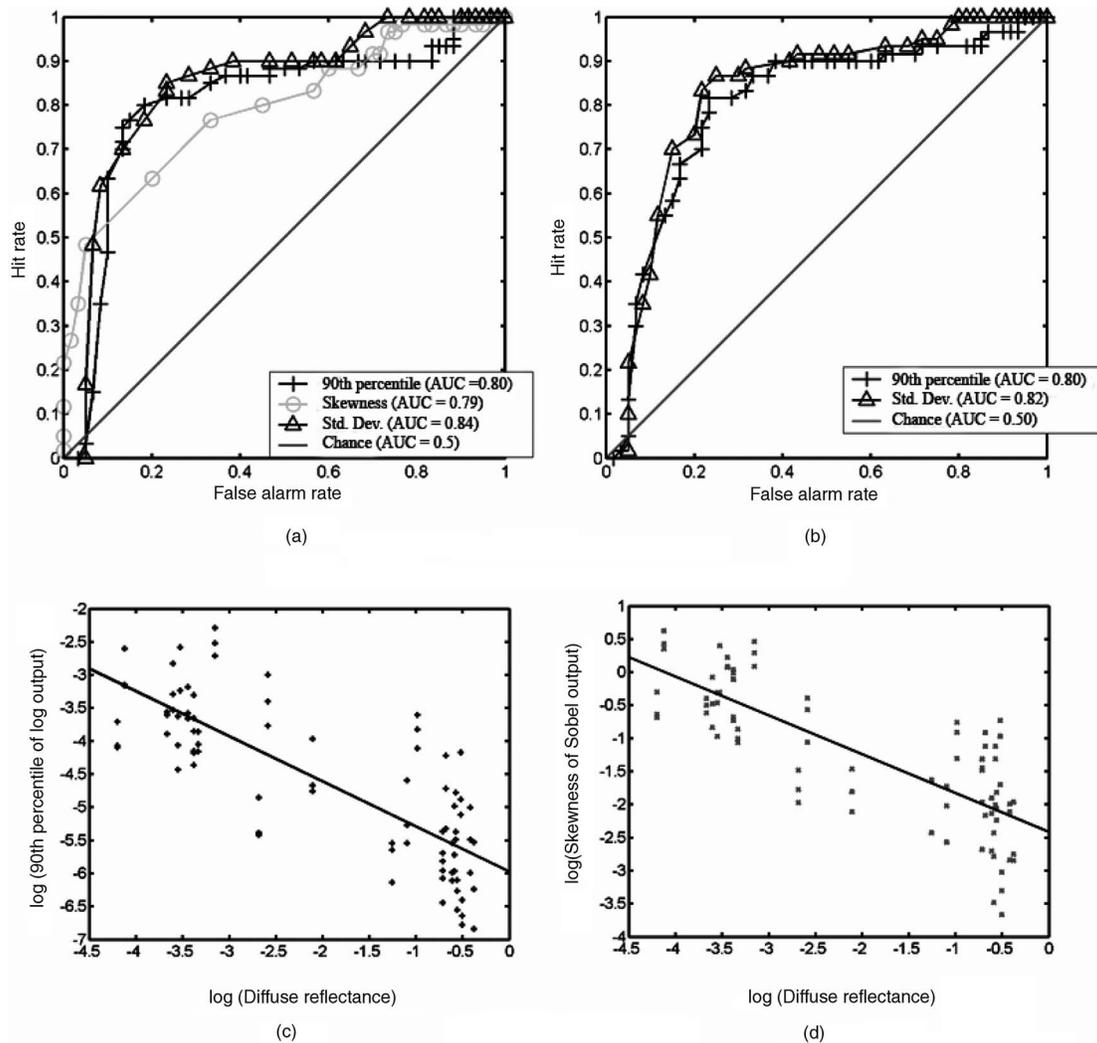


Fig. 8. ROC curves for statistics of filter outputs: (a) 90th percentile, skewness, and standard deviation of LoG filter output; (b) 90th percentile and standard deviation for Sobel filter output are significantly above chance; (c), (d) plot the log of the 90th percentile of the LoG and Sobel filter 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  $512 \times 512$  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  $1280 \times 1024$  resolution, 75 Hz frame rate, and  $70 \text{ cd/m}^2$  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.

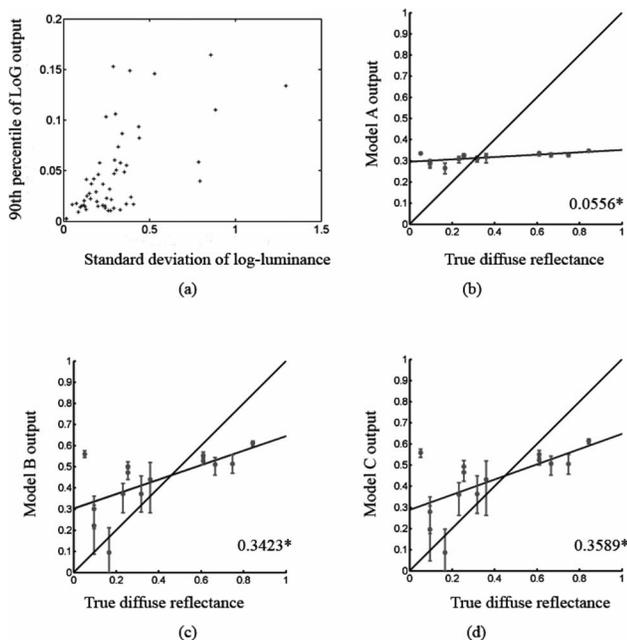


Fig. 9. (a) Standard deviation of log-luminance and the 90th percentile of LoG filter output are correlated ( $r=0.6237$ ,  $p < 0.05$ ). (b), (c), (d) Outputs of three linear models are plotted against the true diffuse reflectance for a subset of our surfaces. Model A uses one statistic, the standard 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

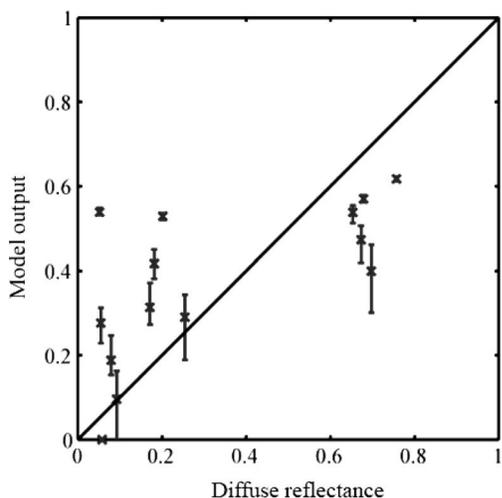


Fig. 10. Output of the regression technique is plotted versus ground truth for diffuse reflectance (albedo). Bars indicate maximum and minimum ratings. If the technique were perfect, all points would lie along the diagonal.

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.

### 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-

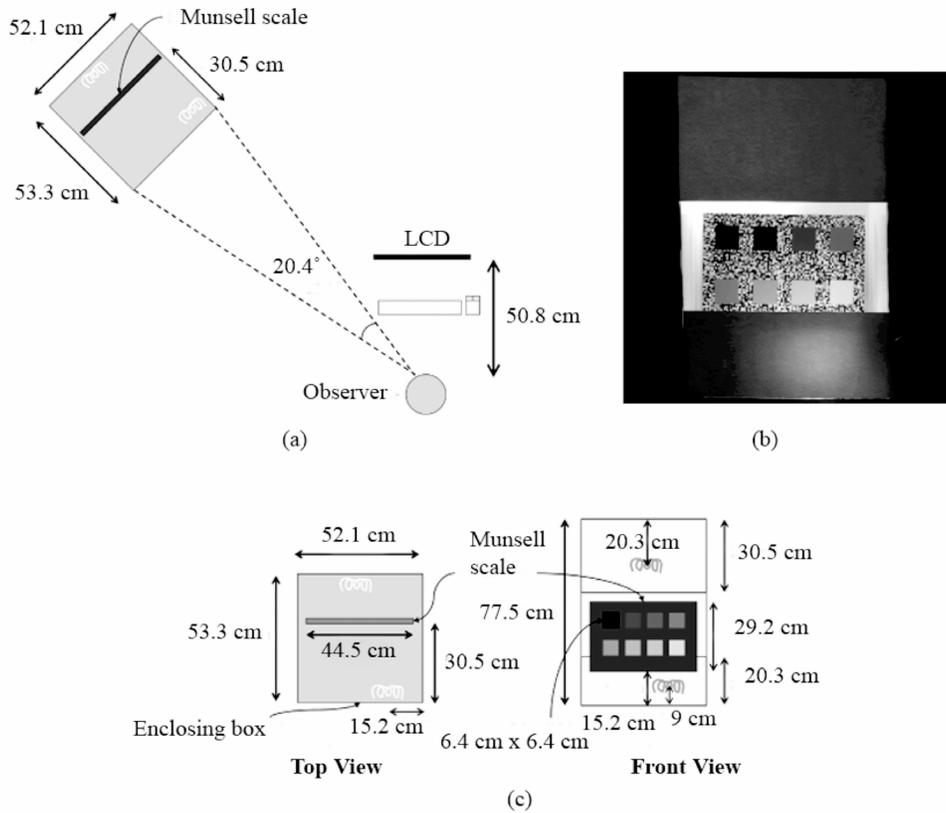


Fig. 11. Observers viewed stimuli on the LCD panel and matched reflectance properties of the surface on the screen to one of the standard surfaces on the Munsell chart. (a) Dimensions of experimental layout, (b) photograph of the reference box, (c) dimensions of the box.

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.

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

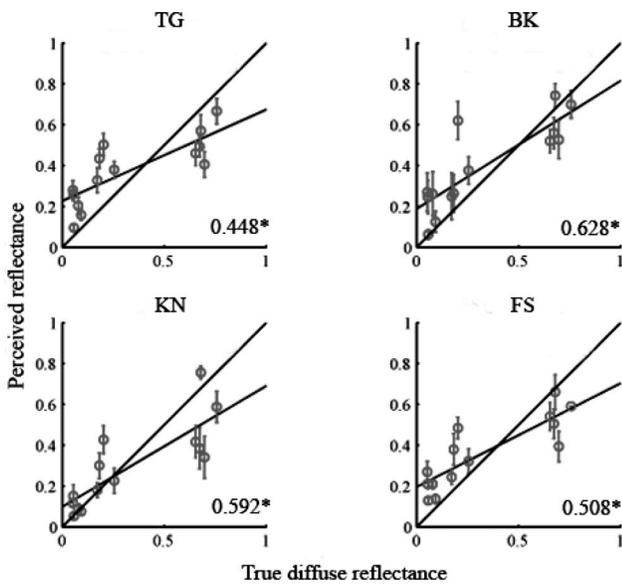


Fig. 12. Results of experiment I. Perceived albedo versus ground truth for four observers. Responses were pooled across all lighting conditions. Error bars indicate 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope=1. The gray line is the linear 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).

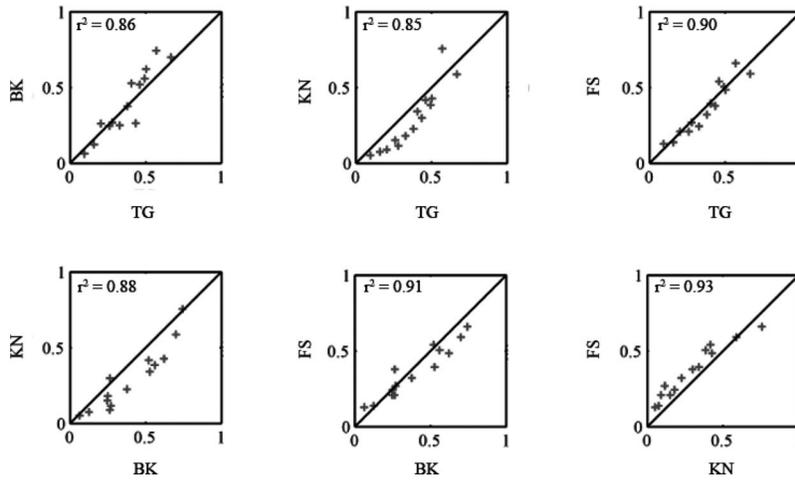
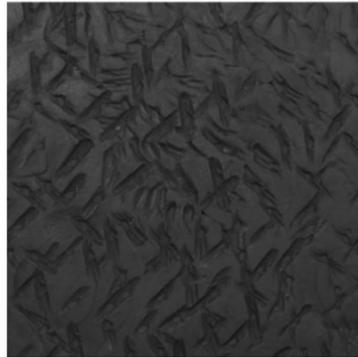
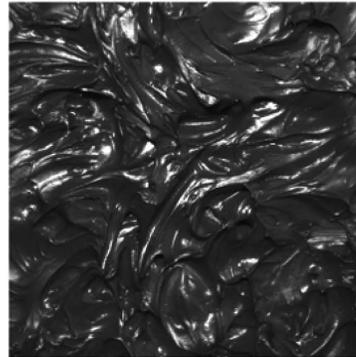


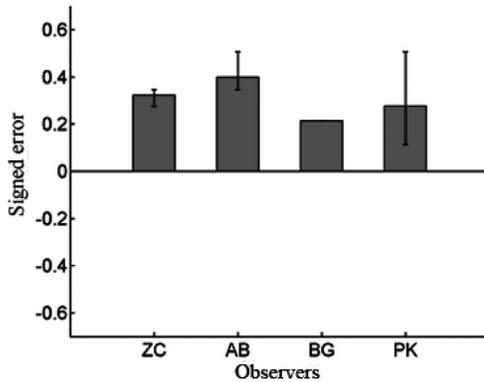
Fig. 13. Observers tend to agree with one another. Perceived reflectance ratings for every pair of observers from Fig. 12 are plotted here. If all observers behaved in the same way, all data points would lie on the black lines with slope=1;  $r^2$  values indicate that there is a great deal of agreement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



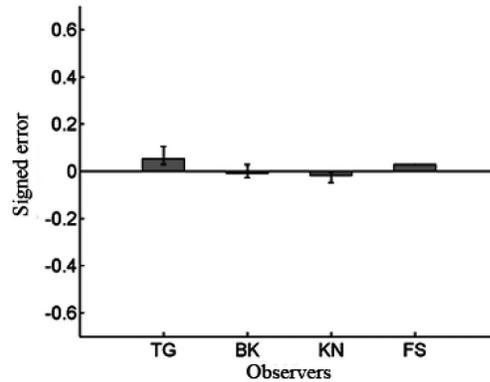
(a)



(b)



(c)



(d)

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-

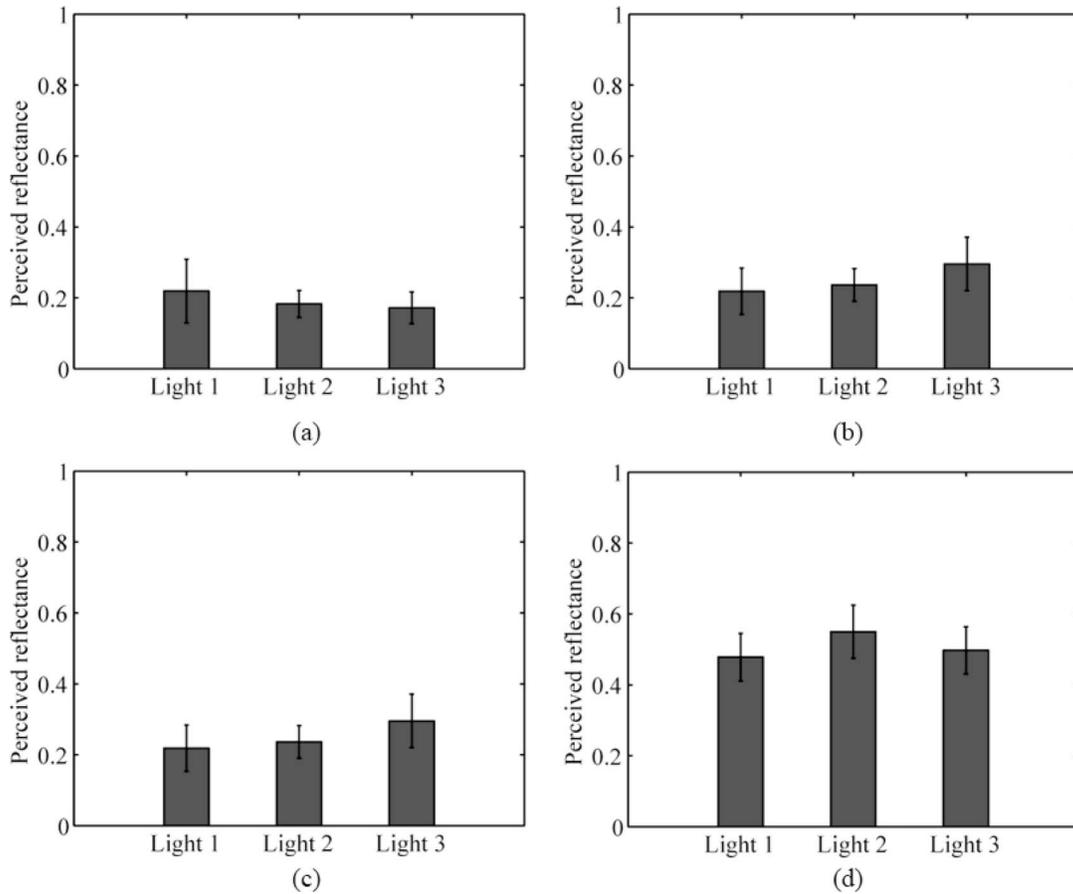


Fig. 15. Effect of lighting conditions. Light 1 is the overhead fluorescent light, light 2 is the halogen spotlight, and light 3 is the diffuse halogen source. (a)–(d) show the averaged observer ratings for four surfaces. Error bars indicate 95% confidence intervals.

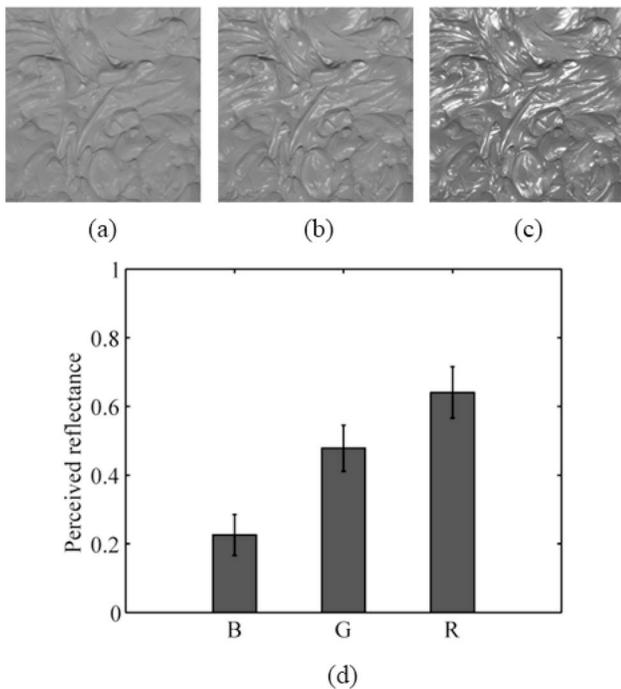


Fig. 16. Observers do not rate individual color channels in the same way. (a) Red channel, (b) green channel, and (c) B channel of an orange surface and the respective averaged observer ratings. (d) Even with the same mean image luminance, illumination conditions, and surface geometry, observers can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

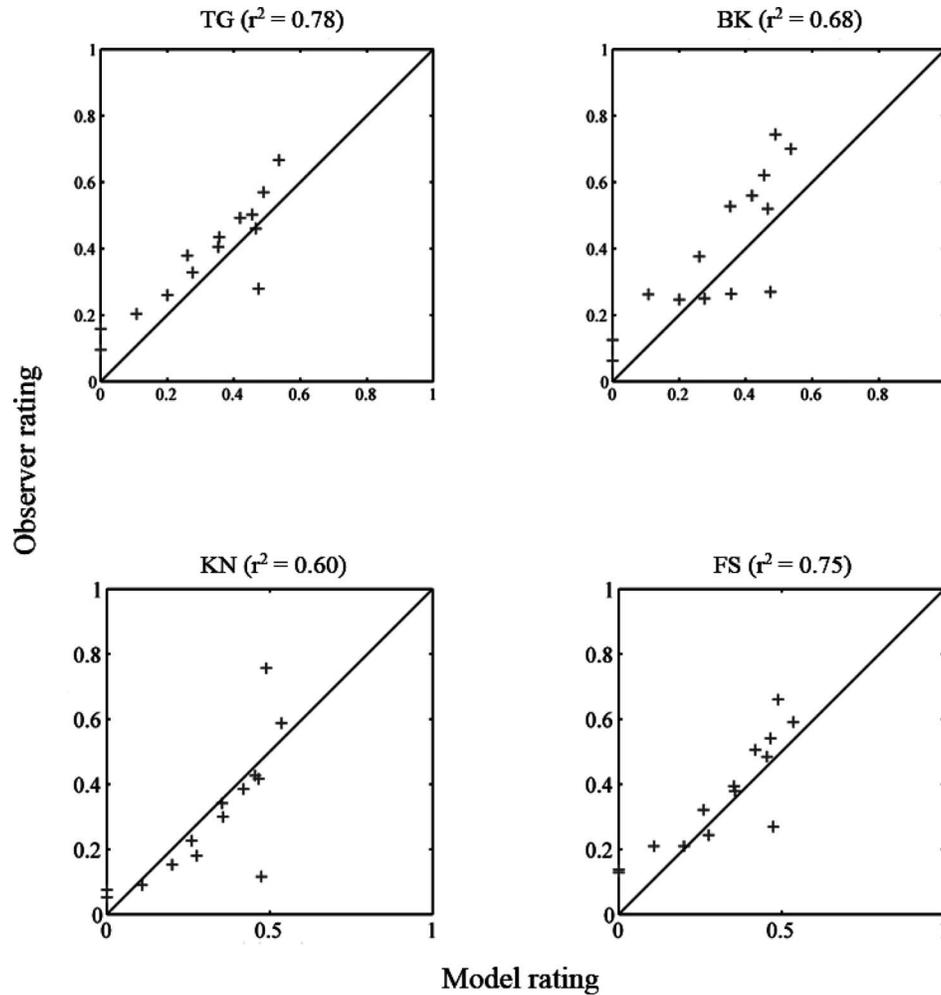


Fig. 17. Agreement between observers and the regression technique (the model) is fairly high. The  $r^2$  values here are somewhat lower than the agreement between observers but are not very different (Fig. 13).

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-

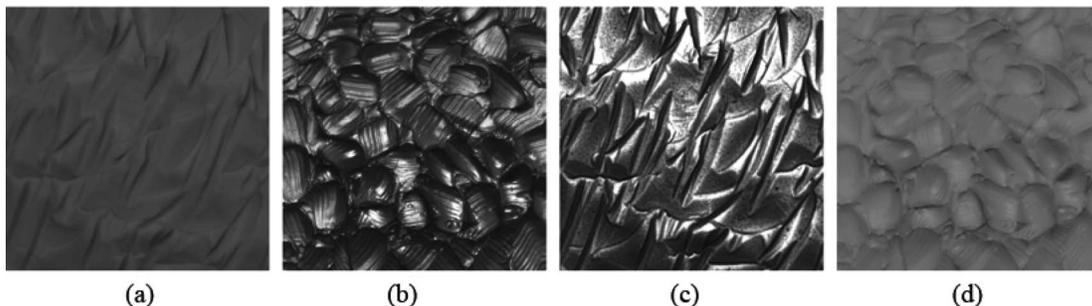


Fig. 18. (a) Light surface and (b) dark surface. (c) Result of matching histogram statistics of (a) to those of (b) and (d) and vice versa.

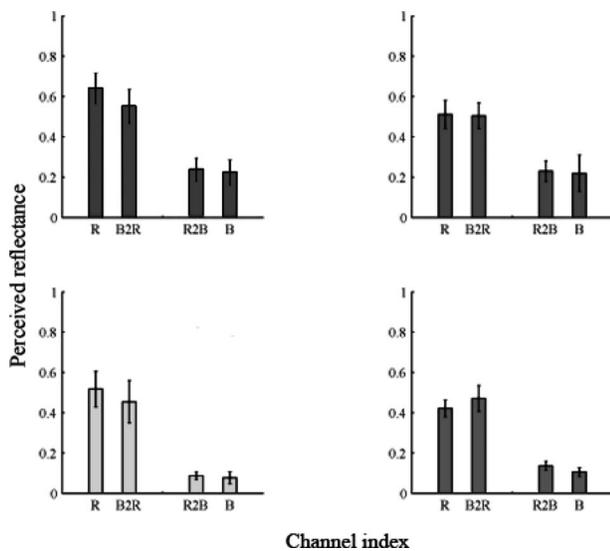


Fig. 19. Observer data for histogram-manipulated images (a) through (d) show four different surfaces in one group. The channels R and B refer to the red and blue color channels, respectively, of the color photographs of each surface. The R2B channel is the result of matching the histogram statistics 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 diagnos-

tic 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 corre-

lated 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).

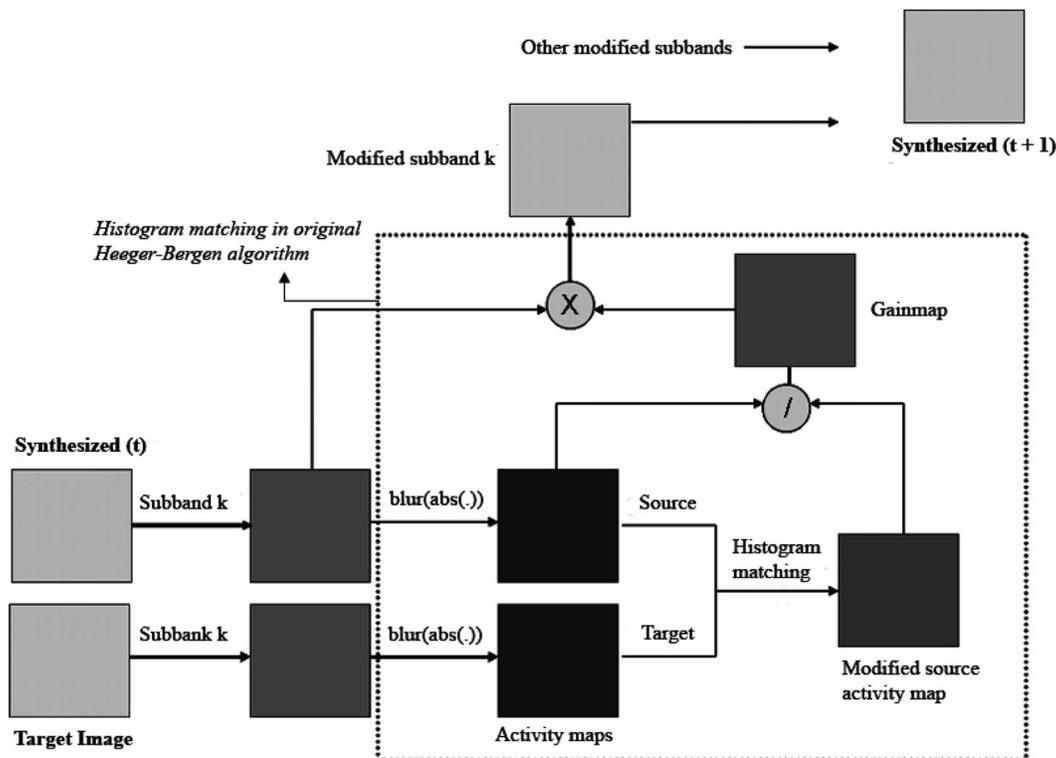


Fig. 20. Block diagram of the activity-map-based Heeger–Bergen technique.

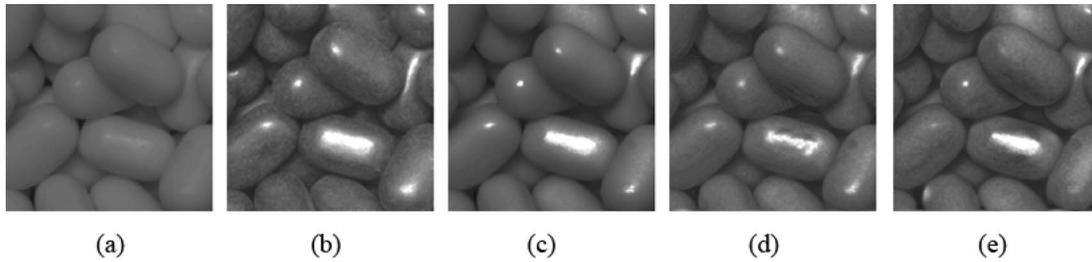


Fig. 21. Comparison of histogram-matching techniques: (a) source image, (b) target image, (c) luminance histogram matching, (d) Heeger–Bergen output, (e) activity-map-based Heeger–Bergen output.

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  $A_{\text{orig}}$  and the histogram matched source activity map be  $A_{\text{modified}}$ . Then the gain map  $G$  is calculated as

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

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  $640 \times 480$  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  $50 \times 50$  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  $100 \times 100$  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  $100 \times 100$  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  $\rho$  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  $512 \times 512$  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

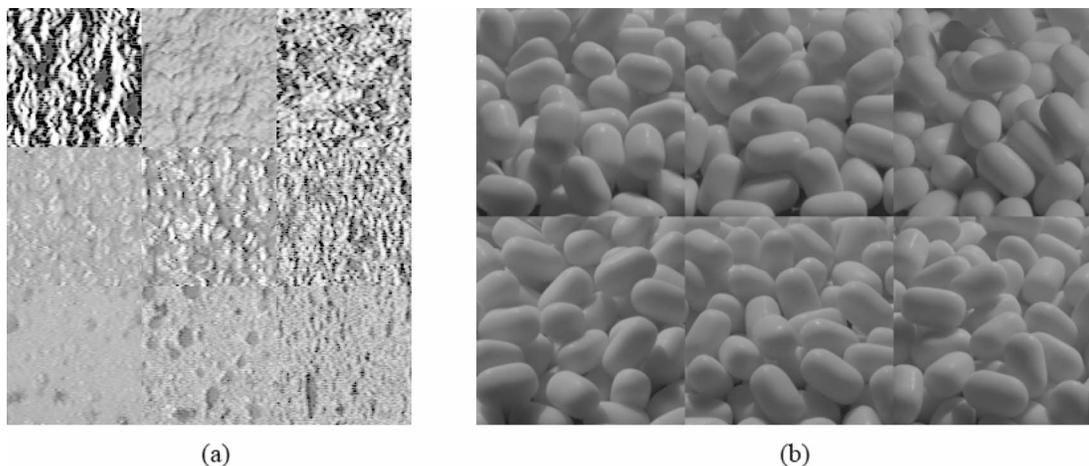


Fig. 22. (a) Some of the CURET images we used. The three rows correspond to materials 11 (plaster), 18 (rug), and 21 (sponge). Three different views are shown for each material. The images were multiplicatively normalized to have the same mean luminance. (b) One of the materials from our data set. The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

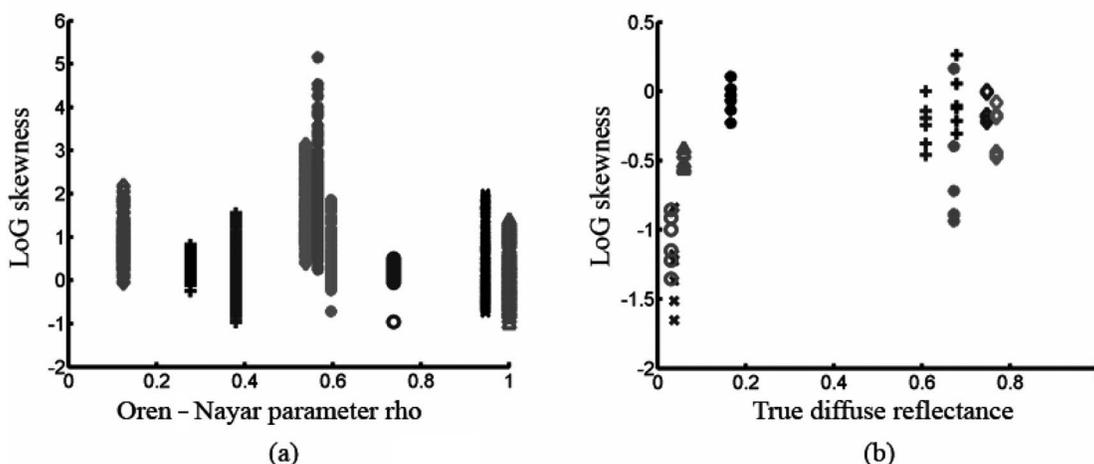


Fig. 23. Effect of varying illumination and viewpoint on image statistics. The skewness of center-surround filtered images is plotted against the ground truth for diffuse reflectance. A Laplacian of a Gaussian filter was used ( $\sigma=0.5$ , size  $5 \times 5$  pixels). (a) Skewness statistics is plotted 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.

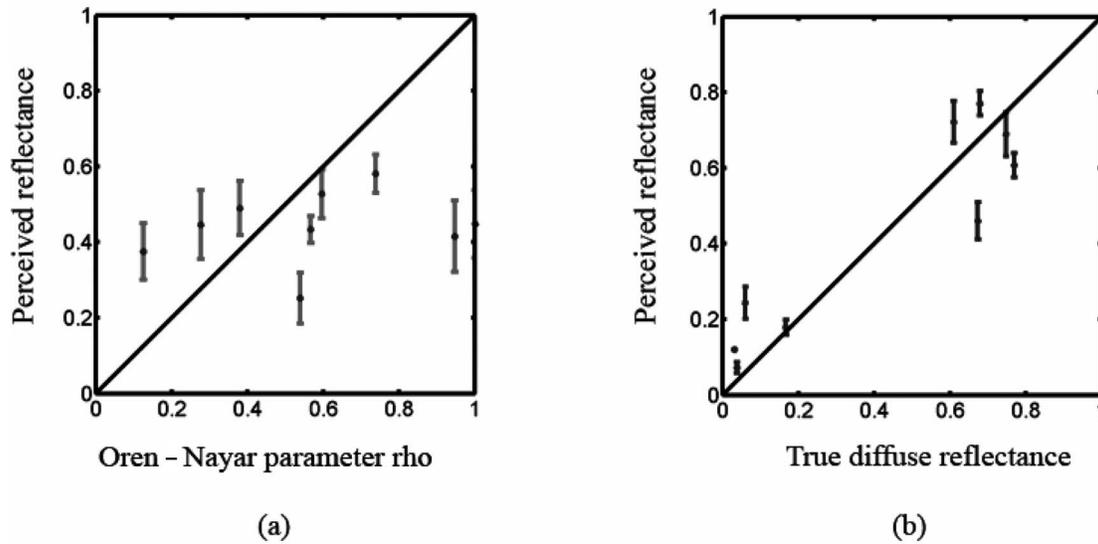


Fig. 24. Perceived reflectance for an observer is plotted against the ground truth. For each material, observer ratings were pooled across all viewing and illumination directions. Error bars are 95% confidence intervals. (a) CURET images: There is no linear relationship between perceived and true reflectance ( $p > 0.05$ ). (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 cho-

sen 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

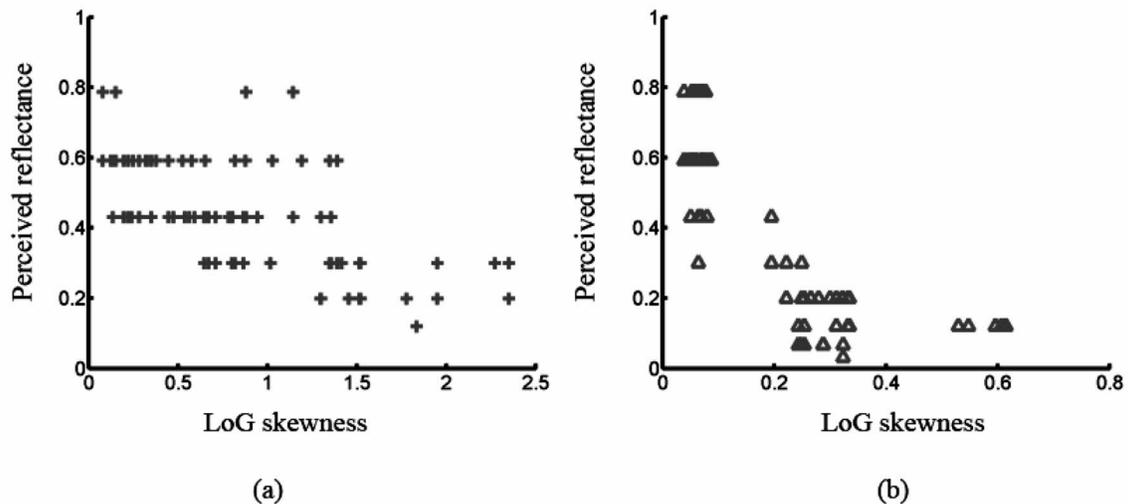


Fig. 25. Perceived reflectance of an image of a surface is plotted against the skewness of the center-surround filter output. A Laplacian of a Gaussian filter was used ( $\sigma = 0.5$ , size  $5 \times 5$  pixels). (a) CURET images and (b) our images. A linear relationship can be observed in both plots.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

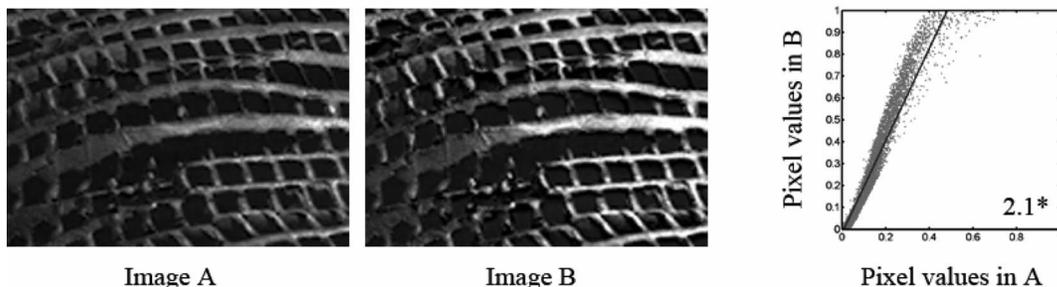


Fig. 26. Verifying linearity correction. Images A and B are the linear outputs of the ddraw program for two exposures of the same scene [37]. The original exposures were recorded in the Canon RAW format (CRW). The exposure time for B was twice that of A. All other camera parameters were the same for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

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 ddraw software [37]; ddraw 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 ddraw 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 ddraw 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 ddraw 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= $2.1055 \pm 0.0084$  ( $p < 0.05$ ,  $r^2 = 0.94$ ). We repeated this analysis for other image pairs for each of our scenes and verified that the accuracy of ddraw 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 supported 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](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).